

# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



COLEGIO DE CIENCIAS Y HUMANIDADES  
PLANTEL VALLEJO  
ÁREA DE MATEMÁTICAS  
ESTADÍSTICA Y PROBABILIDAD I



## ¿QUÉ ES LA ESTADÍSTICA?<sup>1</sup>

Cuando coloquialmente se habla de estadística, se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, ya que hoy día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc., no nos aborde diariamente con cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, etc.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Medicina, Biología, Psicología, ... empezamos a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes deterministas. Podríamos, desde un punto de vista más amplio, definir la estadística como la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.

La **Estadística** se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

Podríamos por tanto clasificar la Estadística en descriptiva, cuando los resultados del análisis no pretenden ir más allá del conjunto de datos, e inferencial cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio.

**Estadística descriptiva:** Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

**Estadística inferencial:** Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

Establecemos a continuación algunas definiciones de conceptos básicos y fundamentales básicas como son: elemento, población, muestra, caracteres, variables, etc., a las cuales haremos referencia continuamente a lo largo del texto.

**Individuos o elementos:** personas u objetos que contienen cierta información que se desea estudiar.

**Población:** conjunto de individuos o elementos que cumplen ciertas propiedades comunes.

**Muestra:** subconjunto representativo de una población.

<sup>1</sup> *Bioestadística: Métodos y Aplicaciones*. Fco. Javier Barón López. Universidad de Malaga  
Recopilación: **Eleazar Gómez Lara**. Agosto/07

**Parámetro:** función definida sobre los valores numéricos de características medibles de una población.

**Estadístico:** función definida sobre los valores numéricos de una muestra.

En relación al tamaño de la población, ésta puede ser:

- **Finita**, como es el caso del número de personas que llegan al servicio de urgencia de un hospital en un día;
- **Infinita**, si por ejemplo estudiamos el mecanismo aleatorio que describe la secuencia de caras y cruces obtenida en el lanzamiento repetido de una moneda al aire.

**Caracteres:** propiedades, rasgos o cualidades de los elementos de la población. Estos caracteres pueden dividirse en cualitativos y cuantitativos.

**Modalidades:** diferentes situaciones posibles de un carácter. Las modalidades deben ser a la vez exhaustivas y mutuamente excluyentes —cada elemento posee una y sólo una de las modalidades posibles.

**Clases:** conjunto de una o más modalidades en el que se verifica que cada modalidad pertenece a una y sólo una de las clases.

### **Variables estadísticas**

Cuando hablemos de **variable** haremos referencia a un símbolo ( $X, Y, A, B, \dots$ ) que puede tomar cualquier **modalidad** (valor) de un conjunto determinado, que llamaremos **dominio de la variable** o **rango**. En función del tipo de dominio, las variables las clasificamos del siguiente modo:

**Variables cualitativas**, cuando las modalidades posibles son de tipo nominal. Por ejemplo, el grupo sanguíneo tiene por modalidades: Grupos Sanguíneos posibles:  $A, B, AB, O$

**Variables ordinales** son las que, aunque sus modalidades son de tipo nominal, es posible establecer un orden entre ellas. Por ejemplo, si estudiamos el grado de recuperación de un paciente al aplicarle un tratamiento, podemos tener como modalidades: Grado de recuperación: *Nada, Poco, Moderado, Bueno, Muy Bueno*.

A veces se representan este tipo de variables en escalas numéricas, por ejemplo, puntuar el dolor en una escala de 1 a 5. Debemos evitar sin embargo realizar operaciones algebraicas con estas cantidades. ¡Un dolor de intensidad 4 *NO* duele el doble que otro de intensidad 2!

**Variables cuantitativas o numéricas** son las que tienen por modalidades cantidades numéricas con las que podemos hacer operaciones aritméticas. Dentro de este tipo de variables podemos distinguir dos grupos:

**Discretas**, cuando no admiten siempre una modalidad intermedia entre dos cualesquiera de sus modalidades. Un ejemplo es el número de hijos en una población de familias:

Número de hijos posibles:  $0, 1, 2, 3, 4, 5, \dots$

**Continuas**, cuando admiten una modalidad intermedia entre dos cualesquiera de sus modalidades, por ejemplo: el peso  $X$  de un niño al nacer.

Ocurre a veces que una variable cuantitativa continua por naturaleza, aparece como discreta. Este es el caso en que hay limitaciones en lo que concierne a la precisión del aparato de medida de esa variable, por ejemplo: si medimos la altura en metros de personas con una regla que ofrece dos decimales de precisión, podemos obtener

Alturas medidas en m.:  $1.50, 1.51, 1.52, 1.53, \dots$

En realidad lo que ocurre es que con cada una de esas mediciones expresamos que el verdadero valor de la misma se encuentra en un intervalo de radio 0.005. Por tanto cada una de las observaciones de  $X$  representa más bien un intervalo que un valor concreto.

Tal como hemos citado anteriormente, las modalidades son las diferentes situaciones posibles que puede presentar la variable. A veces éstas son muy numerosas (v.g. cuando una variable es continua) y conviene reducir su número, agrupándolas en una cantidad inferior de clases. Estas clases deben ser construidas, tal como hemos citado

anteriormente, de modo que sean *exhaustivas* y *excluyentes*, es decir, cada modalidad debe pertenecer a una y sólo una de las clases.

### **Estadística: Conceptos Básicos<sup>2</sup>**

La estadística es comúnmente considerada como una colección de hechos numéricos expresados en términos de una relación sumisa, y que han sido recopilados a partir de otros datos numéricos.

Kendall y Buckland (citados por Gini V. Glas / Julian C. Stanley, 1980) definen la estadística como un valor resumido, calculado, como base en una muestra de observaciones que generalmente, aunque no por necesidad, se considera como una estimación de parámetro de determinada población; es decir, una función de valores de muestra.

“La estadística es una técnica especial apta para el estudio cuantitativo de los fenómenos de masa o colectivo, cuya mediación requiere una masa de observaciones de otros fenómenos más simples llamados individuales o particulares”. (Gini, 1953.

