Estadística en la escuela

Algunos elementos para su trabajo

Carla Damisa | Profesora de Matemática. Formadora de maestros en Enseñanza de la Matemática.

En la nueva formulación de los contenidos programáticos para la Educación Primaria aparece como un contenido específico de Matemática a ser abordado durante toda la escolaridad, el contenido *Estadística*.

Este artículo pretende ofrecer algunos aspectos del tratamiento matemático del contenido Estadística así como establecer algunas relaciones con su enseñanza.

Es un contenido muy amplio, por lo que no pretendemos agotarlo, ni es posible hacer un desarrollo exhaustivo del mismo en una primera presentación. Nuestro propósito es abordar algunas ideas fundamentales de la Estadística y de su tratamiento a nivel de Educación Primaria.

Comenzaremos por presentar algunas referencias históricas, y desarrollaremos el contenido en lo referente a la Estadística Descriptiva.

La Estadística es considerada una ciencia, con una fuerte base matemática, que se dedica a la recolección, organización y tabulación de datos, así como a realizar ciertos cálculos con el objetivo de extraer conclusiones sobre el objeto de estudio. En definitiva se trata de describir ciertos sucesos que son estudiados y a partir de la realización de determinados cálculos poder establecer algunas conclusiones, comprobar las hipótesis propuestas y determinar ciertas generalizaciones con algunas condiciones.

Thomas Bayes

Fue Thomas Bayes, alrededor de 1760, el primero en introducir elementos matemáticos en este proceso de estudio de datos, dando de este modo el puntapié inicial para lo que hoy conocemos como Estadística.

Asimismo fue Gottfried Achenwall (1719-1772) quien utilizó el término “estadística” con en el significado “ciencia de las cosas que pertenecen al Estado”. Etimológicamente, la palabra “estadística” procede del latín *“statisticum collegium”* (consejo de Estado) y de su derivado italiano *“statista”* (hombre de Estado o político). Es desde su origen donde se explicita el vínculo estrecho que tiene la estadística con lo político y las relaciones con el Estado.

Hoy, además de continuar estrechamente ligada al Estado, aparece muchas veces como un contenido “transversal”, utilizado por varias disciplinas. Por ejemplo, la Física, la Biología, la Medicina, etc.

Comúnmente somos usuarios de la Estadística; hoy proponemos, en este artículo, introducirnos en su estudio y conocerla mínimamente para poder entenderla mejor y establecer algunas conclusiones a partir de ella.

Algunos autores (Tapia, 1984) hablan de un “método estadístico”. Estos autores plantean que, partiendo de hechos y observaciones experimentales, se pueden extraer conclusiones generales sobre el objeto que se estudia en cada caso. Al hablar de método estadístico, estos autores citan ciertas etapas del proceso mediante las cuales se llega a establecer conclusiones.

Estas etapas serían:

1) Recuento, relevamiento o compilación de datos.
2) Tabulación y agrupamientos de datos. Gráficos.
3) Medición de datos.
4) Inferencia estadística. Descripción.

Algunos conceptos a tener en cuenta

Más arriba hemos dado una posible definición de Estadística. En ese sentido estudiaremos algunos elementos en relación a ella. Existen dos formas diferentes de estudiar problemas estadísticos, o bien sobre toda la población o sobre una parte representativa de ella.

Se llama **población** o **universo** cuando trabajamos con el conjunto formado por ***todos*** los elementos a estudiar. Cada elemento se llama **unidad de análisis**.

Cuando se trabaja con todos los elementos de una población o universo, decimos que es un censo. Revisando un poco la historia observamos que los primeros censos estadísticos se referían al estudio de poblaciones compuestas por personas. Luego se fue extendiendo el uso de la Estadística no solamente a estudiar algunas características referidas a las personas o propiedades de estas, sino a ser aplicada a la física, la agricultura, la biología, la sociología, la música, el mercado, etc.

Los elementos u observaciones se refieren a diferentes características o atributos. Por ejemplo: si se censan escuelas, cada escuela es una unidad de análisis y el conjunto de todas las escuelas es la población. Si se realiza un censo sobre tipos de películas, cada película (acción, drama, ficción, etc.) sería un individuo o elemento, y la población el conjunto de películas.

La Estadística Descriptiva trabaja sobre toda la población, y la Estadística Inferencial trabaja sobre un subconjunto de la población llamado muestra, y posteriormente extiende sus resultados sobre la población total. Por ejemplo, si queremos saber quién va a ganar las elecciones no realizamos una encuesta (censo) a todos los habilitados a votar, sino que realizamos una entrevista a un subconjunto de esa población. Este subconjunto debe ser representativo, es decir, debe mantener las mismas características que la población total o universo. Cuando se procede de esta manera, se dice que se realiza un muestreo.

Muestreo:
acción de escoger muestras representativas.
Muestra:
subconjunto de una población que se considera representativa de ella.

¿Qué "cosas" se pueden estudiar en una población o muestra?

