

## INTRODUCCIÓN

---

Comenzamos el curso de **Estadística Técnica** de la *Facultad de Ingeniería* de la *Universidad Nacional de Cuyo*. Nuestra asignatura comprende el estudio de la Probabilidad y de la Estadística. Por esto es necesario que sepamos qué es la Probabilidad y qué es la Estadística, además de conocer una serie de conceptos imprescindibles para entender la temática a desarrollar en los distintos ejes.

### ¿Qué es la Probabilidad?

---

Etimológicamente, la palabra *Probabilidad* procede de la palabra latina *probabilitas*, *-atis*, que significa verosimilitud o fundada apariencia de verdad.

“Lanzando al aire una moneda una sola vez nadie puede saber si caerá cara o cruz. Lanzando dos toneladas de monedas, todo el mundo sabe que es de esperar que una tonelada caiga en cara y la otra en cruz”.

El matemático que efectuó esta observación ilustraba las sorprendentes leyes que gobiernan los acontecimientos del azar.

**Los aspectos matemáticos de las leyes del azar, aquellas regularidades y fluctuaciones que se derivan de los acontecimientos que implican inseguridad y riesgo, constituyen la Teoría de las Probabilidades.**

Históricamente, la Teoría de las Probabilidades nació del interés por los juegos de azar.

- El Chevalier de Méré estaba interesado en los acontecimientos relativos a un juego de azar en boga en el siglo XVII, pero como creía que sus cálculos estaban errados, consultó a Blas Pascal, cuyo genio abarcaba la teología, la matemática y las ciencias. Mientras trabajaba en el problema planteado por de Méré, Pascal encontró otros desafiantes problemas de probabilidades.
- Estas cuestiones dieron origen a una provechosa correspondencia entre Pascal y Fermat, otro gran matemático. Entre los dos elaboraron gran parte de la Teoría de las Probabilidades.

## ¿Qué es la Estadística?

Etimológicamente, no hay un acuerdo general acerca del origen de la palabra *Estadística*. Una de las teorías al respecto acepta que procede de la palabra latina *status*, en su significación de país o nación; así la emplearon Shakespeare en 1602 y Achenwald en 1748, pero atendiendo a su connotación actual, bien podría considerársela con el matiz de situación o modo de ser de una persona o cosa (estado físico, estado natural, estado del tiempo, estado de salud, estado económico o financiero, etcétera).

Y hay quienes la hacen derivar de la palabra latina *statere* que, si bien en sentido estricto significa balanza: instrumento para medir el peso de las cosas, por traslación semántica, acabó señalando la acción realizada: medición.

Como dijera David Huntsberger: "La Estadística a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y así sucesivamente". Y tiene razón, pues al escuchar esta palabra, éstas son las imágenes que llegan a nuestra cabeza, pero la Estadística es mucho más que sólo números apilados y gráficas bonitas. Es una ciencia con tanta antigüedad como la escritura, y es por sí misma auxiliar de otras ciencias. Los mercados, la medicina, la ingeniería, los gobiernos, etcétera, se nombran entre los más destacados clientes de ésta. Su ausencia conllevaría muchos problemas, dejando a muchos profesionales sin información vital a la hora de tomar decisiones en condiciones de incertidumbre...

Algunos autores<sup>1</sup> dicen que la **Estadística es la ciencia de los datos**. Esto implica la *colección, clasificación, síntesis, organización, análisis e interpretación* de los datos.

Otros autores<sup>2</sup> definen a la Estadística como un **método de toma de decisiones frente a la incertidumbre**.

**La Estadística es, pues, el método y la forma para que, a partir del análisis de datos, se pueda tomar decisiones en condiciones de incertidumbre.**

La manera en que se trabajan los datos y el objetivo que se persigue en su análisis da lugar a distintas ramas de la Estadística.

- La rama que se dedica a la *organización, síntesis y descripción* de conjuntos de datos es la **Estadística Descriptiva**.
- La rama que se dedica a utilizar datos de muestra para *inferir* algo acerca de una población se denomina **Estadística Inferencial**.

<sup>1</sup> Mendenhall, W. y Sincich, T. (1997). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: Prentice Hall, p.2

<sup>2</sup> Chou, Y. (1990). *Análisis Estadístico*. México: Mc Graw-Hill, p.1

## La evolución histórica de la Estadística

### ▪ La Estadística como recuento de personas y bienes

La Estadística, en su sentido primario y fundamental, es simplemente el recuento de personas y de bienes; es decir, el censo y el inventario de un conglomerado humano, la recolección, ordenación y recuento de la gente y sus pertenencias. Con esta significación elemental, que hoy constituye apenas un aspecto, aparece ya desde la más remota antigüedad:

#### o *La Estadística en el antiguo Egipto*

Herodoto (padre de la Historia según Cicerón), cuyas afirmaciones son unánimemente respetadas, cuenta que en el antiguo Egipto, uno de los trabajos preparatorios para la construcción de las pirámides fue la realización de un censo con el fin de conocer la población, y la confección de un inventario de la riqueza total del país. Se sabe que una de estas construcciones fundamentales, la de Keops, tiene una base cuadrada de 232 metros de lado por 146 metros de altura. La construcción consumió el trabajo de 300.000 hombres durante 30 años. El proyecto exigía una ingente cantidad de obreros y de bienes necesarios. De ahí la necesidad del recuento que, según cálculos posteriores al historiador griego, se habría efectuado en el año 3050 a.C., es decir, hace más de 5.000 años.