Murray R. Spiegel, (1991) dice: “La estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis.

“La estadística es la ciencia que trata de la recolección, clasificación y presentación de los hechos sujetos a una apreciación numérica como base a la explicación, descripción y comparación de los fenómenos”. (Yale y Kendal, 1954).

Cualquiera sea el punto de vista, lo fundamental es la importancia científica que tiene la estadística, debido al gran campo de aplicación que posee.

#### **Población:**

El concepto de población en estadística va más allá de lo que comúnmente se conoce como tal. Una población se precisa como un conjunto finito o infinito de personas u objetos que presentan características comunes.

“Una población es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones”. Levin & Rubin (1996).

“Una población es un conjunto de elementos que presentan una característica común”. Cadenas (1974).

El tamaño que tiene una población es un factor de suma importancia en el proceso de investigación estadística, y este tamaño vienen dado por el número de elementos que constituyen la población, según el número de elementos la población puede ser finita o infinita. Cuando el número de elementos que integra la población es muy grande, se puede considerar a esta como una población infinita, por ejemplo; el conjunto de todos los números positivos. Una población finita es aquella que está formada por un limitado número de elementos, por ejemplo; el número de estudiante del Núcleo San Carlos de la Universidad Nacional Experimental Simón Rodríguez.

Cuando la población es muy grande, es obvio que la observación de todos los elementos se dificulte en cuanto al trabajo, tiempo y costo necesario para hacerlo. Para solucionar este inconveniente se utiliza una muestra estadística.

Es a menudo imposible o poco práctico observar la totalidad de los individuos, sobre todos si estos son muchos. En lugar de examinar el grupo entero llamado **población** o **universo**, se examina una pequeña parte del grupo llamada muestra.

---

<sup>2</sup> JORGE L. CASTILLO T.

### **Muestra:**

“Se llama muestra a una parte de la población a estudiar que sirve para representarla”. Murray R. Spiegel (1991).

“Una muestra es una colección de algunos elementos de la población, pero no de todos”. Levin & Rubin (1996).

“Una muestra debe ser definida en base de la población determinada, y las conclusiones que se obtengan de dicha muestra solo podrán referirse a la población en referencia”, Cadenas (1974).

Ejemplo; El estudio realizado a 50 miembros del Colegio de Ingenieros del Estado de México.

El estudio de muestras es más sencillo que el estudio de la población completa; cuesta menos y lleva menos tiempo. Por último se aprobó que el examen de una población entera todavía permite la aceptación de elementos defectuosos, por tanto, en algunos casos, el muestreo puede elevar el nivel de calidad.

Una muestra representativa contiene las características relevantes de la población en las mismas proporciones que están incluidas en tal población.

Los expertos en estadística recogen datos de una muestra. Utilizan esta información para hacer referencias sobre la población que está representada por la muestra. En consecuencia muestra y población son conceptos relativos. Una población es un todo y una muestra es una fracción o segmento de ese todo.

### **Muestreo:**

Esto no es más que el procedimiento empleado para obtener una o más muestras de una población; el muestreo es una técnica que sirve para obtener una o más muestras de población.

Este se realiza una vez que se ha establecido un marco muestral representativo de la población, se procede a la selección de los elementos de la muestra aunque hay muchos diseños de la muestra.

Al tomar varias muestras de una población, las estadísticas que calculamos para cada muestra no necesariamente serían iguales, y lo más probable es que variaran de una muestra a otra.

**Ejemplo;** Consideremos como una población a los estudiantes del C C H, determinando por lo menos dos caracteres ser estudiados en dicha población;

- Religión de los estudiantes
- Sexo.

### **Tipos de muestreo**

Existen dos métodos para seleccionar muestras de poblaciones; el muestreo no aleatorio o de juicio y el muestreo aleatorio o de probabilidad. En este último todos los elementos de la población tienen la oportunidad de ser escogidos en la muestra. Una muestra seleccionada por muestreo de juicio se basa en la experiencia de alguien con la población. Algunas veces una muestra de juicio se usa como guía o muestra tentativa para decidir como tomar una muestra aleatoria más adelante. Las muestras de juicio evitan el análisis estadístico necesario para hacer muestras de probabilidad.

### **VARIABLES Y ATRIBUTOS:**

Las variables, también suelen ser llamados caracteres cuantitativos, son aquellos que pueden ser expresados mediante números. Son caracteres susceptibles de medición. Como por ejemplo, la estatura, el peso, el salario, la edad, etc.

Según, Murray R. Spiegel, (1992) “una variable es un símbolo, tal como X, Y, H, que puede tomar un valor cualquiera de un conjunto determinado de ellos, llamado

dominio de la variable. Si la variable puede tomar solamente un valor, se llama constante.”

Todos los elementos de la población poseen los mismos tipos de caracteres, pero como estos en general no suelen representarse con la misma intensidad, es obvio que las variables toman distintos valores. Por lo tanto estos distintos números o medidas que toman los caracteres son los “valores de la variable”. Todos ellos juntos constituyen una variable.

Los atributos también llamados caracteres cualitativos, son aquellos que no son susceptibles de medición, es decir que no se pueden expresar mediante un número.

IUTIN (1997). “Reciben el nombre de variables cualitativas o atributos, aquellas características que pueden presentarse en individuos que constituyen un conjunto.

La forma de expresar los atributos es mediante palabras, por ejemplo; profesión, estado civil, sexo, nacionalidad, etc. Puede notar que los atributos no se presentan en la misma forma en todos los elementos. Estas distintas formas en que se presentan los atributos reciben el nombre de “modalidades”.

Ejemplo; El color de ojos de cada uno de los estudiantes del curso de Estadística I, no se presenta en la misma modalidad en todos.

1. Atendiendo a la cobertura; pueden ser **exhaustiva o parcial**.

- **Observación Exhaustiva.** Cuando la observación es efectuada sobre la totalidad de los elementos de la población se habla de una observación exhaustiva.
- **Observación Parcial.** Dado que las poblaciones en general son grandes, la observación de todos sus elementos se ve imposibilitada. La solución para superar este inconveniente es observar una parte de esta población.

