TÍTULO: LA ESTADÍSTICA: FUENTE PARA EL DESARROLLO HUMANO.

Autores: Otoniel Riverón Portela.
Juan Antonio Martín Alfonso.
Idalia González Companionis.
Ángel Gómez Argüelles.
Dirección: Universidad de Ciego de Ávila, Km. 9 carretera Ciego Morón. Cuba.
Teléfonos: 033266434, 2-5337 Fax 033266365
E-mail: oto@centic.unica.cu

RESUMEN:

Este trabajo brinda una recopilación de los principales momentos del surgimiento de la Estadística como ciencia, así como una descripción de los principales métodos y modelos empleados por esta ciencia como soporte del desarrollo alcanzado por el hombre en el dé cursar histórico de la humanidad.

INTRODUCCIÓN:

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 la 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 caloringios 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 científica 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.

DESARROLLO:

Métodos estadísticos

La materia prima de la estadística consiste en conjuntos de números obtenidos al contar o medir cosas. Al recopilar datos estadísticos se ha de tener especial cuidado para garantizar que la información sea completa y correcta.

El primer problema para los estadísticos reside en determinar qué información y cuánta se ha de reunir. En realidad, la dificultad al compilar un censo está en obtener el número de habitantes de forma completa y exacta; de la misma manera que un físico que quiere contar el número de colisiones por segundo entre las moléculas de un gas debe empezar determinando con precisión la naturaleza de los objetos a contar. Los estadísticos se enfrentan a un complejo problema cuando, por ejemplo, toman una muestra para un sondeo de opinión o una encuesta electoral. El seleccionar una muestra capaz de representar con exactitud las preferencias del total de la población no es tarea fácil.

Para establecer una ley física, biológica o social, el estadístico debe comenzar con un conjunto de datos y modificarlo basándose en la experiencia. Por ejemplo, en los primeros estudios sobre crecimiento de la población los cambios en el número de habitantes se predecían calculando la diferencia entre el número de nacimientos y el de fallecimientos en un determinado lapso. Los expertos en estudios de población comprobaron que la tasa de crecimiento depende sólo del número de nacimientos, sin que el número de defunciones tenga importancia. Por tanto, el futuro crecimiento de la población se empezó a calcular basándose en el número anual de nacimientos por cada 1.000 habitantes. Sin embargo, pronto se dieron cuenta que las predicciones obtenidas utilizando este método no daban resultados correctos. Los estadísticos comprobaron que hay otros factores que limitan el crecimiento de la población. Dado que el número de posibles nacimientos depende del número de mujeres, y no del total de la población, y dado que las mujeres sólo tienen hijos durante parte de su vida, el dato más importante que se ha de utilizar para predecir la población es el número de niños nacidos vivos por cada 1.000 mujeres en edad de procrear. El valor obtenido utilizando este dato mejora al combinarlo con el dato del porcentaje de mujeres sin descendencia. Por tanto, la diferencia entre nacimientos y fallecimientos sólo es útil para indicar el crecimiento de población en un determinado periodo de tiempo del *pasado,* el número de nacimientos por cada 1.000 habitantes sólo expresa la tasa de crecimiento en el mismo periodo, y sólo el número de nacimientos por cada 1.000 mujeres en edad de procrear sirve para predecir el número de habitantes en el *futuro.*

Tabulación y presentación de los datos

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.

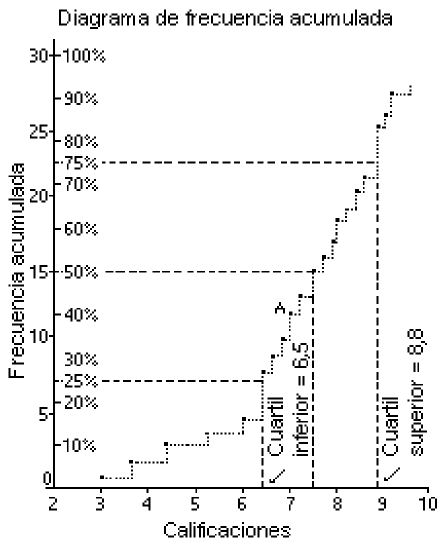


Figura 1

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.

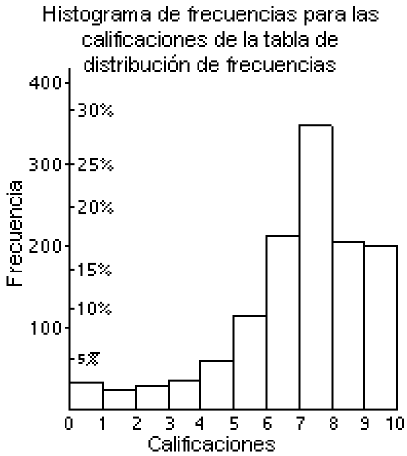


Figura 2

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.

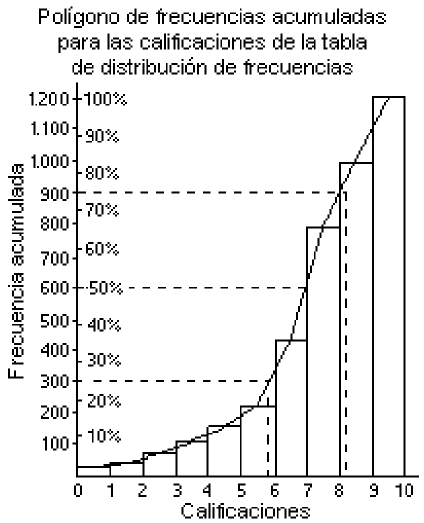


Figura 3

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.