Éste sería el primer trabajo estadístico que registra la historia; aunque su magnitud hace suponer que no fue el primero, en efecto, en varios monumentos pertenecientes también al antiguo Egipto, se han hallado inscripciones que los egiptólogos interpretaron como listas estadísticas de terrenos o fincas próximas al Nilo que se inundaban y eran transitoriamente abandonados por sus moradores durante los desbordes periódicos del gran río, y que luego de pasada la inundación debían ser restituidos a sus antiguos pobladores. Estos registros son también ejemplo de catastros o censos realizados en distintas épocas, anteriores y posteriores a la construcción de las pirámides.

#### o *En la antigua China*

La segunda importante referencia nos viene de la antigua China. Confucio (el moralista, filósofo, político y reformador chino) informa que el emperador Tiao mandó realizar un censo general de la agricultura y de las transacciones comerciales de su vasto imperio. Como este emperador vivió a mediados del siglo XXII a.C. (alrededor del año 2238 a.C.) el censo se habría realizado hace más de 4200 años.

#### o *En el antiguo Israel*

También se conservan las referencias bíblicas: El libro de los Libros, en el llamado de los Números, se relata el recuento que hizo Moisés de los hebreos que cruzaron el desierto con él. Como Moisés, el hombre de las Tablas de la Ley vivió en el siglo XIII a.C. (nació en 1305 a.C.) este recuento se realizó hace más de 3200 años. En tanto que en el Libro de David puede leerse el recuento que hizo de los hebreos de las distintas tribus en edad y condiciones de portar armas para convertirse en guerreros. Como este rey, que mató con su honda, de una pedrada, al gigante Goliat en 1038 a.C. (1055 a.C. - 1014 a.C.), vivió en el siglo XI a.C., el recuento se hizo hace unos 3000 años.

o *En Asiria y Babilonia*

En la biblioteca de Asurbanipal se hallaron documentos oficiales referentes a la producción obtenida en diversas provincias asirias durante el siglo VII a.C., recuentos efectuados hace unos 2.700 años.

o *En Grecia*

Platón en sus "Diálogos" pone en boca de Sócrates (470 a.C. - 400 a.C.) la importancia que los censos tenían para los gobernantes, pues les permitían conocer el número y los bienes de sus gobernados.

o *En Roma*

Se realizaban censos con cierta frecuencia, primero en Grecia y después en Roma, especialmente para saber a quiénes debían cobrar los impuestos o tributos y el monto correspondiente.

o *En la Edad Media y el Renacimiento*

En el año 1000, Guillermo, el Conquistador, al invadir Inglaterra ordenó realizar un censo administrativo de notable importancia. En 1550, el Concilio de Trento impuso la inscripción obligatoria en los libros de las Iglesias, de los matrimonios, nacimientos y defunciones, acumulando un importante material para diversos estudios estadísticos futuros.

▪ **La Estadística como recuento de los bienes del estado**

En el siglo XVII la técnica censal adquirió un gran desarrollo, llegando a constituirse en un eficaz auxiliar de las tareas de gobierno, especialmente en Alemania donde German Könnig (1600 - 1681) fue el sistematizador de los recuentos de los bienes del Estado.

Durante el siglo siguiente la técnica de Könnig se enseñaba en las universidades alemanas, pero fue Godofredo Achenwald (1719 - 1772) quien introduce la palabra estadística dándole el significado de cosas o bienes del estado.

No parece, sin embargo, que este vocablo haya sido creado por él, puesto que un siglo y medio antes, en 1602, William Shakespeare, en su Hamlet, ya

lo había usado con la misma acepción. Lo que hace suponer que la palabra fue empleada antes en la literatura que en las ciencias.

- **La Estadística como medición de hechos aleatorios e inducción de parámetros universales**

A partir del siglo XVII estos recuentos comienzan a realizarse sobre hechos que ya nada tienen que ver con las cosas y los bienes del estado y sí con la medición de hechos sociales, demográficos, económicos, etcétera. Al mismo tiempo su base matemática, hasta entonces meramente aritmética, se amplía y se hace algebraica. Es así como se constituye en un método adecuado e irremplazable para medir sucesos aleatorios a través de variables y parámetros del universo al que pertenecen los hechos observados.

- **Los Aritméticos Sociales**

A mediados del siglo XVII comenzó a conocerse en Inglaterra un grupo de investigadores demográficos que no se conformaron con las escuetas cifras de los censos, sino que intentaron sacar de ellos conclusiones generales.

Estos estudiosos fueron llamados *aritméticos sociales* o *políticos* porque aplicaban la aritmética a tales cuestiones. El más destacado fue John Graunt (1620 - 1674) quien hacia 1650 presentó a la Real Sociedad de Londres unas "Memorias" novedosas e interesantes. En una de ellas relaciona el número de nacimientos en un determinado lapso con el número de habitantes, y el número de defunciones con el de habitantes. Fue el primero en establecer índices de natalidad y mortalidad, así como otros índices y porcentajes. En otra memoria compara el número de varones y mujeres en distintos años y arriba a la conclusión de que en Londres, la relación de los sexos entre adultos se mantiene invariable. Investigó también el número de personas que afluyen del campo a la ciudad en diferentes épocas, y demostró que esta afluencia era constante. Anotaba las estaciones en las que se producen mayor número de defunciones para establecer la influencia de aquéllas sobre la mortalidad, etcétera.

