

Unidad 1. Tema 1.2 Representación estadística.

BIBLIOGRAFÍA BÁSICA

Existen dos tipos de presentación estadística: la tabular y la gráfica.

PRESENTACIÓN TABULAR.

La finalidad de los cuadros es presentar de forma resumida e inteligible determinado material numérico.

Las tablas cumplen generalmente uno de los siguientes propósitos:

- presentar una o más distribuciones de frecuencias tanto de variables cualitativas como cuantitativas.
- Presentar índices asociados a determinadas variables
- Presentar una serie cronológica

Las tablas pueden perseguir más de uno de estos propósitos ya que estos no son excluyentes.

Partes de un cuadro:

- Identificación de la tabla. Consiste en darle un número consecutivo según el orden en que aparezca en el texto del trabajo: “cuadro1, cuadro2...” y a la hora de hacer referencia al mismo se hace por el número asignado, “como puede apreciarse en el cuadro1....”
- Título. Es la parte más importante y difícil de diseñar correctamente. El cuadro o tabla estadística debe ser auto explicativo, debe entenderse sin necesidad de acudir al texto, (si esto no se cumple su diseño no fue correcto). La mejor manera de saber si una tabla tiene errores en su confección es dársela a leer a un colega, si comienza a hacernos preguntas por que no es capaz de entenderla algo no ha quedado claro o bien diseñado. Es por eso que el título juega un papel tan importante, nos brinda la mayor cantidad de información sobre el material numérico que aparece en el cuerpo de la tabla.

Para que un título esté bien elaborado debe cumplir con dos requisitos básicos:

Que sea **completo**. Para que cumpla con este primer requisito debe responder las siguientes preguntas:

¿Qué?, se refiere a los elementos que conforman la muestra o la población. ¿Qué se estudió?, defunciones o nacimientos o niños o camas de un hospital...

¿Cómo?, se estudió esa información. Se refiere a las variables por las que se clasificó la misma. ¿Cómo se estudió...? por sexo, por sexo y edad, por peso al nacer....

¿Dónde? se realizó el estudio, a que "espacio" pertenecen esos datos, Ciudad de la Habana, Cuba, Hospital x, Municipio...

¿Cuándo?, se refiere a la fecha o periodo: año 2000, primer semestre del año, mes de enero, de 1990 a 2000...

Si nos hacemos esas preguntas en ese mismo orden y las vamos contestando tendremos elaborado un título completo. El segundo requisito que debe cumplir el título es que sea **conciso**, que sea **breve**, pero nunca debe dejar de ser completo en aras de lograr una mayor concisión.

Ejemplo

Título 1: Distribución de las defunciones de Cuba ocurridas durante el año 1999 clasificadas según sexo y edad de los fallecidos.

Veamos primero si el título es completo:

¿Qué se estudia?, defunciones

¿Cómo?, por edad y sexo

¿Dónde?, Cuba

¿Cuándo?, 1999

Obviamente el título es completo, responde las 4 preguntas, pero es innecesariamente largo, si nos limitamos a dar respuesta a las preguntas quedará además de completo conciso que era la segunda característica de un buen título:

Defunciones por edad y sexo. Cuba 1999. (Correcto)

Cuerpo de la tabla. Está formado por la unión de filas y columnas, las filas en sentido horizontal y las columnas en sentido vertical.

La primera columna se denomina columna matriz y en ella se coloca la variable con su escala. En la primera fila, denominada “de los encabezamientos” se rotulan los aspectos que denominan los números que aparecen el resto del cuerpo, nombre de los índices, número, % etc.

La última fila y la última columna son las de los totales

Nombre de la variable	FILA DE LOS	ENCABEZAMIENTOS	TOTAL
E			
S			
C			
A			
L			
A			
TOTAL			

Notas aclaratorias

Fuente

Notas aclaratorias. Se acostumbra a colocarlas al pie de la tabla y tienen la finalidad de esclarecer algunos aspectos importantes para la comprensión de la misma. Puede ser las unidades de medidas de algunas variables o el 10” con que se trabajó alguna tasa, por citar algunos ejemplos.