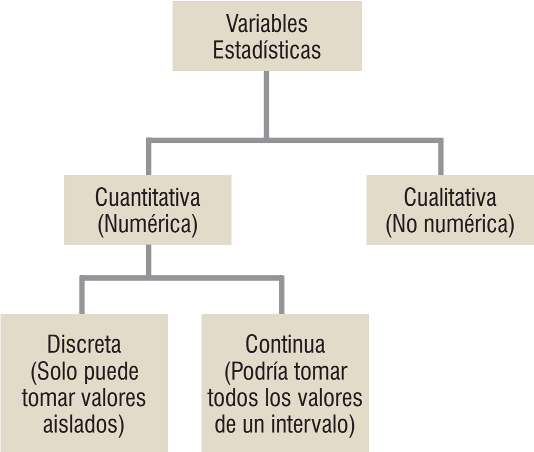
Seguimos con el ejemplo de las películas; si deseamos estudiar el comportamiento de los alumnos de un grupo a la hora de la elección del tipo de películas que miran, entonces la población serían los alumnos de ese grupo (objeto de estudio), la variable estudiada o el atributo de esa población a ser estudiado sería "tipo de película que prefieren". En este caso estamos frente a una **variable** del tipo **cualitativa**, se refiere a una propiedad o característica de las películas, que no toma valores numéricos.

En este caso, lo que estamos estudiando de esa población (alumnos del grupo) es la variable "tipo de película que prefieren". Sin embargo, de esa misma población podríamos estudiar otra variable, por ejemplo, la medida de sus alturas en cm. En este caso, esta variable se expresa en números, entonces estamos frente a una **variable cuantitativa**. Además, aquí la variable puede expresarse por medio del conjunto de los números reales, porque la longitud de las alturas puede tomar todos los valores de un cierto intervalo. En este caso decimos que la **variable** es **cuantitativa continua**.

También podríamos estudiar en esa misma población, la variable "edad". En este caso también estamos frente a una variable cuantitativa, porque se puede expresar numéricamente, si definimos la variable edad como el número de años cumplidos al momento de solicitar esa información. En este caso no toma todos los valores de un intervalo, es decir, las edades pueden ser 5, 6, 7 u 8 años, no 7,3 años. Cuando sucede esto decimos que la **variable es cuantitativa discreta**, porque se representa a través del conjunto de los números naturales.

En suma, algunas características importan especialmente en tanto cualidades cuya posibilidad de mensura no agrega información, porque se trata de categorías o atributos: sexo, religión, de qué cuadro son "hinchas". En otros casos, las características o atributos sí son representados numéricamente y revisten cierto significado, por lo tanto estamos frente a dos tipos de variables: cuantitativas y cualitativas.

A continuación presentamos un cuadro[1] que clarifica lo ejemplificado anteriormente:



Con lo que sabemos hasta ahora, un ejemplo concreto:

"Se hizo una encuesta a todos los alumnos de un 4º año para saber cuántos hijos tienen sus padres. Los resultados obtenidos son los siguientes:

2; 2; 1; 2; 3; 3; 2; 3; 2; 2; 1;
4; 4; 2; 2; 1; 5; 2; 3; 2; 6; 3;
3; 3; 2; 4; 7; 3; 3; 2; 3; 6; 1."

Como se expresa anteriormente, estamos analizando la población total del grupo, estamos trabajando con lo que hemos definido como un censo. La variable que estamos estudiando es el número de hijos de esas familias, por lo tanto, es cuantitativa discreta.

Sin embargo, los datos, así como están presentados, no nos permiten extraer fácilmente conclusiones sobre las características de ese grupo en cuanto al número de hijos de cada familia. En este caso, al estudiar un grupo pequeño, algunas lecturas pueden realizarse directamente; pero si la población fuera mucho mayor (300 alumnos, por ejemplo), ya no sería posible analizar los datos directamente.

Para un análisis profundo es necesaria la organización de los datos con el fin de obtener la información buscada. Para ello, una forma muy cómoda es tabularlos en una tabla como la que mostramos a continuación, de modo que la organización de esos datos nos vaya habilitando a trabajar con ellos de manera más organizada.

En general, en una tabla de frecuencias encontramos los siguientes datos:

- N: indica el número total de observaciones realizadas.

- f: representa la frecuencia absoluta o simplemente frecuencia que indica la cantidad de veces que ocurre cada valor de la variable en la muestra.

- f_r: la frecuencia relativa, indica la fracción del total de la muestra que corresponde a cada valor de la variable y se calcula: $f_r = \frac{f}{n}$

- P: porcentaje, indica el porcentaje del total de elementos de la muestra que corresponde a cada valor de la variable y se calcula: $f_p = f_r \times 100$

- f_a: frecuencia acumulada, indica la frecuencia absoluta que se acumula hasta cada fila que se señala en la tabla y se calcula a través de las sumas de las frecuencias absolutas hasta esa fila. Por ejemplo, hasta con 3 hijos tenemos un total de 26 familias de ese grupo, es decir, sumamos la frecuencia de cada valor de la variable hasta 3 hijos inclusive.

[1] Extraído de J. Colera; M. Guzmán (1994).