### **Censo:**

Se entiende por censo aquella numeración que se efectúa a todos y cada uno de los caracteres componentes de una población.

Para Levin & Rubin (1996) “Algunas veces es posible y práctico examinar a cada persona o elemento de la población que deseamos describir. A esto lo llamamos una numeración completa o censo. Utilizamos el muestreo cuando no es posible contar o medir todos los elementos de la población.

Si es posible listar (o enumerar) y observar cada elemento de la población, los censos se utilizan rara vez porque a menudo su compilación es bastante difícil, consume mucho tiempo por lo que resulta demasiado costoso.

### **Encuesta:**

Se entiende por encuesta las observaciones realizadas por muestreo, es decir son observaciones parciales.

El diseño de encuestas es exclusivo de las ciencias sociales y parte de la premisa de que si queremos conocer algo sobre el comportamiento de las personas, lo mejor, más directo y simple es preguntárselo directamente a ellas. (Cadenas, 1974).

Según Antonio Napolitano “La encuesta, es un método mediante el cual se quiere averiguar. Se efectúa a través de cuestionarios verbales o escritos que son aplicados a un gran número de personas”.

### **Estadística Descriptiva:**

Tienen por objeto fundamental describir y analizar las características de un conjunto de datos, obteniéndose de esa manera conclusiones sobre las características de dicho conjunto y sobre las relaciones existentes con otras poblaciones, a fin de compararlas. No obstante puede no solo referirse a la observación de todos los elementos de una población (observación exhaustiva) sino también a la descripción de los elementos de una muestra (observación parcial).

En relación a la estadística descriptiva, Ernesto Rivas Gonzáles dice; “Para el estudio de estas muestras, la estadística descriptiva nos provee de todos sus medidas; medidas que cuando quieran ser aplicadas al universo total, no tendrán la misma exactitud que tienen para la muestra, es decir al estimarse para el universo vendrá dada con cierto margen de error; esto significa que el valor de la medida calculada para la muestra, en el oscilará dentro de cierto límite de confianza, que casi siempre es de un 95 a 99% de los casos.

### **Estadística Inductiva:**

Está fundamentada en los resultados obtenidos del análisis de una muestra de población, con el fin de inducir o inferir el comportamiento o característica de la población, de donde procede, por lo que recibe también el nombre de Inferencia estadística.

Según Berenson y Levine; Estadística Inferencial son procedimientos estadísticos que sirven para deducir o inferir algo acerca de un conjunto de datos numéricos (población), seleccionando un grupo menor de ellos (muestra).

El objetivo de la inferencia en investigación científica y tecnológica radica en conocer clases numerosas de objetos, personas o eventos a partir de otras relativamente pequeñas compuestas por los mismos elementos.

En relación a la estadística descriptiva y la inferencial, Levin & Rubin (1996) citan los siguientes ejemplos para ayudar a entender la diferencia entre las dos.

Supóngase que un profesor calcula la calificación promedio de un grupo de historia. Como la estadística describe el desempeño del grupo pero no hace ninguna generalización acerca de los diferentes grupos, podemos decir que el profesor está utilizando estadística descriptiva. Graficas, tablas y diagramas que muestran los datos de manera que sea más fácil su entendimiento son ejemplos de estadística descriptiva.

Supóngase ahora que el profesor de historia decide utilizar el promedio de calificaciones obtenidas por uno de sus grupos para estimar la calificación promedio de las diez unidades del mismo curso de historia. El proceso de estimación de tal promedio sería un problema concerniente a la estadística inferencial.

Los estadísticos se refieren a esta rama como inferencia estadística, esta implica generalizaciones y afirmaciones con respecto a la probabilidad de su validez.

Medición Existen diversas definiciones del termino “medición”, pero estas dependen de los diferentes puntos de vista que se puedan tener al abordar el problema de la cuantificación y el proceso mismo de la construcción de una escala o instrumento de medición.

En general, se entiende por medición la asignación de números a elementos u objetos para representar o cuantificar una propiedad. El problema básico está dado por la asignación un numeral que represente la magnitud de la característica que queremos medir y que dicho números pueden analizarse por manipulaciones de acuerdo a ciertas reglas. Por medio de la medición, los atributos de nuestras percepciones se transforman en entidades conocidas y manejables llamadas “números”. Es evidente que el mundo resultaría caótico si no pudiéramos medir nada. En este caso cabría preguntarse de que le serviría la físico saber que el hierro tiene una alta temperatura de fusión.

### **Niveles o Escalas de mediciones**

#### **Escala Nominal:**

La escala de medida nominal, puede considerarse la escala de nivel más bajo, y consiste en la asignación, puramente arbitraria de números o símbolos a cada una de las diferentes categorías en las cuales podemos dividir el carácter que observamos, sin que puedan establecerse relaciones entre dichas categorías, a no ser el de que cada elemento pueda pertenecer a una y solo una de estas categorías. Se trata de agrupar objetos en clases, de modo que todos los que pertenezcan a la misma sean

equivalentes respecto del atributo o propiedad en estudio, después de lo cual se asignan nombres a tales clases, y el hecho de que a veces, en lugar de denominaciones, se le atribuyan números, puede ser una de las razones por las cuales se le conoce como “medidas nominales”.

Por ejemplo, podemos estar interesados en clasificar los estudiantes de la UNESR Núcleo San Carlos de acuerdo a la carrera que cursan.

Carrera	Número asignada a la categoría
Educación	1
Administración	2

Se ha de tener presente que los números asignados a cada categoría sirven única y exclusivamente para identificar la categoría y no poseen propiedades cuantitativas.

### **Escala Ordinal:**

En caso de que puedan detectarse diversos grados de un atributo o propiedad de un objeto, la medida ordinal es la indicada, puesto que entonces puede recurrirse a la propiedad de “orden” de los números asignándolo a los objetos en estudio de modo que, si la cifra asignada al objeto A es mayor que la de B, puede inferirse que A posee un mayor grado de atributo que B.