Graunt fue el primero en no conformarse con los simples datos de los censos, por el contrario, los hacía hablar extrayéndoles conclusiones. Por su trabajo se lo considera un precursor de la Estadística moderna.

- **La base matemática de la Estadística Inductiva**

Nace a mediados del siglo XVII con los estudios de los grandes matemáticos de la época, en especial, los italianos, franceses y suizos.

Es interesante señalar que, al parecer, las primeras sugerencias se debieron a personas dedicadas a los juegos de azar que nada tenían que ver con estu-

dios matemáticos. Se cuenta que un jugador, conversando con Galileo Galilei (1563 - 1642) le comentaba perplejo un hecho conocido de todos los jugadores de dados: que a pesar de la probabilidad de salir que tiene un número cualquiera de cada uno de los dados, tirando tres al mismo tiempo la frecuencia con que aparecen determinadas sumas es mayor que otras. Así, las sumas 10 y 11 aparecen con más frecuencia que las de 3 y 18. Galileo habría comprobado el hecho y explicado la diferencia que hay entre frecuencias y probabilidad, nociones que aún estaban muy confusas.

Pero el verdadero punto de partida de esta estadística algebraica lo hallamos en los estudios del matemático inglés Isaac Newton (1643 - 1727), quien desarrolló y descubrió lo que hoy llamamos binomio de Newton. Su desarrollo fue tomado por otros matemáticos y aplicado al cálculo estadístico. El estudio de este binomio ya había sido objeto de preocupación en otro matemático veinte años más grande que Newton: el francés Blas Pascal (1623 - 1663).

Sin embargo, el matemático de este siglo que más influyó en la evolución de la Estadística fue el suizo Santiago Bernoulli (1654 - 1705), quien dio expresión matemática a la Ley de los grandes números conocida como Teorema de Bernoulli, que más tarde sería generalizada por Poisson.

En el siglo XVIII, el matemático francés Abraham De Moivre (1667 - 1754), refugiado en Inglaterra por razones religiosas, descubre en 1733 la ecuación de la curva simétrica, llamada después curva de Gauss.

Desde la segunda mitad del siglo XVIII y la primera del XIX, el francés Pedro Simon Laplace (1749 - 1827) y el alemán Carlos Federico Gauss (1777 - 1855) descubren que las frecuencias ordenadas de las muestras normales tienen una distribución simétrica que obedecen a una ley llamada, ahora, Ley de las frecuencias normales de Laplace y Gauss, cuya representación gráfica es una curva simétrica y acampanada, hoy llamada curva normal de Gauss.

En 1837, Simeón Dionisio Poisson logra la generalización del teorema de Bernoulli a la que denomina Ley matemática de los grandes números, y estudia las muestras dicotómicas numerosas con medias muy pequeñas, que con justicia llevan su nombre.

En el siglo XIX el astrónomo y sociólogo belga Lambert Quetelet (1796 - 1874) aplica la estadística a las ciencias sociales y a la biología, especialmente a la antropología; mientras que el inglés Francis Galton (1822 - 1911) utiliza métodos estadísticos para el estudio de los caracteres hereditarios, somáticos y psíquicos.

En 1899, el inglés Carlos Pearson encuentra la forma de resolver problemas estadísticos planteados por muestras cuyas frecuencias obedecen a una ley,

hipótesis, razón, proporción, etcétera, y establece un parámetro al que llama  $\chi^2$  (Chi-cuadrado), más una técnica de extraordinaria importancia.

Pearson y Galton establecen la forma de precisar la existencia de una asociación estadística entre dos conjuntos o muestras de observaciones simultáneas, es decir, inician el estudio de las técnicas de correlación y regresión estadística.

En 1908, Gosset, conocido con el pseudónimo Student, publicó su técnica para tratar las muestras normales poco numerosas que escapan a la técnica de Gauss.

En nuestro siglo el número de estudiosos y sus aportes es tan extraordinariamente enorme e importante que sólo vamos a mencionar algunos nombres: Fisher, Snedecor, Yates, Sheppard, Charlier, Bessel, Tchebychev, Cauchy, Mc Nemar, Tukey, May, Keuls, entre otros.

#### ▪ La Estadística en la predicción de hechos futuros

Finalmente la Estadística no sólo mide, cuenta, analiza los resultados, generaliza y mide los parámetros del universo al que pertenece la muestra, sino que se proyecta en el futuro haciendo predicciones matemáticas válidas (Por ejemplo, de seguir el ritmo de crecimiento de la población humana mundial, que en 1975 ya era de 4.000.000.000, en el año 2000 sería de 8.000.000.000 de habitantes según el Anuario Demográfico de las Naciones Unidas, aparecido en 1976).