Fuente. Se refiere al lugar del cual se obtuvo la información, casi siempre hace alusión a una fuente secundaria, un sistema de información, una publicación, un documento oficial etc. El declarar la fuente de donde se obtuvieron los datos obedece a varios motivos, en primer lugar, el lector debe saber cuán confiable es esa información, además puede acceder a la misma si así lo quisiera, y también es un reconocimiento a las personas o instituciones que obtuvieron esa información.

En ocasiones la fuente de la cual se obtuvo la información es una fuente primaria, digamos una encuesta. Si todas las tablas tienen como única fuente esa encuesta, fruto del trabajo del propio investigador, y quedó bien declarada en el método, es innecesario poner en todas las tablas “fuente:

encuesta” como suele ocurrir con mucha frecuencia, es redundante, monótono, y le resta elegancia al trabajo.

Errores más frecuentes al diseñar una tabla

- Título o encabezamiento incorrecto o inadecuado.
- Disposición incorrecta de los datos.
- Errores de cálculo.
- Tablas que muestran solo %
- Cuadros sobrecargados
- No poner la fuente en casos necesarios, o ponerla cuando los datos son originales

Algoritmo para el análisis de una tabla.

Al escribir el informe final de una investigación, un artículo científico o simplemente un informe de trabajo donde aparezca determinado material numérico presentado en un cuadro, este debe ser analizado por el autor y escribir en el texto sólo los aspectos más relevantes, los juicios de valor que se desprenden de ese análisis, las conclusiones a las que se arriba después de realizado el mismo.

Sin embargo, en un número no despreciable de casos nos encontramos que se hace una descripción detallada de la tabla, se describen celda por celda todos los valores que aparecen en las mismas, como nos gusta decir a modo de ilustración, “se hace un retrato escrito de la tabla” y no se da ni un solo juicio de valor.

Para evitar este mal hábito le sugerimos que siga estos pasos para el análisis de la tabla y una vez concluido el mismo escriba en el texto las conclusiones a las que arribó, eso es lo que se espera de usted como autor, los datos numéricos los tiene el cuerpo de la tabla, no hay que repetirlos en forma de texto.

Algoritmo para el análisis de una tabla:

- Leer cuidadosamente el título.
- Leer las notas aclaratorias
- Precisar unidades de medidas
- Fijarse en los totales
- Relacionar los totales con los valores particulares de las categorías de las variables
- Relacionar entre sí los valores de las variables estudiadas

PRESENTACIÓN GRÁFICA.

El gráfico es la representación de los datos numéricos en el plano, con el fin de obtener una impresión visual de conjunto del material presentado, que facilite su rápida comprensión y que permita evidenciar de forma más objetiva que la tabla, la evolución de un fenómeno o hacer más evidentes ciertas relaciones poco aparente en el material tabulado.

El gráfico es un auxiliar de la tabla estadística, no la sustituye: la complementa; además no es necesario tener tantos gráficos como tablas aparecen en el informe final de un trabajo, por el contrario, le restan lucidez al trabajo lejos de proporcionar calidad o rigor científico.

Características generales de los gráficos

- Generalmente se inscriben en los llamados ejes de coordenadas cartesianos o rectangulares.
- En la ordenada (y) se representan las frecuencias o índices
- En la abscisa (x) se representan las clases de la escala de clasificación
- Los ejes deben ser de igual magnitud, solo las abscisas pueden ser 1.5 o 2 veces mayor que las ordenadas. Este aspecto es muy importante porque de no tenerse en cuenta pueden ocurrir las llamadas falacias o distorsiones visuales del fenómeno que se analiza. Si el tamaño del eje de las "y" es más del doble dará la impresión de cambios marcados en el comportamiento del fenómeno, si por el contrario es menor que el eje de las "x" parecerá que apenas hay modificaciones en el comportamiento de las variables estudiadas

- De ser posibles las escalas deben comenzar en cero. En caso de valores muy altos y con pocas oscilaciones puede “partirse la escala”.
- Los ejes de las ordenadas y de las abscisas deben estar rotulados.
- Las escalas no deben tener muchas subdivisiones, deben usarse siempre que sea posible, números redondos. El detalle se encuentra en la tabla, el gráfico da la idea general.