Nº de hijos por familia	f	f_a	f_r	P (%)
Familia con 1 hijo	4	4	$\frac{4}{33} \cong 0{,}121$	12,1
Familia con 2 hijos	12	16	$\frac{12}{33} \cong 0{,}364$	36,4
Familia con 3 hijos	10	26	$\frac{10}{33} \cong 0{,}303$	30,3
Familia con 4 hijos	3	29	$\frac{3}{33} \cong 0{,}091$	9,1
Familia con 5 hijos	1	30	$\frac{1}{33} \cong 0{,}030$	3,0
Familia con 6 hijos	2	32	$\frac{2}{33} \cong 0{,}610$	6,1
Familia con 7 hijos	1	33	$\frac{1}{33} \cong 0{,}030$	3,0
N	33		1	100%

Los datos recogidos y los tabulados son los mismos; sin embargo, ya podemos comenzar a extraer alguna conclusión. Como vemos, la mayoría de los hogares (el 78,8%) tiene entre 1 y 3 hijos. Para calcularlo realizamos la suma de los porcentajes correspondientes desde 1 a 3 hijos respectivamente.

Al analizar la tabla vemos que la suma de la frecuencia relativa es 1 y que la suma de las frecuencias porcentuales es 100.

Como ya mencionamos, la frecuencia relativa es el cociente entre la frecuencia absoluta de cada valor que toma la variable por el número de observaciones totales realizadas. Si analizamos la relación para calcular el porcentaje que corresponde, por ejemplo, a los hogares que tienen 2 hijos, deberíamos realizar la "vieja regla de tres":

12 hogares (tienen 2 hijos) -------- x%
33 hogares -------------------------- 100%

De donde: $x = \frac{100 \cdot 12}{33} = 36{,}3$

Esta última expresión la podemos escribir así:

$$x = \frac{12}{33} \cdot 100 \cong 36{,}3 \ (*)$$

Si observamos la tabla, 36,3 sería el porcentaje correspondiente al número de hogares de ese grupo que tiene 2 hijos (P).

Siguiendo con el análisis, el primer factor de la última expresión (*) ($\frac{12}{33}$) resulta que es el mismo valor de la frecuencia relativa. Es decir que rápidamente, al tener los datos tabulados, podemos directamente calcular el porcentaje sin hacer para cada caso la regla de tres, aunque está implícita, solamente multiplicando por 100 a la frecuencia relativa es suficiente, es decir:

$$P = f_r \cdot 100$$

Si analizamos la columna que corresponde a la frecuencia acumulada (f_a) observamos que la vamos obteniendo al sumar la frecuencia de los diferentes valores. Por lo tanto, fácilmente podemos decir que de 33 casos, 26 son hogares que tienen entre 1 y 3 hijos inclusive.

Al disponer los datos en tablas y comenzarlos a procesar vemos que nos facilitan el establecer relaciones numéricas entre las variables que estamos estudiando. En este caso, número de hijos por hogares del grupo en cuestión.

Ahora a graficar...

Los datos recogidos y organizados en la tabla anterior nos habilitan a presentarlos de otra manera. Por ejemplo, manejando diferentes tipos de gráficos donde esos datos estén registrados. A partir de otras presentaciones de los datos podemos rápidamente extraer información y, por lo tanto, establecer relaciones analizando el tipo de gráfico que se nos presente o que confeccionemos.

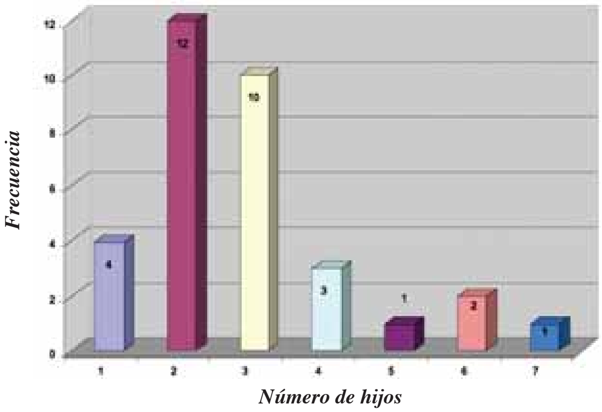
Los gráficos más usuales que normalmente confeccionamos o encontramos son pictogramas, barras (histograma como caso particular), de puntos, de torta.

Comúnmente, el gráfico de barras es el más usado, por ser muy simple su confección e interpretación. Estos gráficos se utilizan generalmente para representar variables cualitativas o cuantitativas del tipo discreto. La longitud de cada barra es igual a la frecuencia de cada observación. Las barras pueden ser tanto horizontales como verticales.

En nuestro ejemplo, Nº de hijos por familia en 4º año, podríamos realizar el siguiente gráfico de barras, donde la "altura" de cada barra representa la cantidad de familias con 1, 2,…, etc. hijos. En definitiva, el eje horizontal o eje de abscisas representa el número de hijos, y el eje vertical o eje de ordenadas representa la frecuencia de cada observación.

Un caso particular de gráfico de barras es el *Histograma*. El gráfico consiste en una serie de rectángulos adyacentes cuya "base" es de igual longitud y la altura coincide con la frecuencia de cada observación. Esto trae como consecuencia que el área de cada rectángulo (barra) es proporcional a la cantidad a representar. Este tipo de gráfico se utiliza tanto para variables cuantitativas como cualitativas. Es muy frecuente usarlo cuando los datos están organizados por intervalos.