La asignación de números a las distintas categorías no puede ser completamente arbitraria, debe hacerse atendiendo al orden existente entre éstas.

Los caracteres que posee una escala de medida ordinal permiten, por el hecho mismo de poder ordenar todas sus categorías, el cálculo de las medidas estadísticas de posición, como por ejemplo la mediana.

Ejemplo: Al asignar un número a los pacientes de una consulta médica, según el orden de llegada, estamos llevando una escala ordinal, es decir que al primero en llegar le asignamos el nº 1, al siguiente el nº 2 y así sucesivamente, de esta forma, cada número representará una categoría en general, con un solo elemento y se puede establecer relaciones entre ellas, ya que los números asignados guardan la misma relación que el orden de llegada a la consulta.

### **Escalas de intervalos iguales:**

La escala de intervalos iguales, está caracterizada por una unidad de medida común y constante que asigna un número igual al número de unidades equivalentes a la de la magnitud que posea el elemento observado. Es importante destacar que el punto cero en las escalas de intervalos iguales es arbitrario, y no refleja en ningún momento ausencia de la magnitud que estamos midiendo. Esta escala, además de poseer las características de la escala ordinal, encontramos que la asignación de los números a los elementos es tan precisa que podemos determinar la magnitud de los intervalos (distancia) entre todos los elementos de la escala. Sin lugar a dudas, podemos decir que la escala de intervalos es la primera escala verdaderamente cuantitativa y a los caracteres que posean esta escala de medida pueden calcularse todas las medidas estadísticas a excepción del coeficiente de variación.

Ejemplo: El lapso transcurrido entre 1998-1999 es igual al que transcurrió entre 2000-2001.

### **Escala de coeficientes o Razones:**

El nivel de medida más elevado es el de cocientes o razones, y se diferencia de las escalas de intervalos iguales únicamente por poseer un punto cero propio como origen; es decir que el valor cero de esta escala significa ausencia de la magnitud que estamos

mediendo. Si se observa una carencia total de propiedad, se dispone de una unidad de medida para el efecto. A iguales diferencias entre los números asignados corresponden iguales diferencias en el grado de atributo presente en el objeto de estudio. Además, siendo que cero ya no es arbitrario, sino un valor absoluto, podemos decir que A. Tiene dos, tres o cuatro veces la magnitud de la propiedad presente en B.

Ejemplo: En una encuesta realizada en un barrio de esta localidad se observó que hay familias que no tienen hijos, otras tienen 6 hijos que es exactamente el doble de hijos que aquellas que tienen 3 hijos.

### Las variables y su medición:

Una variable es un símbolo, tal como X, Y, H, x ó B, que pueden tomar un conjunto prefijado de valores, llamado dominio de esa variable. Para Murray R. Spiegel (1991) "una variable que puede tomar cualquier valor entre dos valores dados se dice que es una variable continua en caso contrario diremos que la variable es discreta".

Las variables, también llamadas caracteres cuantitativos, son aquellas cuyas variaciones son susceptibles de ser medidas cuantitativamente, es decir, que pueden expresar numéricamente la magnitud de dichas variaciones. Por intuición y por experiencia sabemos que pueden distinguirse dos tipos de variables; las continuas y las discretas

Las variables continuas se caracterizan por el hecho de que para todo par de valores siempre se puede encontrar un valor intermedio, (el peso, la estatura, el tiempo empleado para realizar un trabajo, etc.)

Una variable es continua, cuando puede tomar infinitos valores intermedios dentro de dos valores consecutivos. Por ejemplo, la estatura, el peso, la temperatura.

Ejemplo: En el preescolar Blanca de Pérez, ubicado en la urbanización Monseñor Padilla de esta ciudad se procedió a recoger las medidas de talla y peso de los niños que a este asisten.

Niño	Peso	Talla
José	18.30	1.15
Julio	20.50	1.20
Pedro	19.00	1.10
Luís	18.75	1.18

Las variables discretas serán aquellas que pueden tomar solo un número limitado de valores separados y no continuos; son aquellas que solo toman un determinado número de valores, porque entre dos valores consecutivos no pueden tomar ningún otro; por ejemplo el número de estudiantes de una clase es una variable discreta ya que solo tomará los valores 1, 2, 3, 4... Nótese que no encontramos valor como 1.5 estudiantes

### Datos Estadísticos:

Los datos estadísticos no son otra cosa que el producto de las observaciones efectuadas en las personas y objetos en los cuales se produce el fenómeno que queremos estudiar. Dicho en otras palabras, son los antecedentes (en cifras) necesarios para llegar al conocimiento de un hecho o para reducir las consecuencias de este.

Los datos estadísticos se pueden encontrar de forma no ordenada, por lo que es muy difícil en general, obtener conclusiones de los datos presentados de esta manera. Para poder obtener una precisa y rápida información con propósitos de descripción o

análisis, estos deben organizarse de una manera sistemática; es decir, se requiere que los datos sean clasificados. Esta clasificación u organización puede muy bien hacerse antes de la recopilación de los datos.

Ejemplo:

Si se quiere conocer las características de los estudiantes del Colegio San Carlos de la UNAM, que solicitan préstamo a la biblioteca de dicho Colegio la recolección de la información debe clasificar a cada estudiante sobre la base de: Carrera que estudia, edad, semestre de estudios, etc. Vemos pues que la clasificación marca la pauta de la clase de datos que debe ser obtenido.

Clasificación de los datos

Los datos estadísticos pueden ser clasificados en cualitativos y cuantitativos.

Datos Cualitativos: cuando los datos son cualitativos, la diferencia entre ellos es de clase y no de cantidad.

Ejemplo: Si deseamos clasificar los estudiantes que cursan la materia de estadística I, de la Maestría de Ingeniería Mecánica, por su estado civil, observamos que pueden existir solteros, casados, divorciados, viudos.