De todo lo dicho se desprende que la evolución de la Estadística puede dividirse en tres etapas :

- ✓ Una primera, clásica o *descriptiva*, que va desde su iniciación hasta mediados del siglo XVII: el simple recuento, inventario, censo o catastro aplicados a la población, a los recursos naturales, al comercio y luego a la industria, al tráfico y a toda otra manifestación de las actividades de un estado. Para ella las conclusiones no trascienden de la mera suma de los datos.
- ✓ Una segunda o *analítica*, que empieza estableciendo índices y porcentajes para luego constituirse en el único procedimiento idóneo destinado a medir los hechos o sucesos aleatorios, es decir variables por obra del azar; así como el valor de los parámetros universales. En ella la Estadística se ocupa del análisis y de la generalización de los datos procedentes de la observación, analizando conjuntos parciales llamados muestras, e induce el conocimiento del conjunto total, lo compara con otros también totales inducidos o hipotéticos. La tipificación de estos análisis lleva a la sistematización de las técnicas que constituyen los métodos estadísticos. Los hechos que estudia son siempre contingentes y aleatorios, por lo que no es posible predecir con

seguridad su presencia o magnitud, aunque logra establecer su probabilidad. En esta etapa trasciende la mera suma de los datos y abstrae conceptos que le permite generalizaciones.

- ✓ Y una última etapa, en la que la Estadística se hace *predictiva*, haciendo inferencias estadísticas que le permite adelantarse al futuro estableciendo pronósticos matemáticamente válidos. Este gran poder de la Estadística hizo que sus aplicaciones se hayan extendido considerablemente, y en la actualidad no hay rama del conocimiento ni de la actividad humana, en la que no se la utilice. Así la encontramos en las ciencias exactas: física, química, astronomía, mecánica, constituyendo capítulos enteros como el de la física cuántica, o el de la cinética de los gases y emplearse corrientemente en diversas ramas de la ingeniería: control de calidad de materiales, especificación, cibernética, cerebros electrónicos, servomecanismos, etcétera; en las ciencias biológicas, en biometría, tipología, paleontología, genética; en agronomía y veterinaria y en las distintas ramas de la medicina; en las ciencias económicas y financieras: ciclos de producción, pronósticos, etcétera y sus aplicaciones prácticas: el comercio, la industria, las finanzas; en las ciencias políticas: opinión pública; en las ciencias sociales, psicológicas, demográficas; y aún en áreas al parecer ajenas a la Matemática pero que se prestan al recuento de datos: la moral, la filosofía y las letras, en las que se la emplea para el estudio de las costumbres, corrientes filosóficas, estilos literarios, etcétera.

## Algunos conceptos imprescindibles...

Hemos mencionado a la *muestra*, los *datos de muestra* y la *población*, pero no los hemos definido:

**Población** es un conjunto de datos, objetivo de nuestro interés.

En cada caso particular, una investigación se enfocará hacia una colección bien definida de objetos, que forman una *población* de interés.

Por ejemplo, la población, podría consistir en todas las probetas de hormigón con ciertas características, producidas durante un período; otra investigación podría implicar a la población formada por todos los individuos que recibieron el título de ingenieros durante el año académico anterior.

Las poblaciones pueden ser **finitas** o **infinitas**, según la cantidad de elementos que la integren. Por ejemplo, una población *finita* estaría formada por los alumnos de la Facultad de Ingeniería que





están cursando Estadística Técnica en el presente año, mientras que una *infinita* estaría formada por todas las varillas de acero que fabricaría un industrial si continuara trabajando indefinidamente en determinadas condiciones de operación.

Cuando la información deseada está disponible para todos los objetos de la población, se tiene un **censo**. En general, por restricciones de tiempo, dinero y demás recursos limitados, resulta impráctico o imposible realizar un censo, en lugar de ello, se selecciona un subconjunto de la población: una *muestra*, especificada de cierta manera. Por ejemplo, podríamos obtener una muestra de rodamientos de determinada corrida de producción, como base, para investigar si los rodamientos se están apegando a las especificaciones de manufactura, o bien, se podría seleccionar una muestra de ingenieros graduados el año anterior para obtener retroalimentación sobre la calidad de los planes de estudio en ingeniería.

En general, la población está bien definida y es concreta, por ejemplo: *todos los conductores de automóviles equipados con sistemas de cinturones de hombro y en determinada área metropolitana*. Sin embargo, cuando se analizaron las causas del estallido del transbordador espacial Challenger, en 1986, la atención se dirigió al comportamiento de los anillos O del motor del cohete y, en particular, se estudió la temperatura del anillo O en cada ignición de prueba o de despegue real del motor del cohete, como el del transbordador. El análisis se hizo en base a una muestra de cien temperaturas de anillos O, pero proviene de una población que no existe en realidad; en lugar de esto, la población en este caso es una que está formada por todas las posibles mediciones que pudieran hacerse bajo condiciones experimentales similares. Tal población se conoce como **población conceptual** o **hipotética**.

**Muestra** es el subconjunto de datos seleccionados de una población.

Para que la información obtenida de los datos de una muestra nos permita hacer inferencias valederas, debe cumplir ciertas condiciones:

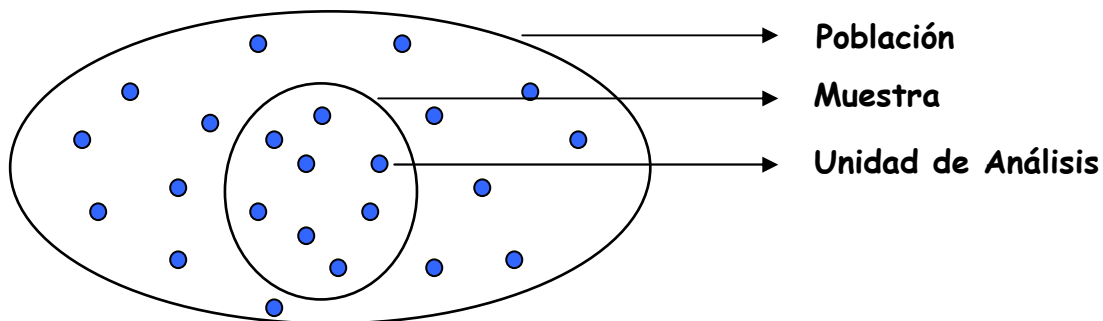
- **Representatividad:** La muestra debe revelar las características de la población de la cual proviene lo más aproximadamente posible. Debe ser una auténtica representación de toda la población, y por lo tanto no sirve cualquiera porción de la misma.