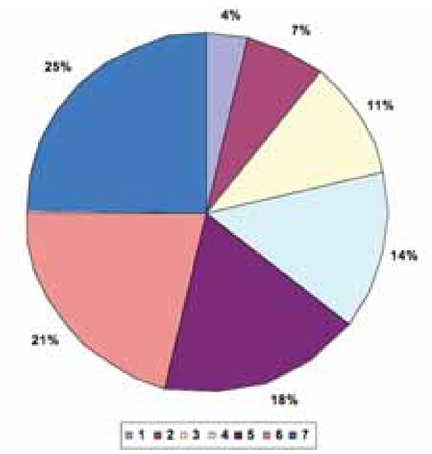
Gráfico de Barras de Frecuencias



El gráfico de *torta* o de *sectores* resulta útil cuando se quieren comparar datos a "simple vista". Mediante los sectores circulares se representa la proporción entre los distintos valores que toma la variable. Generalmente se expresan en porcentajes.

En nuestro caso, la variable estudiada en la población es Nº de hijos por familia, de donde cada sector representará esa cantidad. Para confeccionar el gráfico de torta o circular se establece una relación de proporcionalidad entre el ángulo al centro correspondiente al círculo que elegimos con el porcentaje que representa cada valor de la variable.

Gráfico Circular de Frecuencias



Medición de datos

En el comienzo hicimos referencia al método estadístico y a partir de allí hemos trabajado en base a "esas etapas". Ahora tendríamos que comenzar a trabajar con la medición de datos, es decir, comenzar a trabajar realizando algunos cálculos donde podamos identificar los parámetros estadísticos; en definitiva, calcular las medidas de centralización y las de dispersión.

Las **medidas de centralización** son la *media* o *promedio*, la *mediana* y la *moda*. Las **medidas de dispersión** son el *rango*, la *desviación* o *dispersión*, la *desviación media* o *desviación promedio*, la *varianza* y la *desviación estándar*.

Para el espíritu de nuestro artículo, es decir, el tratamiento de la estadística a nivel escolar, nos centraremos en el trabajo con las tres *medidas de centralización*, y con el *rango*, y la *desviación* como medidas de dispersión.

Cálculo de las medidas de centralización

Media aritmética o Promedio ($\bar{X}$): es el cociente entre la suma de los valores de las variables por el número de observaciones realizadas.

En nuestro ejemplo deberíamos sumar todos los valores de la variable Nº de hijos por familia, y dividirlos entre las 33 observaciones realizadas. Al sumar todos esos valores de forma desorganizada, podemos cometer errores. Es por eso que al realizar los cálculos a partir de la tabulación hecha en la tabla anterior, donde se encuentran los datos organizados, se podrá garantizar el intento de no cometer errores en dichos cálculos.

Si organizamos la suma que debemos realizar, tenemos que el valor de la variable "1 hijo" se repite 4 veces, por lo tanto, lo que debemos sumar es el resultado de 1 x 4; de igual manera, el valor de la variable "2 hijos" se repite 12 veces, de donde debemos sumar 12 x 2, de ahí que lo organicemos de la siguiente manera:

Nº de hijos (x)	f	x.f
1	4	4
2	12	24
3	10	30
4	3	12
5	1	5
6	2	10
7	1	7
N	33	92

Para realizar el cálculo de la media aritmética o promedio ($\overline{X}$):

$\overline{X} = \dfrac{\Sigma x \bullet f}{n}$, el signo Σ indica la suma de los sumandos ($x \bullet f$)

$$\overline{X} = \frac{92}{33} \cong 2{,}78 \cong 3$$

$$\overline{X} = 3$$

Si observamos el cálculo realizado, debimos aproximar el resultado a 3 hijos porque la variable es cuantitativa discreta y "no podemos" decir que la media de los datos recogidos es 2,78 hijos; de donde debemos siempre vigilar y relacionar con la magnitud que estamos trabajando.

Si bien el promedio es un parámetro muy importante, no siempre la media o el promedio es característica de la colección de datos. Por eso es necesario introducir otras medidas de centralización como lo son la *mediana* y la *moda*.

Mediana (M_e): si todos los valores observados de la variable se ordenan en sentido creciente (o decreciente), la mediana es el valor de la variable que ocupa el lugar central, es decir, el que deja a un lado y a otro el mismo número de observaciones.

Para nuestro caso tenemos 33 observaciones, las ordenamos todas en forma creciente; resulta:

Mediana

1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,**3**,3,3,3,3,3,3,3,3,3,4,4,4,5,6,6,7

$$\mathbf{M_e = 3}$$

Ordenar así los datos sería costoso, por lo tanto podemos utilizar la tabla anterior. Como sabemos, el número de observaciones realizadas es 33; por lo tanto, ordenadas, el valor que ocupa el lugar 17 estaría dejando 16 observaciones hacia cada lado. Por lo tanto, ese lugar corresponde al valor de la observación de la mediana. Nos valemos de la tabla para ir recorriendo la columna de la frecuencia hasta llegar al lugar de frecuencia 17. En este caso tenemos 16 lugares que corresponden 4 a 1 hijo, y 12 a 2 hijos, en total 16 observaciones. Por lo tanto, la observación siguiente será la mediana. Para nuestro ejemplo, el primer valor correspondiente a 3 hijos corresponde al lugar 17, dejando así 16 observaciones de cada lado.