Datos cuantitativos: cuando los valores de los datos representan diferentes magnitudes, decimos que son datos cuantitativos.

Ejemplo: Se clasifican los estudiantes del Colegio San Carlos de de acuerdo a sus calificaciones, observamos que los valores (calificación) representan diferentes magnitudes.

## ¿PARA QUÉ SIRVE LA ESTADÍSTICA?

La Estadística puede dar respuesta a muchas de las necesidades que la sociedad actual nos plantea. Su tarea fundamental es la reducción de datos, con el objetivo de representar la realidad y transformarla, predecir su futuro o simplemente conocerla.

**La Estadística responde a las necesidades fiscales de los gobernantes.** Esto se puede conseguir con un conocimiento claro de la población con la que se cuenta. La herramienta para conseguirlo es el CENSO DE POBLACIÓN y su hermano pequeño, el PADRÓN MUNICIPAL DE HABITANTES.

La práctica del recuento de la población y de algunas características de esta por los Estados es muy antigua (se remonta a 3000 años antes de Cristo en Egipto y Mesopotamia). En palabras de Bielfed, *la Estadística es la ciencia que nos enseña el ordenamiento político de todos los estados del mundo conocido*, es decir, está al servicio del Estado, de hecho, la palabra Estadística deriva de Estado.

**La Estadística responde a la actividad planificadora de la sociedad.** Con la Revolución Industrial aparecen nuevos problemas, sobre todo de desigualdades sociales. La Estadística es un instrumento para identificar estas injusticias y para producir información en el llamado Estado del Bienestar.

**La Estadística responde a nuevas demandas sociales.** Para realizar investigaciones exhaustivas sobre temas sociales surgen tres problemas básicos a la hora del trabajo de campo, como el tiempo que tardaríamos en entrevistar a toda la población y el costo económico y de personal de estas entrevistas. Con las técnicas de MUESTREO se consigue hacer buenas investigaciones sobre una pequeña parte de esa población, obteniendo resultados válidos para toda ella.

**La Estadística responde a las necesidades del desarrollo científico y tecnológico de la sociedad.** Tras la Revolución Industrial se produce un desarrollo de la sociedad en todos sus ámbitos y, en particular, en el Científico y Tecnológico. Las Comunicaciones, la Industria, la Agricultura, la Salud... se desarrollan rápidamente y se exige el máximo rendimiento y la mejor utilización de estos sectores.

Las técnicas de **Investigación de Mercados** permiten saber si un producto cualquiera será bien acogido en el mercado antes de su salida a este, o bien medir la audiencia en Televisión y Radio.

El **Control de Calidad** permite medir las características de la calidad de un producto, compararlas con ciertos requisitos y tomar decisiones correctivas si hay diferencias entre el funcionamiento real y el esperado. Con estudios estadísticos aplicados a la Agricultura y a la Pesca podemos estimar los rendimientos obtenidos en una cosecha, o encontrar bancos de peces...

En **Medicina e Investigación** farmacológica es imprescindible la Estadística, probando nuevos tratamientos en grupos de pacientes o bien, obteniendo conclusiones sobre ciertas enfermedades observando durante un tiempo un grupo de pacientes (saber si para el tratamiento de cierto tipo de cáncer es más efectiva la cirugía, la radioterapia o la quimioterapia, sin más que observar un grupo de pacientes tratados con estas técnicas).

Con el estudio de los **Procesos Estocásticos** se puede tener una mejor comprensión de fenómenos de comportamiento aleatorio como meteorología, física nuclear, campañas de seguridad...

## ¿MIENTEN LAS ESTADÍSTICAS?

Cuando través de una muestra pretendemos obtener información de una población entera los datos obtenidos pueden ser diferentes a los reales. Son valores aproximados del parámetro desconocido. A estos valores se les llama Estimaciones.

Al dar una estimación estoy cometiendo un error llamado error de muestreo debido a que no estoy considerando a toda la población, sino a una parte de ella. Existen procedimientos que pueden determinar de antemano el error que puedo cometer.

Aparte de estos, existen otros errores que se presentan tanto en encuestas por muestreo como en las encuestas por censo. Este tipo de errores son mayores y de difícil corrección. Son errores ajenos al muestreo. Se dividen en:

**Errores de observación:** Debidos a la recogida, registro o procesamiento incorrecto de los datos. Pueden ser de **sobrecobertura**, cuando el listado de entrevistados contiene unidades que no pertenecen a la muestra investigada, de **medida**, que son la diferencia entre el valor observado y el verdadero, y errores de **procesamiento**, debidos a los errores de entrada de datos, edición, tabulación y análisis.

**Errores no de observación:** No es posible obtener la información deseada para ciertos individuos de la población. Estos son errores de **cobertura**, cuando hay una parte de la población que no está en el listado, errores de **falta de respuesta** por parte del entrevistado.

El principal problema de las encuestas es la falta de respuesta por parte del entrevistado. Suelen ser personas que se consideran acosadas para que proporcionen información de su entorno social y de sus actividades. Esto produce un rechazo a

responder cuestiones sobre las que en la mayoría de los casos se desconoce su utilidad.

Para solventar estos problemas debe contarse con la formación de los entrevistadores, conociendo estos perfectamente el tema que están tratando y estar preparados para responder cualquier cuestión sobre el tema que puede plantear el entrevistado.

La formulación de las preguntas debe ser clara, poco influenciadas y cómodas para los entrevistados. En temas polémicos como el consumo de drogas o las afinidades políticas, el entrevistado suele ser reacio a contestar. De manera sugestiva hay que formular ese tipo de preguntas para que el entrevistado no se vea comprometido ni violento con su contenido.

Por último, decir que hay otros tipos de falta de respuesta en una encuesta debidos a situaciones como:

- Ausencia temporal del individuo seleccionado para entrevistar (los no en casa).
- Negativa absoluta a colaborar (los hueso duro)
- Falta de conocimientos del entrevistado o incapacidad del entrevistador para explicar el contenido de la respuesta.
- Pérdida involuntaria de la información.
- No cubrimiento debido a condiciones ambientales, escasa facilidad de transporte...