Para que sea representativa de una población, el tanto por ciento de individuos de la muestra que poseen una propiedad determinada debe ser el mismo que el tanto por ciento de individuos con esta propiedad en la población.

- **Aleatoriedad:** Cada elemento de la población debe tener la misma probabilidad de ser elegido. Sólo si satisface este requisito los métodos estadísticos serán razonables.

- **Independencia:** Esto equivale a decir que la probabilidad de que cualquier miembro de la población aparezca en la muestra, no depende de la aparición de los otros miembros de la población en la muestra.  
 Para que se cumpla la independencia, la población debe ser infinita. Cuando la población es finita, y se efectúa el muestreo aleatorio sin reemplazo, muestreo exhaustivo o sin reposición, no podemos hablar de que los elementos de la población sean independientes, en este caso decimos que hay dependencia.

**Unidad de análisis** es cada uno de los elementos seleccionados de la población y que integran la muestra.



**Experimento aleatorio** es el experimento que proporciona diferentes resultados aun cuando se repita siempre de la misma manera. Además, los resultados obtenidos deben depender sólo del azar.

En el ámbito de la ingeniería es interesante analizar las poblaciones, las muestras y los procesos<sup>3</sup>:

Los ingenieros están sometidos a conjuntos de hechos, o de **datos**, tanto en sus actividades profesionales como en las cotidianas. La Estadística les proporciona métodos para organizar y resumir datos, y para sacar conclusiones basadas en la información que contienen los datos.

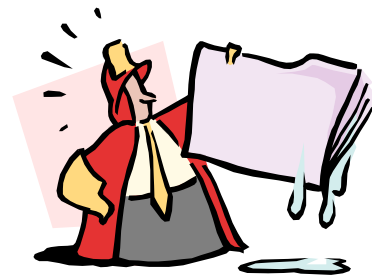


## Fuentes de datos

Existen cuatro fuentes de datos para la recopilación de datos:

<sup>3</sup> Devore, J. (2001). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: International Thomson Editores, S.A., pp.3-7

- Se pueden obtener datos ya publicados por fuentes gubernamentales, industriales, individuales.
- Se pueden diseñar experimentos para obtener los datos necesarios.
- Se puede realizar una encuesta.
- Se puede realizar un estudio observacional.



Las fuentes de datos se pueden clasificar de distintas maneras:

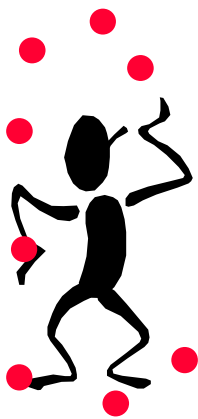
- Tipos de fuentes de datos según el *origen de la información*
  - o **Fuentes primarias**  
Recolección de información propia para un estudio. Es información que se obtiene expresamente para un estudio concreto.
    - **Internas**  
La información procede de la propia empresa.
    - **Externas**  
La información procede del mercado, de la sociedad, de entes gubernamentales, etcétera.
  - o **Fuentes secundarias**  
Revisión de fuentes de información ya existentes. Es información que se encuentra disponible antes de que se haga necesaria su obtención. Son estudios publicados por otros, o realizados por la empresa pero con otros objetivos.
    - **Internas**  
La información procede de la propia empresa.
    - **Externas**  
La información procede del mercado, de la sociedad, de entes gubernamentales, etcétera.
- Tipos de fuentes según la *técnica de obtención de información*
  - o **Técnicas Cualitativas**  
Se recomiendan especialmente en investigaciones exploratorias, en las primeras aproximaciones al problema.
  - o **Técnicas Cuantitativas**  
La información obtenida mediante técnicas cuantitativas puede ser cuantificada y medida de alguna manera.
  - o **Técnicas Mixtas**  
Siguen planteos cualitativos o cuantitativos, según el caso, o ambos conjuntamente.

## Tipos de datos y escalas de medición

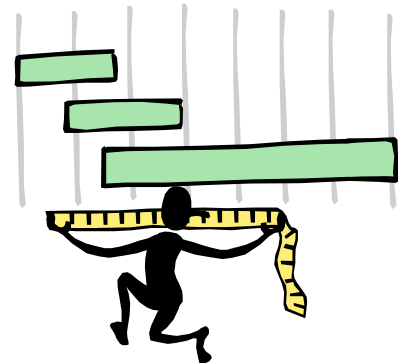
Generalmente, sólo nos interesan ciertas *características* de los objetos de una población: la cantidad de defectos en la superficie de un perfil, el espesor de pared de una cápsula, el género de un graduado de ingeniería, la edad a la que se graduó, etcétera. Estas *características* pueden ser: **categorías** o **cualitativas**, como el género (femenino, masculino) o tipo de defecto (como soldadura insuficiente), o bien pueden ser de naturaleza numérica, es decir, **cuantitativas**, como la edad (por ejemplo, 23 años) o el espesor (como 0,02 cm). Las características mencionadas son llamadas **variables**, es decir, son aquellas características cuyos valores pueden cambiar de un objeto al siguiente en la población. Cuando estos valores son los resultados de un experimento estadístico, la llamaremos *variable estadística*, y representa generalmente un cierto carácter de los individuos de una población.