¿Qué sucedería si el número de observaciones recogidas fuese par?

Si en vez de tener 33 observaciones tuviésemos 34, por ejemplo, habría en nuestra clase un alumno que pudiera tener 1 hermano, por lo tanto, su familia tendría 2 hijos. Esto afectaría en nuestro análisis a que en vez de que la frecuencia para 2 hijos fuese 12, sería 13. Por lo tanto, tendríamos la siguiente tabla:

Nº de hijos (x)	f
1	4
2	13
3	10
4	3
5	1
6	2
7	1
N	34

Entre número de hijos 1 y 2, tenemos 17 observaciones, pero están quedando otras 17 con más de dos hijos, lo que implica que no podemos elegir el valor de la mediana porque, al ser el número de observaciones par, no obtenemos igual cantidad de observaciones para cada lado y, por lo tanto, deberemos proceder eligiendo dos valores y hacer el promedio de los mismos.

En este caso, el número de observaciones es 34 (agregamos una), por lo tanto, tomamos el valor que corresponde al lugar 17 y el 18 porque dejan 16 valores por debajo de esos y 16 por encima, lo que indica que tenemos dos valores centrales, por lo tanto, deberemos realizar el promedio de ellos para obtener el valor correspondiente a la mediana:

- el lugar 17 corresponde al último valor de la variable 2 hijos,
- el lugar 18 corresponde al valor de la variable 3 hijos,
- realizando el promedio de los valores correspondientes a cada lugar obtenemos

$$M_e = \frac{2+3}{2} = 2{,}5 \cong 3 \text{ hijos}$$

$$\mathbf{M_e = 3}$$

Por lo tanto, en los dos ejemplos trabajados, tanto para número de observaciones pares como impares podemos calcular la mediana.

Ahora nos falta trabajar con la *moda*. Este concepto se relaciona con la idea de "moda" en el lenguaje cotidiano. La moda es lo que se "usa más".

En Estadística, el concepto de **Moda o Modo** (**M_o**) es el valor de la variable que más veces se repite, o sea, el valor que presenta mayor frecuencia.

Es útil como medida de tendencia central, solamente en aquellos casos en que un valor de la variable es mucho más frecuente que el resto. Se basa en la idea de "lo que es moda" o en el "comportamiento de la mayoría" para tomar a cierto valor como representativo del comportamiento de los datos evaluados.

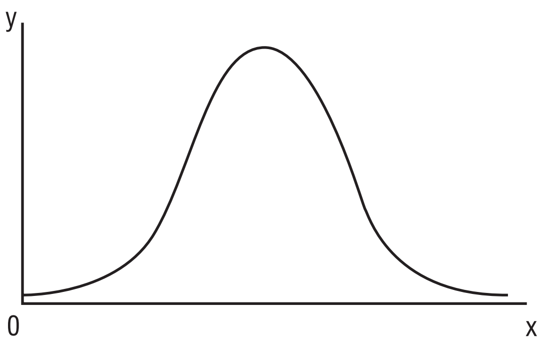
Para nuestro ejemplo, la *moda* sería el valor correspondiente a 3 hijos:

$$M_o = 3$$

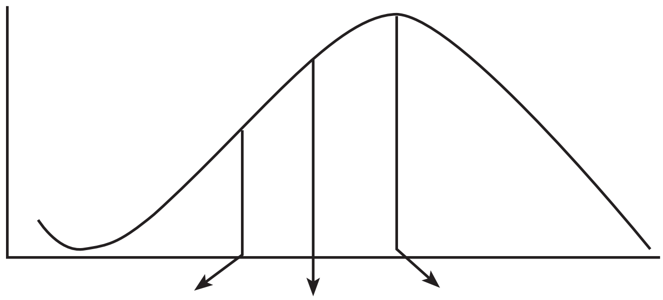
A partir de los cálculos de medidas de tendencia central que hemos determinado en nuestro ejemplo, podemos establecer algunas relaciones. Por ejemplo, en general, con la *media* conocemos el valor promedio de las observaciones realizadas. Con la *moda*, por su parte, tenemos una idea de qué valor es el más frecuente; y con la *mediana*, el valor que "equidista de los extremos".

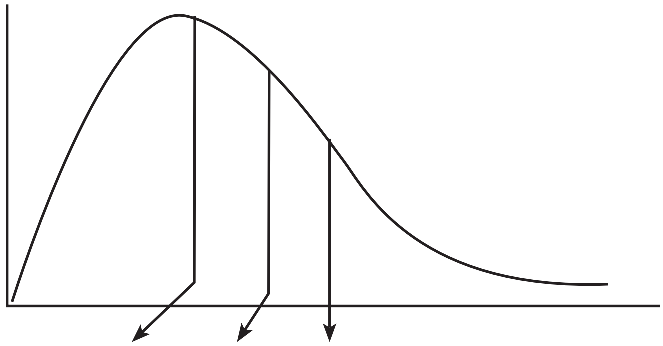
Estos valores nos indican cómo se agrupan los valores de las observaciones realizadas. Si los tres valores coinciden, como en nuestro caso, estamos en presencia de una distribución simétrica de los valores de las observaciones realizadas. Si los valores de la *media*, *mediana* y *moda* no coinciden, las distribuciones son asimétricas. Los gráficos que siguen nos muestran estas características:

Simétrica:
la *media*, la *mediana* y la *moda* coinciden



Asimétricas:





Cálculo de las medidas de dispersión

Conceptualmente, los valores de tendencia central tienen sentido cuando son complementados con la lectura de los valores de medidas de dispersión. En este artículo no nos encargaremos de todos los parámetros de dispersión, sino de algunos, para poder usarlos a nivel escolar complementando las medidas de tendencia central.