### Un panorama de las aplicaciones en los negocios<sup>3</sup>.

Hay varias áreas donde se usa la estadística en forma más notoria

- mercadeo e investigación de opiniones.
- control de calidad industrial (control de procesos, muestreo de inspección).
- en responsivas legales.
- diseño y análisis de experimentos en algunas industrias, a veces conducente a la mejora de la calidad.

La estadística se usa en la presentación de resultados, en reportes anuales o de avance.

- se usa en pronósticos económicos para planeación estratégica o a largo plazo.
- muestreos varios aunque no siempre estadísticos (mercadeo y auditoría).
- investigación de operaciones (en compañías grandes).

Diez tópicos generales de relevancia para los negocios.

1. Muestreo de poblaciones finitas y encuestas.
2. Teoría estadística de decisiones.
3. Análisis de series de tiempo autocorrelacionadas para pronóstico.
4. Diseño y análisis de experimentos para la toma de decisiones en negocios.
5. Inferencia sobre datos no experimentales.
6. La mejora de la calidad y la productividad.
7. Mejora en gráficas y tablas para comunicar información y para análisis.
8. El análisis exploratorio de datos.
9. Mezcla de información de diferentes fuentes.
10. Simulación de negocios.

La "revolución" de la calidad que ha mostrado que sí se puede entrenar masivamente a la gente para que haciendo uso de técnicas estadísticas sencillas tenga una influencia en la mejora de las empresas. La cuestión de la calidad es, frecuentemente, un síntoma de una mejora en la administración de la empresa.

---

<sup>3</sup> Dr. Carlos E. Mendoza Durán.

Las aplicaciones no las suelen hacer estadísticos. Son muy raras las empresas en que hay estadísticos "de planta", en la mayoría de los casos en que los estadísticos tienen que ver con la industria es a nivel de asesoría. La toma de datos y el análisis suelen hacerlo no-estadísticos.

Los paquetes de estadística y las PC. Qué se puede hacer por los negocios.

Hay dos tipos de uso, importantes

- vigilancia y mejora de procesos entendida en sentido amplio.
- mejora del producto y del diseño, no sólo para la industria sino para los servicios.

En 6 áreas de utilidad:

1) *Descripción de resultados pasados*. En la mayoría de los casos se tienen registros que es preciso mantener y resumir de manera efectiva a la gerencia o dirección.

2) *Pronóstico*. Para prever y planear. Se necesita algún tipo de pronóstico numérico para poder hacer estas actividades. Hacer un pronóstico no es adivinar. Para pronosticar se toma información del pasado, muchas veces directamente (a veces indirectamente), y se usa algún algoritmo para llevarla al futuro. La mayoría de las veces los pronósticos en negocios se hacen a juicio pero se pueden hacer con técnicas estadísticas elementales con una ventaja: frialdad.

3) *Evaluación del desempeño*. Hay una gran variedad de medidas "intermedias" del desempeño:

- ventas, calidad del producto, satisfacción del cliente, costos, trabajo en proceso, inventarios, flujo de efectivo, rotación del personal.

Todas ellas no son definitivas pero sí indicativas del desempeño. No hay que optimizarlas "per se", pero indican el desempeño. La mayor parte de las veces se usan para establecer comparaciones.

4) *Causa-efecto*. Las relaciones de causa-efecto o por mejor decirlo las asociaciones, se obtienen de dos tipos de estudio.

- estudios observacionales (pasivos).
- estudios experimentales (activos).

A veces un buen análisis de datos observacionales sirve, pero no es lo usual. Los experimentos requieren una cuidadosa planeación (diseño) Se pueden usar no sólo en producción y diseño del producto sino también en procesos administrativos.

5) *Datos discordantes y chiripa*. Algún dato inconsistente con los otros puede, por acaso, dar la clave de algo importante. Hay técnicas estadísticas que señalan las aberraciones para analizarlas con mayor cuidado.

6) *Muestreo*. Para tomar sólo las observaciones necesarias y no gastar recursos en observaciones innecesarias.

## ¿En cuál empresa?

Estadística y cultura organizacional. Si no hay "ambiente receptivo", aún una estadística bien hecha es inútil. La idea central para el uso de la estadística en los negocios es la de proceso y su mejora. Se deben considerar las actividades de la empresa como un proceso y la mejora del proceso tendrá como señales distintivas: mediciones *empíricas* y juicios con base a *datos*.

Respecto al estilo de pensamiento estadístico hay varios puntos sobresalientes:

- selección de una muestra para inferir.
- localización de una muestra y medición de la dispersión.
- reconocimiento de la variabilidad y de la distribución de las mediciones que hagamos.

- probabilidad condicional y conocimiento imperfecto.
- "errores" de medición en los datos.
- variabilidad y qué hacer ante ella o cómo actuar ante situaciones variables
  - variabilidad normal
  - variabilidad especial.

Hay dos tipos de causas de variabilidad: causas comunes y causas especiales. De unas se encargan los capataces pero de las otras quien se debe encargar es la administración.

## ABUSOS DE LA ESTADÍSTICA

Estadística mágica. Amuletos usuales:

- Si el histograma es normal, la cosa está bien.
- $R^2$  significativa (o algún otro número mágico).
- No prestar atención a las hipótesis.
- Parámetros y no pronóstico de observaciones futuras.
- No tener una definición operacional de las cantidades medidas y ser descuidado a la hora de hacer mediciones.

DEFINICIONES:

- Hay tres clases de mentiras: la mentira, la maldita mentira y las estadísticas - (Mark Twain)
- La estadística es una ciencia que demuestra que si mi vecino tiene dos coches y yo ninguno, los dos tenemos uno - (George Bernard Shaw)
- La estadística es una ciencia según la cual todas las mentiras se tornan cuadros (Pitigrilli)
- La estadística es la primera de las ciencias inexactas - (Edmond Goucourt)
- Democracia: es una superstición muy difundida, un abuso de la estadística (Jorge Luís Borges).