Partes de un gráfico

- Identificación. Al igual que en la tabla su objetivo es numerarlos, darles un número consecutivo si es más de uno el que aparece en el trabajo.
- Título. Debe cumplir los mismos requisitos que el de las tablas.
- Gráfico propiamente dicho. Su forma y características dependen del tipo de variables como ya veremos más adelante.
- Leyenda. Se utiliza cuando por razones de espacio o estética se desea identificar colateralmente alguna parte del gráfico.
- Notas aclaratorias. Tienen el mismo objetivo que en la tabla.
- Fuente. La fuente de todo gráfico es la tabla estadística que le dio origen. Una tabla en ocasiones puede ser fuente de varios gráficos.

Principales tipos de gráficos

1- Diagramas de barras. Los diagramas de barras son los gráficos de elección para representar distribuciones de frecuencias de variables en escalas **cuantitativas y cuantitativas discretas**. El tipo gráfico de barras que se utilice dependerá del número de variables a graficar.

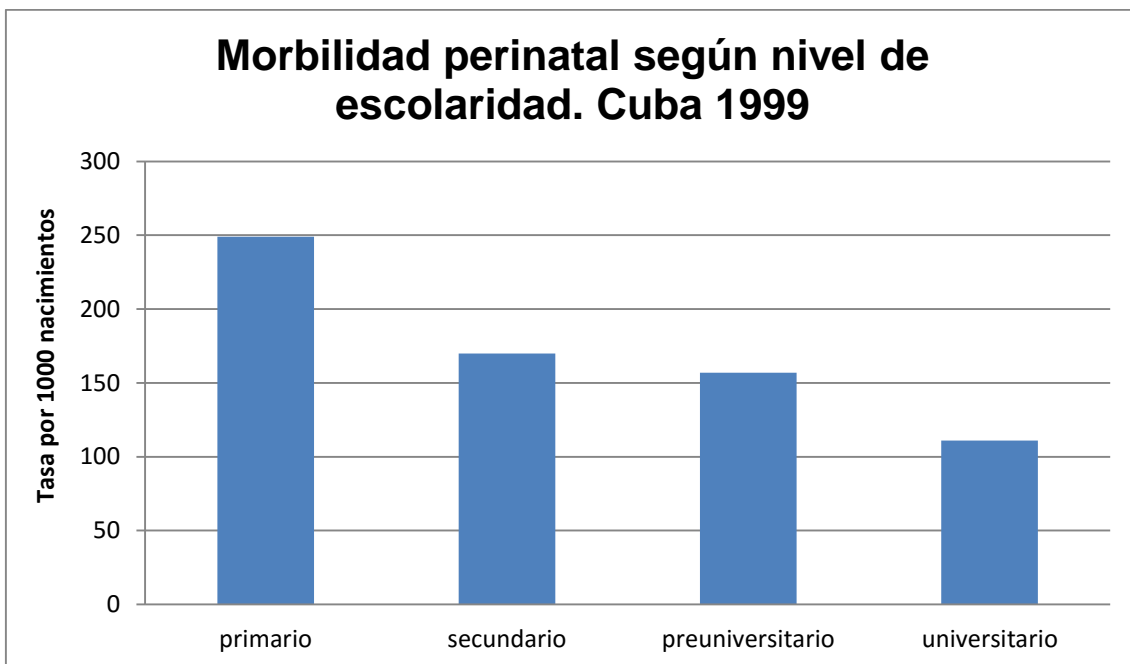
a) Barras simples: Se utilizan para representar **una sola variable en escala cualitativa o cuantitativa discreta**.