Para ejemplificar el comentario anterior podríamos pensar en una muestra representativa en varias clases sobre el número de hijos de cada familia y obtener los siguientes 20 valores en cada clase:

Nº de hijos	f
1	5
2	2
3	3
4	3
5	2
6	5
N	20

Tabla 1

Nº de hijos	f
1	2
2	3
3	5
4	5
5	3
6	2
N	20

Tabla 2

Nº de hijos	f
1	3
2	3
3	4
4	4
5	3
6	3
N	20

Tabla 3

Las series de datos anteriores tienen el mismo rango[2] o amplitud, la misma media y la misma mediana:

$$\overline{X} = 3,5 \qquad M_e = 3,5$$

Rango o amplitud: 6-1 = 5

Sin embargo, analizando la distribución de los datos en las tablas, o si realizáramos los gráficos correspondientes, podemos observar que en la serie correspondiente a la Tabla 1, los datos se concentran en los extremos, porque de 20 datos, 10 están concentrados en el número de hijos 1 y 6. Es decir que el 50% de los valores que toma la variable se encuentran en los extremos.

En cambio, en la Tabla 2, los datos se concentran en el medio, para valores 3 y 4 como Nº de hijos. Es decir que el 50% se concentra en el centro de los valores.

Asimismo, analizando la Tabla 3, vemos que los valores tienen una distribución pareja, es decir, más o menos todos los valores que toma la variable tienen la misma frecuencia.

Esto nos lleva a reflexionar que solamente con los valores de tendencia central no es suficiente para analizar los sucesos que estamos estudiando. Los valores de desviación nos ayudan a caracterizar dichas distribuciones. También en este análisis hemos agregado el rango como parámetro de dispersión pero, en nuestros ejemplos, todas tienen el mismo rango; por lo tanto, no será suficiente explicar el comportamiento de lo que estamos estudiando solamente con el rango como parámetro de dispersión. En consecuencia, vamos a analizar la *desviación* o *desvío* (d) y la *desviación media* (d_m) como parámetros de dispersión para dar una posible explicación.

Como ya hemos mencionado, no son los únicos parámetros de dispersión, pero nos dan pistas, en este nivel, para intentar algunas explicaciones.

Llamaremos *desvío* o *desviación* (d) de un valor de la variable a la diferencia entre el valor de la variable y la media.

La *desviación media* (d_m) es el promedio de los valores absolutos de las desviaciones.

Analizaremos, a partir de las tres tablas presentadas en nuestro ejemplo anterior, la *desviación* y la *desviación media*. Para ello necesitamos primero calcularlas, para luego analizar el comportamiento de cada serie de datos.

Nº de hijos	f	$\overline{X}$	Desvío $d = X - \overline{X}$	f . d	Valor absoluto d $\lvert X - \overline{X} \rvert$
1	5	3,5	1 - 3,5 = -2,5	5 x -2,5 = -12,5	12,5
2	2	3,5	2 - 3,5 = -1,5	-3,0	3,0
3	3	3,5	3 - 3,5 = -0,5	-1,5	1,5
4	3	3,5	4 - 3,5 = 0,5	1,5	1,5
5	2	3,5	5 - 3,5 = 1,5	3,0	3,0
6	5	3,5	6 - 3,5 = 2,5	12,5	12,5
N	20			0	34,0

La suma de los valores absolutos para esta distribución de datos es 34; ahora, como hay que calcular la desviación media, debemos dividir por el número de observaciones realizadas, resultando:

$$d_m = \frac{34}{20} = 1,7 \quad \text{(para la Tabla 1)}$$

Realizando los mismos cálculos en las Tablas 2 y 3, obtenemos las siguientes desviaciones medias:

$$d_m = \frac{24}{20} = 1,2 \quad \text{(para la Tabla 2)}$$

$$d_m = \frac{34}{20} = 1,7 \quad \text{(para la Tabla 3)}$$

El análisis toma las tres medidas: media, mediana y desviación media. Especialmente la media (o promedio) y la desviación media, porque la lectura conjunta de estos parámetros nos permite conocer cómo se agrupa la información (en torno a qué valores), cómo la serie de datos se desvía o se "aleja" de la media. Por este motivo es fácil advertir que ante la misma media, los valores de la tabla 2 se desvían o se "alejan" sólo 1,2 del valor medio. Esto significa que los valores van a estar concentrados en un entorno de $3,5 \pm 1,2$.

$$3,5 - 1,2 = 2,3$$
$$3,5 + 1,2 = 4,7$$

[2] Rango o Amplitud: es la diferencia entre el mayor y menor valor, tomados por la variable. En este caso, la diferencia entre el mayor y el menor número de hijos en las observaciones recogidas.

La mayoría de las familias estudiadas va a estar concentrada entre 2,3 hijos y 4,7 hijos (entre 2 y 5 hijos).