### Estadísticas

(Publicado en la columna Milenio del [Diario de Noticias](#) el lunes 28 de octubre de 2002)

La frase ha sido atribuida a muchas personas diferentes, pero parece proceder de Benjamin Disraeli: "there are three kind of lies - lies, damned lies and statistics". O sea, que hay tres tipos de mentiras: mentiras, mentiras enormes y estadísticas. He traducido así a mi aire, porque normalmente he visto la cita incorporando el término "mentiras piadosas y mentiras de mala fe", o "malditas mentiras": posiblemente el adjetivo se refiere más a lo extremo de algunas mentiras. Claro que, por encima de todo, según Disraeli, estaban las estadísticas. Siempre he pensado que, si bien el adagio se entiende (interpretándolo en el sentido del uso que alguien puede hacer de una estadística, no por la técnica matemática en sí), es ciertamente injusto. Quizá porque la gente de la calle no comprendemos las estadísticas, y así, nos da la sensación de que cualquiera nos puede engañar con ellas.

Lo cierto es que somos normalmente nosotros los que nos engañamos con la estadística. Sobre todo cuando hablamos de correlaciones, es decir, cuando ponemos en relación estadística dos sucesos. Casi siempre hay alguien, o casi todos, que piensan que de ahí se deriva que uno es causa del otro, cuando a menudo tal relación causal no existe. Por ejemplo, cuando el gobierno habla de que la delincuencia la producen en un porcentaje importante inmigrantes ilegales, la gente lee una relación causal: los inmigrantes vienen a delinquir. ¿La hay realmente? En absoluto, aunque hay causas subyacentes a los dos fenómenos que explican esa correlación.

Últimamente muchas políticas pretenden basarse en estadísticas, estableciendo causas y efectos allí donde sólo existen correlaciones. Quizá sería bueno que algunos tomaran unas clases de estadística, de matemáticas, y dejaran de hacer demagogia.

### **Mentiras, pecados y abusos estadísticos**

**Por Bartolo Luque Serrano.-** *No es de extrañar que, con las manipulaciones y los malos usos estadísticos, el público en general acabe navegando entre la fascinación y la repudia por las cifras. "Existen medias mentiras, mentiras y estadísticas", oímos con frecuencia decir satisfecho al tertuliano de turno. La frase correcta debería ser: "Existen medios mentirosos, mentirosos y estadísticos embaucadores".*

La estadística no es ni torpe ni taimada. Lo son aquellos que la usan ignorando sus supuestos más sencillos o con anhelo de que sirvan a sus intereses particulares. Como apunta el estadístico Stephen K. Campbell, esta continua perversión de la estadística hace que el ciudadano "en un principio acepte las conclusiones estadísticas sin ejercer crítica alguna, por suponer que las cifras no mienten. A veces nos desalentamos con el solo hecho de que se nos ofrezcan afirmaciones que empiecen "según las estadísticas..." o "las estadísticas demuestran que..." Pero conforme maduramos pasamos al extremo opuesto. Ya nos han engañado demasiadas veces publicistas, políticos, ciudadanos prominentes que tratan de darnos gato por liebre, periodistas que buscan el sensacionalismo, etc. Por lo que tendemos a creer que con las estadísticas se puede probar cualquier cosa, y por lo tanto no prueban nada. Mientras que en un momento creímos que las cifras no podían mentir, ahora se deduce que lo único que pueden hacer es engañar".

Muchos malos usos estadísticos son pequeños pecados. Por ejemplo, cuando un medio de comunicación quiere impresionar a su audiencia con la gravedad de una situación, que puede afectar a gran parte de la población, suele emplear números absolutos en vez de porcentajes. Así leemos en un titular: "50 muertos en el puente de cuatro días." Si la noticia se acompañara de las estadísticas de muertes por accidente de tráfico, observaríamos que se trata de aproximadamente el mismo valor que el número de víctimas por accidente de tráfico en cualquier periodo de cuatro días. Pero entonces el titular no sería noticia... Se trata de lo que los matemáticos llaman falacia de "base extensa". La argucia se suele emplear en forma inversa. Por ejemplo: "El número de asesinatos en la ciudad tal aumentan un 60% respecto al pasado año." Si el periodista o el político nos dijeran que el año pasado se cometieron 5 homicidios y éste 8, probablemente el dato no nos impactaría de la misma manera. Y sin embargo, en ninguno de los dos ejemplos podemos decir que el periodista o el político hayan mentido.

Pero en otras cuestiones, el pecado puede ser mayor. Para determinar el número de pobres de un país, la tasa de desempleo o el índice de precios al consumo (IPC), se debe acordar una definición de pobreza, desempleo o consumo. El término apalancamiento estadístico hace referencia al grado en que puede cambiar el conteo en cuanto se altera un poco la definición de aquello que estamos contando. Nuestro ejemplo más cercano es la disociación entre el IPC presentado por el gobierno y la sensibilidad popular respecto al impacto en los precios con la entrada del euro.

Una redefinición del IPC puede cambiar las cifras apalancándolas a la baja. No nos ha de extrañar que varias instituciones de prestigio puedan ofrecer cifras verdaderamente dispares sobre la misma cuestión. Todo depende de la definición y métodos de conteo

que cada institución haya empleado. Sin esas definiciones y métodos presentados de forma explícita, los datos deberían levantar nuestra suspicacia. Sobre todo teniendo en cuenta que, la mayoría de las veces, los resultados están asombrosamente adecuados a la ideología o intereses particulares de las instituciones.

A propósito de la polémica por la disparidad de cifras ofrecidas por el gobierno y los sindicatos, en referencia al seguimiento de la última huelga general, ¿recuerdan si gobierno, sindicatos o algún medio de comunicación dio explicaciones de los métodos por los que se determinaron las cifras? Tal vez, si el ciudadano hubiera sido informado en estos términos, su juicio hubiera sido más claro.