Como en todo gráfico en el eje de las “y” se inscribe la frecuencia absoluta o relativa o el índice utilizado para el análisis de la información y en el eje de las “x” la variable con su escala. La frecuencia es representada por la altura de la barra, todas las barras tienen el mismo ancho, a libre elección, y la separación entre barras es también arbitraria pero la misma para todas; con frecuencia se toma el mismo ancho de la barra o la mitad del ancho como medida de separación.

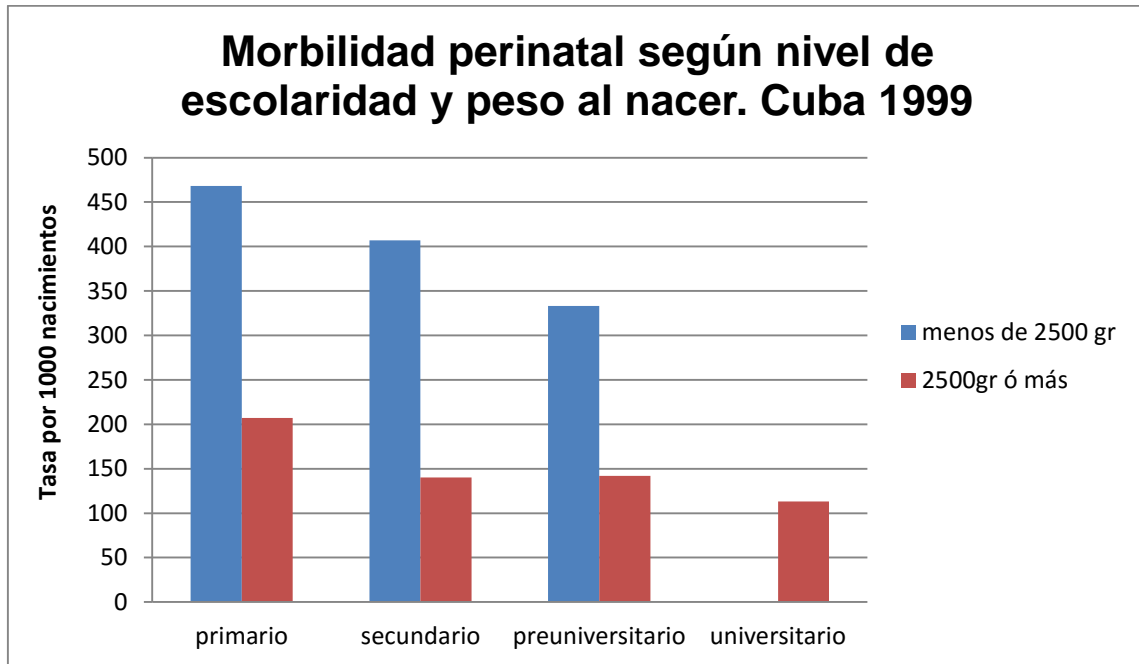
Cada barra se rotula con el nombre de la clase que representa. Cuando el número de categorías de la escala es muy numerosa se hace difícil el rotulado de las barras y en estos casos se acostumbra a “rotar los ejes”, se colocan las barras en posición vertical para facilitar la identificación de las barras.

Si los datos que se representan no están en una secuencia que deba conservarse, es aconsejable ordenar las barras de mayor o menor o viceversa, con lo que se mejora la comprensión y el diseño.

Ejemplo de gráfico de barras simples.