Usualmente, las variables estadísticas se clasifican en *cualitativas* y *cuantitativas*, según que las modalidades del carácter que representan sean o no numéricas.

Dentro de las variables cuantitativas se distingue entre variables *discretas* y *continuas*, siendo discretas aquellas que, por su naturaleza, sólo pueden tomar valores aislados, generalmente números enteros, y continuas a las que pueden tomar todos los valores de un cierto intervalo.



Así, los experimentos que consisten en el recuento de objetos, como pueden ser: número de miembros de una familia, número de empleados de una empresa, etcétera, dan lugar a variables discretas, mientras que al medir magnitudes tales como el peso, el tiempo, capacidad, longitud, etcétera, se obtienen variables continuas.



Como resultado de nuestras medidas sobre individuos o unidades de análisis de la población bajo estudio, obtenemos un conjunto de datos, o resultados del experimento estadístico.

Pueden obtenerse muchas características diferentes para un mismo individuo. Estas características, dependiendo del tipo de valores que originan, pueden medirse con cuatro tipos distintos de *escalas de medición*.

- **Escala nominal**

La forma más simple de observación es la clasificación de individuos en clases que simplemente pueden distinguirse entre sí pero no compararse ni realizar entre ellas operaciones aritméticas. En este tipo se incluyen características tales como la profesión, nacionalidad, sexo, etcétera.

- **Escala ordinal**

A veces, las categorías obtenidas pueden ser ordenadas, aunque diferencias numéricas iguales, a lo largo de la escala numérica utilizada para medir dichas clases, no correspondan a incrementos iguales en la propiedad que se mide. Por ejemplo, puede asignarse un número de orden de nacimiento a un grupo de hermanos, sin que la diferencia de edad entre el primero y el segundo sea la misma que entre el segundo y el tercero. Características de este tipo son, las puntuaciones en una prueba de aptitud, categorización de una vivienda, etcétera.

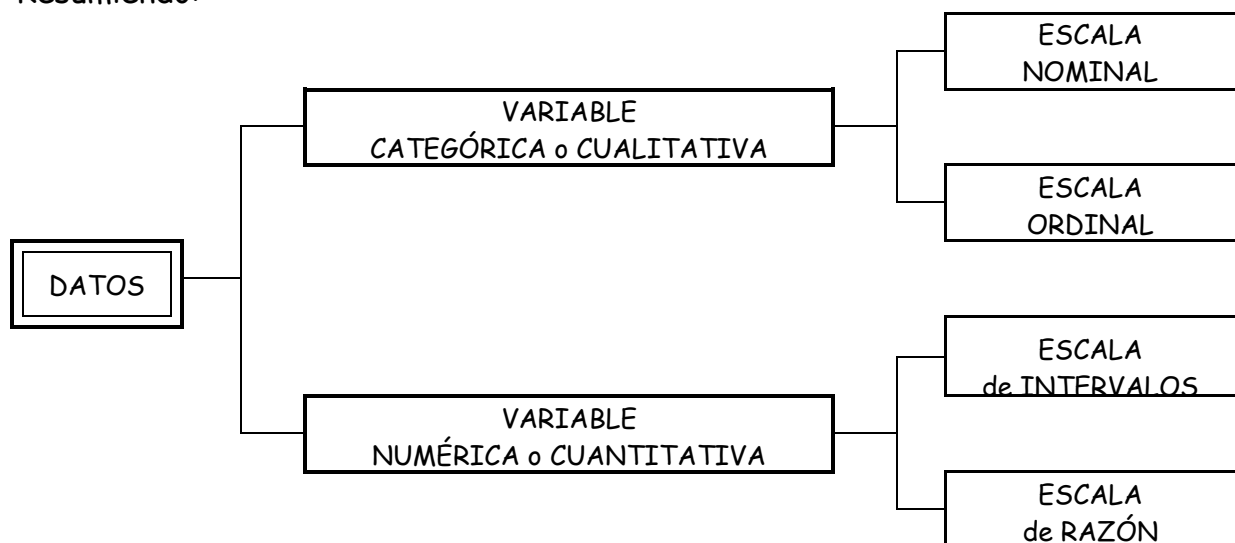
- **Escala de intervalo**

Esta escala, además de clasificar y ordenar a los individuos, cuantifica la diferencia entre dos clases, es decir, puede indicar cuánto más significa una categoría que otra. Para ello es necesario que se defina una unidad de medida y un origen, que es, por su naturaleza, arbitrario. Tal como ocurre con la temperatura y la escala cronológica.

- **Escala de razón**

Es idéntica a la anterior, pero además existe un cero absoluto. En el caso del tiempo, no podemos usar esta escala porque nuestro 'cero' no es absoluto, sólo denota un punto en la historia, el nacimiento de Cristo. Ejemplos de características medidas a nivel de razón son el volumen de ventas, costos de producción, edad, cotización de un cierto tipo de acciones, etcétera.