El quid de la estadística consiste en deducir opiniones, propiedades, cualidades, etc. del conjunto de una población, a partir de informaciones de una muestra. Estimar el tanto por ciento de población que prefiere a cierto candidato político, parece, en principio, simple. Se encuesta a una muestra de la población seleccionada al azar y se determina qué porcentaje de la muestra prefiere al candidato. De ahí hemos de estimar qué porcentaje de toda la población prefiere al candidato. Lamentablemente, no es tan fácil: tanto el tamaño de la muestra, como la forma de seleccionarla, pueden conducirnos a engaños. Es evidente que el tamaño de la muestra es determinante a la hora de sacar conclusiones. Cuanto mayor sea la muestra, mejor serán nuestros datos. De hecho, lo ideal sería conocer la opinión de toda la población...

### **Intervalos de confianza**

Para controlar el efecto del tamaño de la muestra los estadísticos usan intervalos de confianza y niveles de significación. Una frase estadísticamente más correcta que: "El 66% de los ciudadanos prefieren al candidato X", sería: "Con un nivel de significación del 95%, el 66% más/menos el 6% de los ciudadanos prefirieron al candidato X". Quiere decir que con una seguridad del 95%, entre el 60% y el 72% de la población (el intervalo de confianza) prefiere al candidato X.

Lamentablemente si queremos afinar la estimación, reducir el intervalo de confianza, perdemos seguridad, disminuimos el nivel de significación. La única manera de aumentarla es aumentando el tamaño de la muestra encuestada, pero eso es caro. Como nos comenta el profesor de estadística de la Universidad Politécnica de Madrid, José Olarrea: "Los periódicos no se atreven a publicar los datos estadísticos de forma más correcta. La formación estadística de los periodistas es nula. Probablemente piensen que si ellos no lo entienden, exponerlos sería apabullar al lector. Pero en el fondo le hacen un flaco favor. Por ejemplo, es común que una compañía contrate los servicios de varias empresas de estadística para analizar su posición en el mercado respecto a sus competidores. Una vez la compañía dispone de los resultados de varias encuestas, puede escoger aquella que más apoye sus intereses jugando con los intervalos de confianza y los niveles de significación. Puesto que finalmente sólo nos ofrecen un porcentaje sin ninguna referencia más, probablemente no podemos decir que mientan, pero el dato así presentado carece de toda credibilidad".

Si una noticia nos dice que el desempleo ha cambiado del 9,1 al 8,9% de la población activa, parece claro que se trata de una buena noticia. Pero si resulta que el intervalo de confianza es de más/menos 1%, el error asociado al muestreo nos está diciendo que esa aparente disminución puede ser inexistente. Es más: ¿es posible que el paro haya aumentado!

La segunda cuestión importante, después del tamaño, a la hora de realizar estadísticas fiables es escoger una muestra representativa de la población. Para que así fuera, deberíamos escoger al azar. Pero eso es mucho más difícil de lo que parece. Verbigracia: si la encuesta es telefónica, dejaremos fuera a aquellos que no tienen teléfono. Si pedimos a los lectores de una publicación, ya sea en papel o electrónica, que den su opinión, estaremos construyendo una muestra autoseleccionada. Sólo los lectores de esa publicación, y entre ellos los verdaderamente comprometidos con la cuestión, perderán su tiempo. ¿Qué validez tiene entonces una información tan sesgada?

El bis cómico de la selección la sufrimos diariamente en los programas televisivos. Cuando el presentador dice: "escuchemos la opinión del hombre de la calle", significa que el reportero ha molestado a los viandantes hasta que puede obsequiarnos con una declaración a favor y otra en contra. Pero el premio se lo llevan los anuncios publicitarios donde un famoso nos aconseja tal o cual producto. ¡En ese caso la estadística se reduce a un solo individuo y desde luego no escogido al azar!

### **Correlaciones espúreas**

Si obtener las cifras estadísticas ya es un problema, inferir de ellas conclusiones parece haberse convertido en el todo vale. Para demostrar que un hecho es causa de otro, no basta con establecer que tienden a producirse de forma simultánea. Si razonamos de forma tan cándida, seremos víctimas de lo que los estadísticos llaman correlaciones espúreas. Por ejemplo: una serie de tests sobre 1.000 niños demuestra sin ambages que los que poseen los pies más grandes, son capaces de sumar mejor. Estamos correlacionando la habilidad para sumar con el tamaño del pie. ¡Y la correlación es positiva! ¿Cómo es posible? Sencillamente los niños de mayor edad suman mejor que los más pequeños. Y por supuesto tienen los pies más grandes... Es una correlación espúrea.

Sin una teoría del por qué, que sustente una correlación, debemos levantar las defensas. O podemos acabar estableciendo conclusiones absurdas, a partir de hechos estadísticamente fiables. Todas las estadísticas nos dicen que la mayor parte de los accidentes de tráfico se producen a menos de 40 Km. del lugar de residencia y por debajo de los 80 Km. por hora. De aquí podríamos deducir que para conducir, es conveniente hacerlo lo más alejado posible de nuestra casa y a grandes velocidades. Un absurdo.

Somos víctimas del abuso estadístico por nuestro analfabetismo numérico, por nuestro "anumerismo", como gusta decir al lógico Douglas Hofstadter. La brecha entre las dos culturas, las humanidades y las ciencias, nos ha conducido a situaciones tan paradójicas como que el ciudadano medio se jacte de su analfabetismo matemático: "Yo de números no entiendo..." y sin embargo le parezca signo de profunda incultura el desconocimiento de quién era Cervantes. Ambas negligencias son incultura. Y la ignorancia siempre ha sido la puerta de entrada para las injusticias, los atropellos y el engaño.

Como escribe el matemático A. K. Dewdney: "Aquellos que abusan de las matemáticas también abusan de nosotros. Nos convertimos en presas de las triquiñuelas comerciales, las estafas financieras, la charlatanería médica y el terrorismo numérico de los grupos de presión, todo porque somos incapaces (o no estamos dispuestos) a pensar con claridad durante unos momentos."