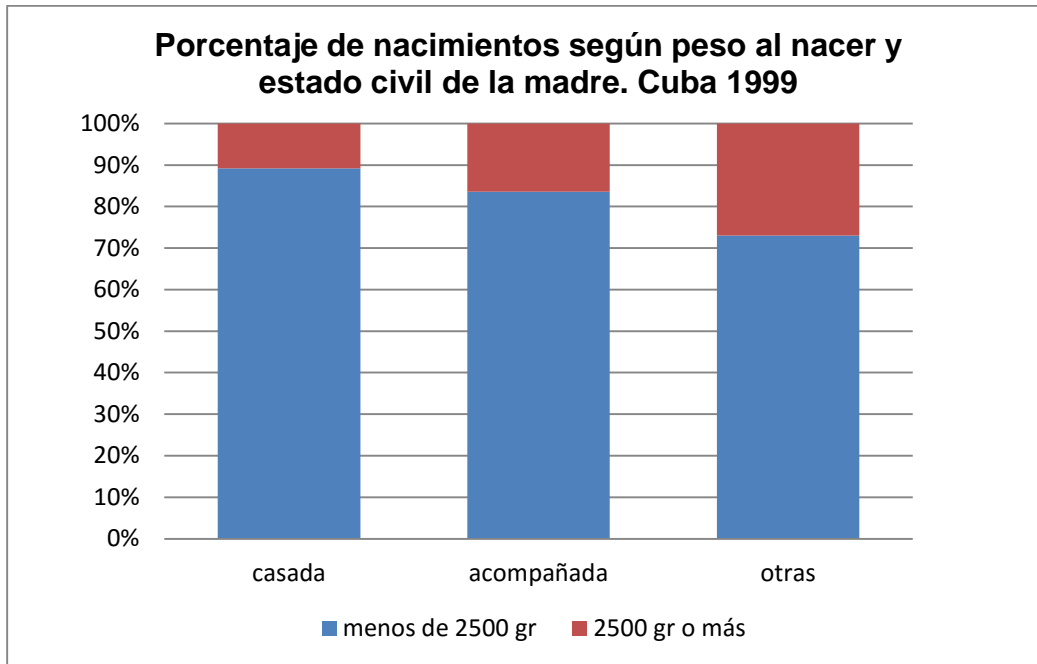


- b) Barras múltiples: Se usan para representar **dos variables en escala cualitativa o cuantitativa discreta**. Las barras pueden ser dobles, triples etc., depende del número de clases o categorías de la escala de una de las variables y el número de grupos de barras estará dado a su vez por el número de clases de la escala de la otra variable. Se sugiere que la variable que más interés se tenga en resaltar será la que formará los grupos de barras.



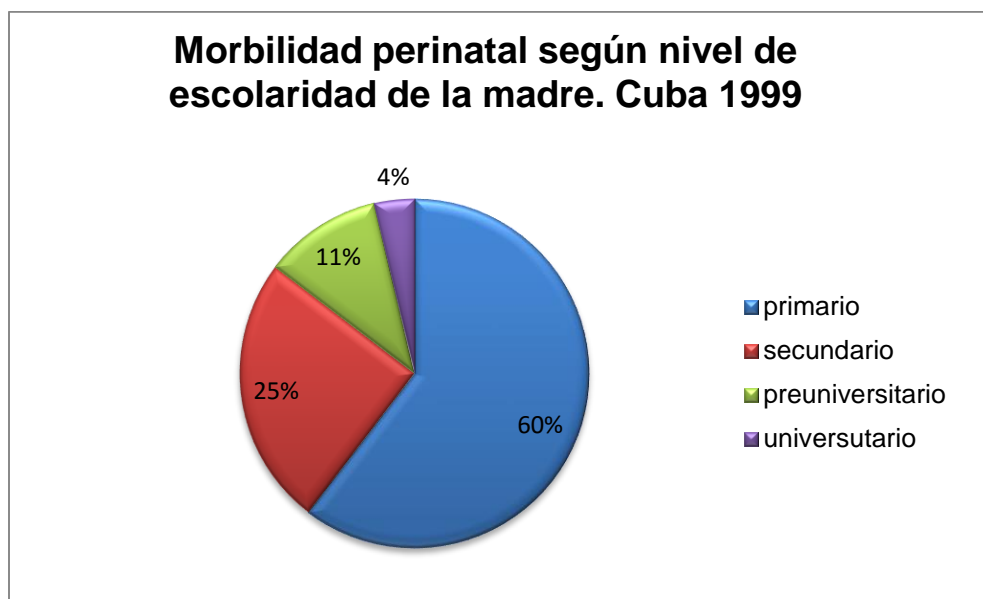
Observe en el ejemplo, que se trata de dos variables, una en escala cualitativa ordinal, la escolaridad y la otra aunque por su naturaleza es continua se trabajó con una escala discreta, menos de 2500gr y 2500gr o más. Como la variable que más se quería destacar era el peso de los recién nacidos es la que forma los grupos de barras y al ser la escala dicotómica da origen a un gráfico de barras dobles. Al tener la escolaridad 4 clases en la escala se formaron 4 grupos de barras dobles. Si nuestro interés hubiese sido poner como variables principal la escolaridad tendríamos entonces dos grupos de barras cuádruples.

