

ESTADÍSTICA:

Rama de las matemáticas que se ocupa de reunir, organizar y analizar datos numéricos y que ayuda a resolver problemas como el diseño de experimentos y de toma de decisiones.

HISTORIA

Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas. Hacia el año 3000 a.C. los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y de los géneros vendidos o cambiados mediante trueque. Los egipcios analizaban los datos de la población y la renta del país mucho antes de construir las pirámides en el siglo XXXI a.C. Los libros bíblicos de Números y Crónicas incluyen, en algunas partes, trabajos de estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. En China existían registros numéricos similares con anterioridad al año 2000 a.C. Los griegos clásicos realizaban censos cuya información se utilizaba hacia el 594 a.C. para cobrar impuestos.

El Imperio romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. Durante la edad media sólo se realizaron algunos censos exhaustivos en Europa. Los reyes carolingios Pipino el Breve y Carlomagno ordenaron hacer estudios minuciosos de las propiedades de la Iglesia en los años 758 y 762 respectivamente. Después de la conquista normanda de Inglaterra en 1066, el rey Guillermo I de Inglaterra encargó un censo. La información obtenida con este censo, llevado a cabo en 1086, se recoge en el Domesday Book. El registro de nacimientos y defunciones comenzó en Inglaterra a principios del siglo XVI, y en 1662 apareció el primer estudio estadístico notable de población, titulado *Observations on the London Bills of Mortality* (Comentarios sobre las partidas de defunción en Londres). Un estudio similar sobre la tasa de mortalidad en la ciudad de Breslau, en Alemania, realizado en 1691, fue utilizado por el astrónomo inglés Edmund Halley como base para la primera tabla de mortalidad. En el siglo XIX, con la generalización del método científico para estudiar todos los fenómenos de las ciencias naturales y sociales, los investigadores aceptaron la necesidad de reducir la información a valores numéricos para evitar la ambigüedad de las descripciones verbales.

En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El trabajo del experto estadístico no consiste ya sólo en reunir y tabular los datos, sino sobre todo en el proceso de interpretación de esa información. El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchos conjuntos de datos se pueden aproximar, con gran exactitud, utilizando determinadas distribuciones probabilísticas; los resultados de éstas se pueden utilizar para analizar datos estadísticos. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

OBJETO DE ESTADÍSTICA:

La estadística tiene por objeto el desarrollo de técnicas para el conocimiento numérico de un conjunto.

La estadística se divide en 2 ramas principales:

- Estadística descriptiva: Cuyo objetivo es examinar a todos los individuos de un conjunto.
- Estadística inferencial: Por la que, mediante el estudio de una muestra se sacan conclusiones válidas

para la totalidad.

POBLACIÓN Y MUESTRA: Se llama *población* al conjunto formado por todos los elementos cuyo conocimiento nos interesa. A cada uno de los elementos se lo llama individuo.

Muestra es un subconjunto limitado extraído de una población, con objeto de reducir el campo de experiencias. Las propiedades que obtengamos se harán extensivas a toda la población.

CARACTERES: Para el conocimiento de una población estadística, deberemos analizar a cada uno de sus individuos (o a cada individuo de una muestra). Pero ese análisis no puede ser exhaustivo; deberemos seleccionar uno o varios detalles (caracteres) y ver cómo se manifiesta ese carácter en cada uno de los individuos. Por ejemplo, de una población de mazorcas de maíz, los caracteres dignos de estudio pueden ser: color, número de granos, longitud, peso, etc.

Los caracteres pueden ser:

- Cualitativos: que se presentan bajo varias cualidades no medibles. Ej: color
- Cuantitativos: cuando son medibles. Ej: longitud de una mazorca, su peso, número de granos, etc.

Un carácter cuantitativo es **discreto** (o de variación discreta) cuando sólo puede tomar determinados valores. Es el caso del número de granos de una mazorca. Y se llama **continuo** (o de variación continua) cuando puede tomar valores tan próximos como se quiera (longitud, peso, etc.)

ESTADÍSTICA DESCRIPTIVA:

Esta rama de la estadística trata de describir y analizar un grupo dado sin sacar conclusiones para un grupo mayor. Trabaja en poblaciones, no en muestras propiamente dichas.