Razonando de manera análoga, es fácil advertir que la conducta de las familias analizadas en las Tablas 1 y 3 tienden a dispersarse más, es decir, la serie de datos tiene una distribución más "pareja". Lo interesante de estas lecturas anteriores es que pueden realizarse "sin leer la tabla", porque al relacionar las medidas de centralización y de desviación es posible realizar una aproximación real al comportamiento de las familias que estamos estudiando.

Algunas relaciones para establecer

Todos estos cálculos previos para extraer las conclusiones o establecer algunas relaciones en la población estudiada con la variable elegida no son el centro del estudio, son necesarios para poder concluir. El foco lo debemos poner en el análisis que realizamos posteriormente a los cálculos. Para ello, en estudios con bases de datos muy grandes, contamos con algunas ayudas tecnológicas como es la planilla de cálculo y paquetes informáticos estadísticos para realizarlos con facilidad y no centrarnos en ellos.

Por lo tanto, la idea no es centrarnos en los cálculos, sino enfocarnos en las posibles interpretaciones a partir de la organización de los datos en tablas y gráficos. Estas interpretaciones son las relaciones potentes que queremos construir en Estadística a nivel escolar. Son elementos que se espera que los alumnos manejen a partir de la gestión docente, a la hora de poder analizar algunos sucesos.

Estos análisis no se van a producir a lo largo de toda la escolaridad, sino que irán creciendo a la vez que avanzamos en ella.

En los primeros años nos centraremos en recolectar, organizar y, si es posible, graficar los datos recogidos. También se podrán identificar los tipos de variables a estudiar en la población seleccionada y establecer algunas relaciones entre ellos con esos elementos.

Al avanzar en la escolaridad, el uso de la planilla de cálculo, la determinación de los parámetros de medidas de centralización y algunos de dispersión nos habilitarán a confirmar las sospechas de las relaciones establecidas en años anteriores. Es decir, a través del cálculo podremos verificar si son verdaderas o no esas conclusiones provisorias que fuimos estableciendo.

El contenido Estadística, al ser trabajado en la escuela primaria, no será conveniente realizarlo de forma aislada, es decir, por sí mismo, sino en relación a otros contenidos y disciplinas como, por ejemplo, Probabilidad, entre otros. Establecer estas relaciones será tema de otro artículo.

A continuación veremos otros ejemplos de actividades referidas a un contexto de la vida cotidiana. Cuando desde los informativos o la prensa nos invaden con información, es imprescindible haber desarrollado una "mirada estadística": cuánta carne comemos los uruguayos, cómo se lee una tabla de posiciones de cuadros de fútbol, a cuántos uruguayos se consulta para saber quién ganará las próximas elecciones. Analizaremos qué conceptos estadísticos están vinculados en cada una, y algunas relaciones que se pueden establecer entre ellos.

Hay que vacunarse…

Esta gráfica es del Boletín Epidemiológico del Ministerio de Salud Pública. El punto señalado con una flecha indica el momento en que comenzó la vacunación en nuestro país. Sin conocer los datos de la tabla que dio lugar a esta gráfica:

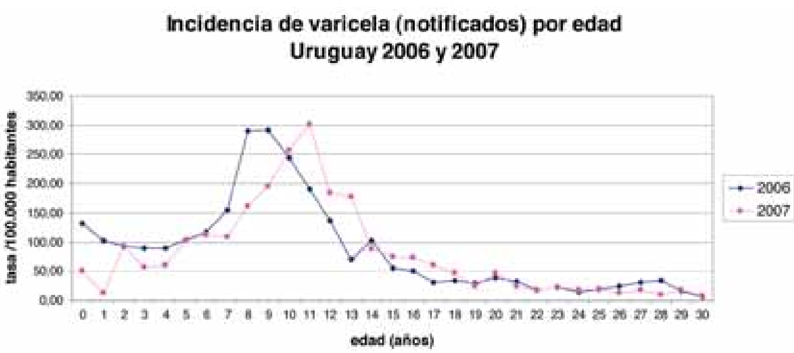
- ¿Qué podrías concluir observándola?
- ¿Por qué es importante la vacuna de la varicela?
- ¿En qué años se registraron más casos de varicela? ¿Y menos?
- Fíjate en el año que naciste; compara la cantidad de casos de varicela con el año anterior.

Fuente: MSP. http://www.msp.gub.uy

A partir de esta actividad tenemos la posibilidad de interpretar gráficos donde no conocemos la tabla que los generó y, por lo tanto, "no tenemos todos los datos". Sin embargo se podrán establecer algunas relaciones entre ellos, por ejemplo:

- consecuencias de la acción de la vacunación;
- comparar el comportamiento de la enfermedad después de la vacunación;
- comparar algunos años del desarrollo de la varicela;
- establecer la medida de la moda;
- analizar a partir del gráfico sin calcular la dispersión media que hay en cuanto al desarrollo de la enfermedad estudiada;
- habilitar la lectura de la "estadística en nuestra vida cotidiana".

Esta gráfica es del mismo informe. Analiza la incidencia de la varicela en los años 2006 y 2007. ¿Cuáles son las edades en que la varicela es más frecuente? Los abuelos, ¿tienen varicela? Y si preguntamos en cada clase de la escuela: ¿quiénes ya tuvieron varicela?, ¿cómo graficaríamos los datos recogidos?