- c) Barras compuestas: Al igual que las barras múltiples se usan para variables en escala cualitativa o cuantitativa discreta, por lo que deviene en una alternativa para variar el diseño.



Sólo tiene que cumplirse un requisito adicional y es que las distribuciones de frecuencias tienen que ser relativas, expresadas en %. La barra en su totalidad representa el 100% y cada parte en que se divide la barra el aporte que hace la otra variable.

2- Gráfico circular o de pastel. Se usa para representar la distribución de **una sola variable expresada en escala cualitativa o cuantitativa discreta**, por lo que constituye una alternativa de presentación de las barras simple, pero al igual que las barras compuestas la distribución tiene que estar en %



Cuando la variable tiene un número considerable de categorías en la escala no es aconsejable utilizar este tipo de gráfico porque tendría tantas subdivisiones que obstaculizarían su comprensión, en estos casos el gráfico de barras simples es el de elección. Esta misma situación se daría en las barras compuestas y se prefiere en este caso las barras múltiples.

3- Histograma. Hasta ahora los gráficos que hemos visto se emplean para representar información referente a variables expresadas en escala cualitativa o cuantitativa. Las variables **cuantitativas continuas** también tienen sus gráficos para ser representadas, y es el caso del histograma que se utiliza cuando queremos graficar **una sola variable continua**.

No es precisamente este gráfico el que más se usa, sino, el polígono de frecuencias como veremos más adelante. Si nos detenemos a dar una explicación exhaustiva del mismo, es porque nos ayudará mucho a entender las distribuciones teóricas de frecuencias en la parte inferencial. En el eje de las **x** se inscriben los límites inferiores de los intervalos de clase y en el de las **y** la frecuencia absoluta o relativa. La figura que se usa en la presentación es también el rectángulo, como en los gráficos de barras. Cuando todos los intervalos de clase tienen la misma amplitud, simplemente la altura de la barra dependerá de la frecuencia del intervalo, el ancho a criterio del diseñador y todas las barras tendrán el mismo ancho, pero para dar el sentido de continuidad que tiene la variable las barras no se separan, sino que van unidas.

Sin embargo, basta con que un solo intervalo de clase tenga una amplitud diferente al resto para que no podamos asumir el diseño del histograma como lo vimos anteriormente, hay que “promediar por cada unidad de escala” y veamos qué quiere decir esto con un ejemplo.

Defunciones por accidentes en menores de 40 años según edad. Matanzas 2000

Edad en años	Número
<1	10
1-4	80
5-9	125
10-19	240

20-29	50
30-39	10

Si realizáramos el gráfico dándole la altura al rectángulo en dependencia del valor de la frecuencia, el grupo de edad de 20 a 29 años sería el mayor y por ende, diríamos que los accidentes son más frecuentes en ese grupo de edad.

Pero ¿cuántos años contempla ese grupo de edad? ¿tienen todos los grupos la misma amplitud y por consiguiente son esos grupos comparables?

Evidentemente no, las amplitudes cambian, los tres últimos intervalos tiene 10 de amplitud, abarcan 10 edades, mientras que los primeros intervalos tienen 1, 4 y 5 respectivamente. Es lógico que los intervalos de mayor amplitud, que contemplan un mayor número de edades, que abarcan una mayor parte de la población tengan más fallecidos.