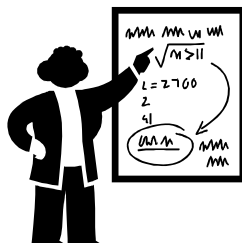
Resumiendo:



Una variable es **discreta** si sus valores forman un conjunto finito, o bien, se pueden listar en una sucesión infinita (infinita numerable). Una variable es **continua** si sus posibles valores forman un intervalo real. Para entender mejor estos conceptos analicemos algunos ejemplos:

- Una variable *discreta* cuyos valores forman un conjunto finito es "la cantidad de elementos defectuosos observados en una muestra de  $k$  elementos", y los valores que puede tomar serán,  $0, 1, 2, \dots, k$ .
- Una variable *discreta* cuyo conjunto de valores es infinito numerable es "la cantidad de veces que es necesario arrojar una pelota hasta encestar", y los valores que puede tomar serán,  $1, 2, 3, \dots$
- Una variable *continua* puede ser "la estatura de los estudiantes universitarios", y los valores que puede tomar pertenecen a un intervalo real como  $[100 \text{ cm}, 250 \text{ cm}]$ , otro ejemplo sería, "el tiempo de duración de un determinado tipo de lámpara" y los valores que puede tomar serán,  $[0 \text{ h}, 2000 \text{ h}]$

A partir de los datos obtenidos al relevar una variable en una muestra o en una población, podemos presentar estos datos de manera organizada y sintética, ya sea a través de gráficos, tablas o de valores numéricos que describan al conjunto de datos en cuestión, éstos últimos, son llamados *medidas descriptivas*.



Las medidas descriptivas que caracterizan a una población se llaman **parámetros**, mientras que las que se refieren a la muestra se llaman **estadísticos** o **estadísticas**.

## Relación entre Probabilidad y Estadística Inferencial

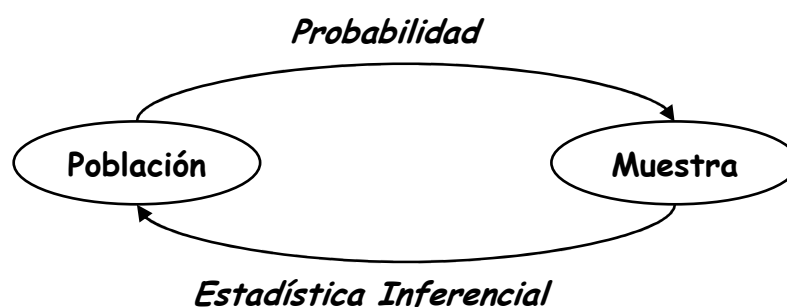
Pero nuestra temática no se refiere sólo al estudio estadístico, sino que también estudiaremos **Probabilidad**.

La Probabilidad<sup>4</sup> es un puente entre las técnicas descriptiva e inferencial de un estudio estadístico, permitiendo una mayor comprensión de cómo los procedimientos inferenciales se desarrollan y utilizan, cómo es que las conclusiones estadísticas pueden traducirse e interpretarse en el lenguaje diario, y cuándo y dónde pueden ocurrir riesgos ocultos al aplicar los métodos.

Tanto la Probabilidad como la Estadística se relacionan con cuestiones donde intervienen poblaciones y muestras, pero lo hacen de manera inversa.

<sup>4</sup> Devore, J. (2001). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: International Thomson Editores, S.A., pp. 6-7

En un problema de probabilidad, las propiedades de la población en estudio se suponen conocidas (es decir, en una población numérica puede suponerse alguna distribución específica de los valores de la población) y se formulan y responden preguntas en relación con una muestra tomada de la población. En un problema estadístico, las características de una muestra están disponibles para el experimentador, y esta información lo capacita para sacar conclusiones respecto a la población. La relación entre dos disciplinas se puede resumir al decir que la Probabilidad razona desde la población hacia la muestra (razonamiento deductivo), mientras que la Estadística razona desde la muestra hacia la población (razonamiento inductivo).



En la teoría de probabilidades deducimos la composición probable de la muestra a partir de la composición de la población original, pero la estadística se edifica sobre la teoría de las probabilidades, posibilitando la inversión de este razonamiento, es decir, *inferimos* la composición de la población original partiendo de una muestra convenientemente elegida.

El campo de la estadística concierne también a una variedad de otros temas, tales como el análisis y presentación de datos, diseño de experimentos y decisiones prácticas.

La estadística puede ayudar en el planeamiento económico y efectivo de experiencias. Muchas experiencias no satisfacen los propósitos originales, por falta de los debidos controles y debido a que su tamaño es inadecuado para la determinación de conclusiones de interés, con un nivel dado de seguridad.

Antes de que podamos comprender lo que nos dice una muestra en particular acerca de la población, primero debemos entender la incertidumbre asociada con el hecho de tomar una muestra de una población dada. Ésta es la razón por la que primero estudiamos Probabilidad y después Estadística Inferencial.

Como ejemplo de los distintos enfoques de la Probabilidad y la Estadística Inferencial, veamos el uso de los cinturones de seguridad, por parte de los conductores, en automóviles equipados con sistemas de cinturones de hombro (El artículo "Automobile Seat Belts: Usa-



ge Patterns in Automatic Belt Systems" en *Human Factors*, 1998, pp. 126-135, resume los datos de uso de estos cinturones).