Esta parte de la actividad establece sobre el mismo tema, "el estudio de la varicela en Uruguay", el foco en otras relaciones. La gráfica muestra el comportamiento de la varicela en función de las edades. Aparecen dos curvas que se pueden comparar, la de referencia de 2006 y la de 2007.

La actividad exige responder una serie de preguntas que relacionan la frecuencia de la enfermedad con las edades. Esto nos ayuda a establecer la relación edad-varicela.

Se podrá identificar nuevamente la moda y la dispersión, en qué edades se concentra la enfermedad y establecer algunas vinculaciones con otras áreas del conocimiento como salud y educación social.

Además, con respecto a la primera parte de la actividad, se pide que se confeccione una gráfica a partir de recoger datos en la escuela y discutir qué tipo de gráfico será más conveniente y por qué, de acuerdo a lo que se quiera comunicar. Esto exigirá recoger los datos, tabularlos, realizar algunos cálculos, decidir qué gráfica hacer, establecer relaciones con los datos proporcionados por el MSP en relación a la población de la escuela.

¿Quién gana las elecciones? La importancia de las muestras representativas

No sabemos.

Pero Factum, un Centro de Investigaciones Sociales de nuestro país (www.factum.edu.uy), le pregunta a 933 personas mayores de 18 años (¡nada más!) de todo el país (urbano y rural) y lo averigua.

Realizar algunas reflexiones sobre la relación entre las muestras representativas da insumos para establecer las vinculaciones con los conceptos estadísticos trabajados. Este es un ejemplo de que no se necesita realizar un censo (o las elecciones, en este caso) para anticipar, con poco margen de error, la opinión de los uruguayos sobre este tema. Así funciona el trabajo de la estadística inferencial.

Los uruguayos, ¿comemos mucha carne?

En el diario *La República* del 1º de julio de 2007 se puede leer que, en el año 2006, el consumo promedio de carne bovina fue de 52,8 kg por persona. ¿Vos dirías que cada uruguayo comió 1,100 kg de carne por semana?

Nuevamente, poner en funcionamiento los conceptos que nos brinda el soporte estadístico para establecer la lógica de esa afirmación y poder decidir que no necesariamente el promedio nos da lo que nos pasa personalmente, pero sí una idea del comportamiento frente al consumo de carne de la población uruguaya. En la actividad anterior, la media o promedio por persona sería $\bar{X}$ = 1,100 kg. Este dato explica la conducta colectiva, pero no las individuales. Por eso es necesario analizar los límites de cada parámetro, como hemos intentado en el artículo.

De igual manera, al analizar la actividad referida a la vacuna de varicela y relacionarla en un plano personal, individual, surge la idea de inscribir los registros personales. Inscribir los registros personales es ver si estamos en la media o somos parte del desvío. Las estadísticas en el fondo, las sociológicas, trabajan analizando esos dos planos: el general o estructural, y el personal. En algún punto se tocan: ¿en cuál? Para responder esta pregunta es que damos estos elementos, para *"pensar estadísticamente"* y para *"pensarnos estadísticamente"*.

En síntesis

Analizar las situaciones anteriores, constatar la veracidad o no de las informaciones brindadas, reflexionar sobre los cálculos realizados en relación al problema estudiado, trabajar con diferentes representaciones de la misma información, son elementos imprescindibles a considerar para el trabajo con Estadística. La idea es poder manejar algunos de los conceptos básicos de Estadística para lograr que en las aulas se puedan establecer estas relaciones y se logre un manejo con sentido de estos conceptos.

Bibliografía consultada

ABDALA, Carlos; GARAVENTA, Luis; REAL, Mónica (2007): *Carpeta de Matemática 2*. Buenos Aires: Aique Grupo Editor.

BRESSAN, Ana P. de; BRESSAN, Oscar (2008): *Probabilidad y Estadística: cómo trabajar con niños y jóvenes*. Buenos Aires: Ed. Novedades Educativas. Colección Biblioteca Didáctica.

COLERA, José; DE GUZMÁN, Miguel (1994): "Bloque 4. Estadística, Azar y Combinatoria" en M. de Guzmán; J. Colera; A. Salvador: *Matemáticas. Bachillerato 1*. Madrid: Ed. Anaya.

RAYNER, David; CABRILLO, Ezequiel (1998): *Estadística y probabilidad: Secundaria*. Madrid: Oxford University Press. Biblioteca de Recursos Oxford Educación.

SPIEGEL, Murray R. (1993): *Estadística*. México: Mc Graw Hill.

VÁZQUEZ DE TAPIA, Nelly; TAPIA DE BIBILONI, Alicia; TAPIA, Carlos Alberto (1984): *Matemática 4,* Cap. 14. Buenos Aires: Editorial Estrada.

Revista:

UNO. Didáctica de las Matemáticas, Nº 5 (Setiembre 1995): "Probabilidad y Estadística". Barcelona: Ed. Graó.

Sitios visitados:

www.factum.edu.uy

http://www.msp.gub.uy

www.tenfieldigital.com.uy

http://es.wikipedia.org/wiki/Wikipedia