Para hacer esos grupos comparables se divide la frecuencia de cada intervalo entre la amplitud correspondiente y así obtenemos las defunciones promedio por año (promedio por unidad de escala) y este resultado es lo que nos da la altura de la barra.

Otro aspecto a tener en cuenta que las barras no pueden tener el mismo ancho, el ancho de la barra dependerá de la amplitud de la clase.

Es por ello que se dice que el histograma es un diagrama de superficie, o sea que no es la altura de la barra la que representa la frecuencia, sino, que la frecuencia de la clase está representada por toda el área de la barra.

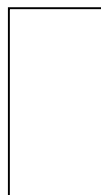
Si recordamos de la enseñanza media la fórmula de cálculo del área de un rectángulo está es:

$$A = b \times h$$

Donde:

A= área del rectángulo

b= base del rectángulo



h= altura del rectángulo

h

b

Tenemos el área, que es la frecuencia del intervalo, tenemos la base que es la amplitud, lo que no conocemos es que altura darle a cada rectángulo para que se cumpla que toda el área represente la frecuencia.

Si despejamos de la fórmula tenemos:

$$h = A/b$$

A= frecuencia

b= amplitud

Sustituyendo:

$$h = \text{frecuencia} / \text{amplitud}$$

Podríamos pensar, que, en buena ley, siempre se debe dividir la frecuencia entre la amplitud para tener una altura que asegure que toda el área sea la frecuencia. Teóricamente debe ser así, lo que ocurre es que al tener todos los intervalos igual amplitud en la práctica este es un paso innecesario, porque al dividir por el mismo valor todas las frecuencias la diferencia entre la altura de las barras es la misma.

Si se sumaran todas las áreas de los rectángulos, esta suma es igual al total de casos, si se trata de la frecuencia absoluta. Si se trabajó con la frecuencia relativa, la suma de todas las áreas de los rectángulos tiene que ser igual a la unidad, si se trata de proporciones o a 100 si se trabajó con porcentajes.

4-Polígono de frecuencia. Se utiliza más que el histograma y permite no sólo representar la distribución de una variable continua, sino, que posibilita tener más de una distribución en el mismo gráfico, o sea, permite tener dos variables, una en escala cuantitativa continua- que es la variable principal y otra en escala cualitativa, casi siempre, aunque puede ser discreta y valorar comparativamente los cambios en la distribución de la variable principal.

Por ejemplo: el peso por sexo, para hombres y mujeres en un mismo gráfico porque en este caso la forma de representación es una curva, o como se suele decir es un gráfico de tipo lineal.

El polígono se construye exactamente igual que el histograma, lo único que cambia es la forma de representación, en lugar de dibujar un rectángulo a la misma altura que tendría la barra se ubica un "punto" en el centro del intervalo de clase, a nivel de la marca de clase, en el punto medio del intervalo, luego esos puntos se unen y queda una línea curva que sustituye a los rectángulos.

Como el polígono es también un diagrama de superficie, el área bajo la curva es igual al tamaño de la muestra o de la población, al total de observaciones, si se trabajó con frecuencias absolutas, y si fue con frecuencia relativas, el área bajo esa curva es equivalente a 1 si se usó la proporción o a 100 si fue el porcentaje

5- Gráfico Aritmético Simple (GAS). Es el diagrama de elección cuando se quiere representar series de tiempo. Su empleo rebasa los límites de la representación para convertirse en algunos casos en técnicas de análisis como ocurre con los canales endémicos.

En el eje de las "x" aparece el tiempo expresado en años, meses, semanas, días u horas y en el eje de las "y" la frecuencia o el índice calculado para el análisis de la información. El valor de la frecuencia para cada unidad de tiempo queda representado por un punto, que posteriormente se unen para dar lugar a una curva que representa el comportamiento del fenómeno estudiado.