Para esa descripción y análisis de los individuos, que son el objeto de la estadística descriptiva, esta utiliza el siguiente método:

- **Selección de caracteres:** Dignos de estudio.
- **Análisis de cada carácter.** Este análisis consiste en:
 - examinar cada individuo y anotar el valor de cada carácter.
 - Establecer las clases de individuos que se desean distinguir respecto a ese carácter.
 - **Clasificar y contar** los individuos incluidos en cada clase
 - Calcular determinados valores numéricos (los **parámetros estadísticos**) a partir de los datos contenidos en las distribuciones anteriores.

Cuando se pretende dar a conocer a un gran público los resultados de estos análisis, se suelen utilizar **representaciones gráficas** de los datos tabulados en el apartado anterior.

ESTADÍSTICA INFERENCIAL:

Cuando hay problemas cuya resolución se nos torna:

- Imposible debido a lo inmenso del problema (millones de individuos).
- Posible, pero ruinoso.

Recurrimos a la estadística inferencial (diseño de experimentos, teoría de muestras, teoría de la estimación, etc...) para hallar respuesta de los interrogantes.

Los datos recogidos deben ser organizados, tabulados y presentados para que su análisis e interpretación sean rápidos y útiles. Por ejemplo, para estudiar e interpretar la distribución de las notas o calificaciones de un examen en una clase con 30 alumnos, primero se ordenan las notas en orden creciente: 3,0; 3,5; 4,3; 5,2; 6,1; 6,5; 6,5; 6,5; 6,8; 7,0; 7,2; 7,2; 7,3; 7,5; 7,5; 7,6; 7,7; 7,8; 7,8; 8,0; 8,3; 8,5; 8,8; 8,8; 9,0; 9,1; 9,6; 9,7; 10 y 10. Esta secuencia muestra, a primera vista, que la máxima nota es un 10, y la mínima es un 3; el rango, diferencia entre la máxima y la mínima es 7.

En un *diagrama de frecuencia acumulada*, como el de la figura 1, las notas aparecen en el eje horizontal y el número de alumnos en el eje vertical izquierdo, con el correspondiente porcentaje a la derecha. Cada punto representa el número total de estudiantes que han obtenido una calificación menor o igual que el valor dado. Por ejemplo, el punto A corresponde a 7,2, y según el eje vertical, hay 12 alumnos, o un 40%, con calificaciones menores o iguales que 7,2.

Para analizar las calificaciones obtenidas por 10 clases de 30 alumnos cada una en cuatro exámenes distintos (un total de 1.200 calificaciones), hay que tener en cuenta que la cantidad de datos es demasiado grande para representarlos como en la figura 1. El estadístico tiene que separar los datos en grupos elegidos previamente denominados *intervalos*. Por ejemplo, se pueden utilizar 10 intervalos para tabular las 1.200 calificaciones, que se muestran en la columna (a) de la tabla de distribución de datos adjunta; el número de calificaciones por cada intervalo, llamado *frecuencia* del intervalo, se muestra en la columna (c). Los números que definen el rango de un intervalo se denominan *límites*. Es conveniente elegir los límites de manera que los rangos de todos los intervalos sean iguales y que los puntos medios sean números sencillos. Una calificación de 8,7 se cuenta en el intervalo entre 8 y 9; una calificación igual a un límite de intervalo, como 9, se puede asignar a cualquiera de los dos intervalos, aunque se debe hacer de la misma manera a lo largo de toda la muestra. La *frecuencia relativa*, columna (d), es la proporción entre la frecuencia de un intervalo y el número total de datos. La *frecuencia acumulada*, columna (e), es el número de estudiantes con calificaciones iguales o menores que el rango de cada intervalo sucesivo. Así, el número de estudiantes con calificaciones menores o iguales a 3 se calcula sumando las frecuencias de la columna (c) de los tres primeros intervalos, dando 53. La *frecuencia acumulada relativa*, columna (f), es el cociente entre la frecuencia acumulada y el número total de notas.