En *Probabilidad* se podría suponer que el 50% de todos los conductores de autos con este equipo, de determinada área metropolitana, usan con regularidad su cinturón de hombro (es una hipótesis sobre la población), así que podríamos preguntar: "¿Qué tan probable es que en una muestra de cien conductores, al menos setenta de ellos usen su cinturón?", o bien, "en una muestra de cien conductores, ¿cuántos de ellos se espera que usen con regularidad su cinturón?".

Por otro lado, en la *Estadística Inferencial* tenemos disponible la información de la muestra; por ejemplo, "en una muestra de cien conductores de autos con este equipo, se vio que sesenta y cinco usan con regularidad su cinturón, entonces, ¿con esto tenemos evidencia suficiente para decir que más del 50% de todos los conductores en esta área usan con regularidad su cinturón?". En este último escenario tratamos de usar información de la muestra para contestar una pregunta acerca de la estructura de toda la población de la que se seleccionó la muestra.

## Aspectos éticos

Se dice que existen tres tipos de mentiras: *Las mentiras comunes, las mentiras piadosas y la estadística...* Este juicio tremendamente severo se hace realidad cuando la Estadística es utilizada por personas inescrupulosas que manipulan los datos y sacan conclusiones según sus deseos y necesidades. Darrel Huff en "Cómo mentir con la Estadística" dice: "...los bribones ya conocen tales trucos; los hombres honrados deben aprenderlos para defenderse."

Uno de los objetivos de este curso es revisar algunas de las formas más comunes en que se utiliza la Estadística de manera incorrecta, ya sea por falta de conocimiento o debido al intento de engañar al usuario. En cualquier caso, los que utilizan la Estadística y no saben cómo tratar tales prácticas fraudulentas, no pueden obtener mucha utilidad de esta disciplina.

Claro está que, desde nuestra posición, no aceptamos ninguna de las razones para el uso incorrecto y el abuso de las técnicas estadísticas. El ignorante hace tanto daño como el delincuente.

Todos estamos hartos de las mentiras descaradas que se escudan en un informe estadístico de vaya uno a saber qué datos y de quién sabe cómo fueron analizados.

Estamos abocados a prepararlos para hacer un uso responsable de la Estadística, además de prevenirlos para que no sean víctimas de ciertos estafadores de guante blanco que andan por ahí...



## Un momento para la recapitulación...

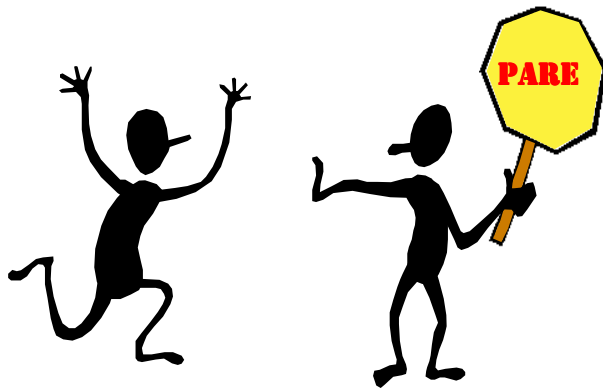
Sabemos que para ser sólo una introducción hemos dado muchas definiciones y conceptos, pero es preciso que los domine antes de comenzar el estudio de los distintos ejes, para que tenga fluidez en su aprendizaje.



Para autoevaluarse, responda las preguntas que están a continuación. Puede hacerlo con el material de estudio, pero asegurándose que "entiende" cada palabra, a tal punto que usted podría explicarle a un amigo, que no conoce el tema, de manera simple, los conceptos estudiados:



- De manera sencilla, cómo se define a la Probabilidad y a la Estadística.
- ¿Cómo se divide la Estadística, según el objetivo y la forma de estudio?
- ¿Cuál es el objetivo de la Estadística?
- ¿Qué es una población, en el marco de un estudio estadístico?
- ¿Qué es una población hipotética?
- ¿Qué es un censo?
- ¿Cuándo una población es infinita y cuándo no lo es?
- ¿Qué es una muestra? ¿Debe poseer características especiales?
- ¿A qué llamamos unidad de análisis?
- ¿A qué llamamos experimento aleatorio?
- ¿Qué relación existe entre la Probabilidad y la Estadística Inferencial?
- ¿Qué papel juegan los datos en un estudio estadístico?
- ¿Qué tipo de datos conoce?
- ¿Qué fuentes de datos conoce?
- ¿Cómo se clasifican las fuentes de datos?
- ¿A qué se llama variables?
- ¿Cómo puede clasificar a las variables y a las escalas de medición?
- ¿Podría dar ejemplos de los distintos tipos de variables y de las escalas de medición?
- Piense ejemplos de variables discretas y continuas. Además, piense qué valores podría tomar, es decir, defina el rango de la variable.
- ¿Qué son los parámetros? ¿Y las estadísticas?
- ¿Qué aspectos éticos son importantes para trabajar en Estadística?



**Por favor...  
¡No avance si tiene dudas  
o no recuerda  
las nociones aquí volcadas!**



Si está listo para continuar, realice las autoevaluaciones propuestas para la parte conceptual.

Una vez que las haya terminado y sienta confiado para continuar, realice las aplicaciones prácticas. Éstas le servirán para aplicar lo aprendido, destacar y reafirmar los conceptos.



Recuerde que en todas las instancias puede consultar sus dudas con el tutor.

**¡¡¡Por fin  
terminé la  
Introducción!!!**