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

$$\bar{x} = \frac{\Sigma x}{n}$$

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

$$\frac{\Sigma f_i m_i}{\Sigma f_i}$$

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

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, …, x_n la desviación de x_1 es x_1 - □, y el cuadrado de la desviación es (x_1 - □)^2. La *varianza* es la media del cuadrado de las desviaciones. Por último, la desviación típica, representada por la letra griega sigma (□), es la raíz cuadrada de la varianza, y se calcula de la siguiente manera:

$$\sigma = \sqrt{\frac{1}{n}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right]} = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

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

Correlación

Cuando dos fenómenos sociales, físicos o biológicos crecen o decrecen de forma simultánea y proporcional debido a factores externos, se dice que los dos fenómenos están *positivamente correlados.* Si uno crece en la misma proporción que el otro decrece, los dos fenómenos están *negativamente correlados.* El grado de correlación se calcula aplicando un coeficiente de correlación a los datos de ambos fenómenos. El coeficiente de correlación más utilizado es

$$\frac{\Sigma\left(\frac{x}{\sigma^x} \cdot \frac{y}{\sigma^y}\right)}{N}$$

donde *x* es la desviación de una variable con respecto a su media, e *y* es la desviación de la otra variable con su media; *N* es el número total de casos en las series. Una correlación positiva perfecta tiene un coeficiente +1, y para una correlación negativa perfecta es -1. La ausencia de correlación da como coeficiente 0. Por ejemplo, el coeficiente 0,89 indica una correlación positiva grande, -0,76 es una correlación negativa grande y 0,13 es una correlación positiva pequeña.

Modelos matemáticos

Un modelo matemático es una representación ideal (en la forma de un sistema, proposición, fórmula o ecuación) de un fenómeno físico, biológico o social. Así, un dado teórico perfectamente equilibrado, que se puede lanzar de forma aleatoria, es un modelo matemático de un dado real. La probabilidad de que en *n* lanzamientos de un dado matemático se obtenga *k* veces un 6 es

$$P(k) = \binom{n}{k}\left(\frac{1}{6}\right)^{k}\left(\frac{5}{6}\right)^{n-k}$$

donde (□) es la representación de un número combinatorio

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{1\cdot 2\cdot\ \ldots\ \cdot k}$$

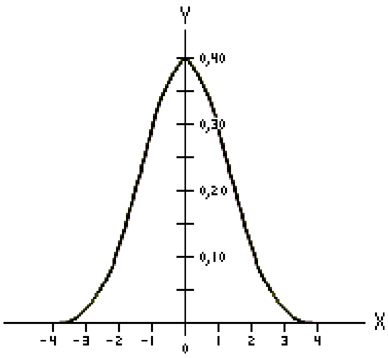
$$\left(\binom{n}{0} = 1\right)$$

El estadístico que utiliza un dado real debe diseñar un experimento, como lanzar el dado un gran número de veces, para determinar, a partir de los resultados obtenidos, la posibilidad de que el dado esté perfectamente equilibrado y de que el lanzamiento sea aleatorio.

Muchos conjuntos de medidas experimentales presentan el mismo tipo de distribución de frecuencias que se pueden representar con un modelo matemático único. Por ejemplo, el número de veces que sale un 6 al lanzar un dado *n* veces, el peso de *N* garbanzos tomados al azar de una bolsa o las presiones atmosféricas medidas por distintos estudiantes sucesivamente en el mismo barómetro. En todos los casos los valores presentan patrones de frecuencia muy similares. Los estadísticos adoptan un modelo que es un prototipo o idealización matemática de todos esos patrones o distribuciones. Una forma de modelo matemático puede ser una ecuación de la distribución de frecuencias, en la que el número de medidas o valores se considera infinito:

$$y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

donde e es la base de los logaritmos neperianos, e *y* representa la *frecuencia* del valor *x.* La gráfica de esta fórmula (figura 4) es una curva en forma de campana llamada curva de probabilidad normal o *gaussiana*:



Curva de probabilidad normal

Figura 4

CONCLUSIONES:

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.

BIBLIOGRAFÍA

- "Matemáticas", Enciclopedia Microsoft® Encarta® 98 © 1993-1997 Microsoft Corporation. Reservados todos los derechos.
- Historia e Historias de matemáticas
Autor: Mariano Perero.
Editorial: Grupo Editorial Iberoamérica.
- *Libro de matemáticas 1º B.U.P*
Autores: M. de Guzmán, J. Colera, A. Salvador.
Editorial: Anaya.
- *Historia de la matemática*
Volumen I.
Autores: J. Rey Pastor y José Belsini.
Editorial: Gedisa. 2ª edición.
- Revista "Ciencia y Vida"
Número 2. Abril de 1998
Páginas: 56-67.
Autor: C. Chauvenau.
- Gran Enciclopedia LAROUSSE Tomo 7 (pág. 40 - 41) .
Enciclopedia DURVAN Tomo 12 - 774 .
- "HISTORIA E HISTORIAS DE MATEMÁTICAS" (pág.3 - 4- 5 - 8- 136 - 137 - 138) Autor: Mariano Perero. Grupo Editorial Iberoamérica.

Internet:
Web sobre Historia de las Matemáticas de la Universidad de St. Andrews (Escocia)
Web sobre matemáticas de Kevin Brown (USA)