Los datos de una tabla de distribución de frecuencias se pueden representar gráficamente utilizando un histograma o diagrama de barras (como en la figura 2), o como un *polígono de frecuencias acumuladas* (como en la figura 3). El histograma es una serie de rectángulos con bases iguales al rango de los intervalos y con área proporcional a sus frecuencias. El polígono de la figura 3 se obtiene conectando los puntos medios de cada intervalo de un histograma de frecuencias acumuladas con segmentos rectilíneos.

En los periódicos y otros medios de comunicación los datos se representan gráficamente utilizando símbolos de diferente longitud o tamaño que representan las distintas frecuencias.

PARÁMETROS ESTADÍSTICOS:

Las tablas estadísticas son una forma organizada de dar toda (o casi toda) la información, todos los datos de que disponemos.

Con las gráficas estadísticas se pierde algo de información (mucho o poco, según los casos) pero el mensaje entre pos los ojos, que es lo que se pretende.

En cualquiera de los dos casos, la cantidad de datos que se dan es excesiva para que sea operativo, para poder hacer referencias concisas a esa distribución o comparaciones rápidas con otras distribuciones.

Esa es la razón de ser de los parámetros estadísticos, el resumir en un número un aspecto relevante de la distribución que pueda dar una idea de la misma o compararla en ese aspecto con otras.

Hay muchos parámetros estadísticos. Suelen clasificarse en:

VALORES DE LA TENDENCIA CENTRAL

Una vez que los datos han sido reunidos y tabulados, comienza el análisis con el objeto de calcular un número único, que represente o resuma todos los datos. Dado que por lo general la frecuencia de los intervalos centrales es mayor que el resto, este número se suele denominar valor o medida de la *tendencia central*.

Sean $x_1, x_2, \dots, x_n$ los datos de un estudio estadístico. El valor utilizado más a menudo es la *media aritmética* o *promedio aritmético* que se escribe $\bar{x}$, y que es igual a la suma de todos los valores dividida por n :

El símbolo Σ , o *sumatorio*, denota la suma de todos los datos. Si las x se agrupan en k intervalos, con puntos medios $m_1, m_2, \dots, m_k$ y frecuencias $f_1, f_2, \dots, f_k$, la media aritmética viene dada por

donde $i = 1, 2, \dots, k$.

La *mediana* y la *moda* son otros dos valores de la tendencia central. Si las x se ordenan según sus valores numéricos, si n es impar la mediana es la x que ocupa la posición central y si n es par la mediana es la media o promedio de las dos x centrales. La moda es la x que aparece con mayor frecuencia. Si dos o más x aparecen con igual máxima frecuencia, se dice que el conjunto de las x no tiene moda, o es *bimodal*, siendo la moda las dos x que aparecen con más frecuencia, o es *trimodal*, con modas las tres x más frecuentes.

MEDIDAS DE LA DISPERSIÓN

Normalmente la estadística también se ocupa de la *dispersión* de la distribución, es decir, si los datos aparecen sobre todo alrededor de la media o si están distribuidos por todo el rango. Una medida de la dispersión es la diferencia entre dos *percentiles*, por lo general entre el 25 y el 75. El percentil p es un número tal que un p por ciento de los datos son menores o iguales que p . En particular, los percentiles 25 y 75 se denominan *cuartiles* inferior y superior respectivamente. La *desviación típica* es otra medida de la dispersión, pero más útil que los percentiles, pues está definida en términos aritméticos como se explica a continuación. La *desviación* de un elemento del conjunto es su diferencia con respecto a la media; por ejemplo, en la sucesión $x_1, x_2, \dots, x_n$ la desviación de x_1 es $x_1 - \bar{x}$, y el cuadrado de la desviación es $(x_1 - \bar{x})^2$. La *varianza* es la media del cuadrado de las desviaciones. Por último, la desviación típica, representada por la letra griega sigma (σ), es la raíz cuadrada de la varianza, y se calcula de la siguiente manera:

Si la desviación típica es pequeña, los datos están agrupados cerca de la media; si es grande, están muy dispersos.

Medidas de posición: cuartiles, deciles, centiles. Señalan la situación de algunos valores importantes en la distribución

Medidas de asimetría: para señalar si la distribución está sesgada hacia uno u otro lado.

Medidas de apuntamiento: indican si la distribución es más o menos puntiaguda.

•