

## **Descriptive Statistics IV: Presentation of data I (Tables, histograms, polygons and ogives).**

## **Estadística Descriptiva IV: Presentación de datos I (Tablas, histogramas, polígonos y ojivas).**

**Juan Luis Soto Espinosa**<sup>1</sup>  <https://orcid.org/0000-0003-2600-9292>

<sup>1</sup> *Facultad de Estudios Superiores Zaragoza*

*Dirección (autor principal): Batalla 5 de Mayo s.n. esq. Fuerte de Loreto., Col. Ejército de Oriente, Ciudad de México.*

*Correo electrónico de contacto: soej@unam.mx*

**Palabras clave:** *Estadística, tablas de frecuencia, tablas de contingencia, histograma.*

Fecha de envío: 01/09/2021

Fecha de aprobación: 05/11/2021

Continuando con estadística descriptiva, en este número se abordará la forma ordenada de presentar los datos obtenidos como parte de un estudio o investigación de tipo estadístico.

La presentación de datos es un aspecto fundamental en estadística descriptiva pues es a través de estas herramientas que podemos mostrar cómo se presentan los valores de las variables bajo estudio dentro de los elementos de la población o muestra bajo análisis.

Para una correcta presentación ordenada de datos estadísticos, existen diferentes herramientas y técnicas, cuya utilidad y enfoque dependen del tipo de información con que se cuente y del mensaje que se desee transmitir a los lectores del trabajo.

La presentación correcta de los datos estadísticos empieza desde el texto de introducción al trabajo que se presenta; es necesario proporcionar información general acerca del tipo de datos que se presenta, los alcances del trabajo y brindar un contexto claro acerca del enfoque y ámbito de conocimiento que se dará al tratamiento de la información

Un documento que trate sobre un análisis estadístico va más allá de la mera enumeración y listado de los datos encontrados, debe proporcionar un marco conceptual que permita entender la importancia del estudio y los alcances que se pretenden en el mismo.

Para que un dato estadístico tenga sentido, debe presentarse dentro de un contexto, de forma ordenada y organizada junto con el resto de los datos producto de la investigación.

Dentro de las herramientas más utilizadas para la presentación de datos estadísticos están las TABLAS y las GRÁFICAS.

Las TABLAS Y LAS GRÁFICAS constituyen dos formas equivalentes de presentar conjuntos de datos de forma ordenada, abreviada y facilitan su comprensión.

A continuación, procederemos a la definición de cada una de ellas, así como a revisar sus características, utilidad, ventajas y desventajas, así como los casos en los que su uso se recomienda.

*Documentos Educativos*

## Tablas simples

Se les denomina TABLAS SIMPLES a los arreglos de información organizadas en filas y columnas en las que se presentan datos de una sola variable, a estas tablas también se les conoce como TABLAS DE FRECUENCIA, cuya estructura se revisó en números anteriores de RIST, pero que, con motivos didácticos, a continuación, se presentarán los elementos que las constituyen.

Una TABLA DE FRECUENCIA presenta la información organizada en columnas, como se muestra a continuación:

Nivel de estudios	Frecuencia (Número de veces que un valor se repite en el conjunto de datos)
Primaria	Frecuencia 1
Secundaria	Frecuencia 2
Bachillerato	Frecuencia 3
Licenciatura	Frecuencia 4

Dónde:

- Variable
- Frecuencias
- Clases o categorías

Una tabla de frecuencia típica, presenta FRECUENCIA ABSOLUTA, FRECUENCIA RELATIVA, FRECUENCIA ACUMULADA Y FRECUENCIA RELATIVA ACUMULADA, un ejemplo se muestra en la tabla siguiente:

## Coronavirus: Casos presentados por mes en la Zona Metropolitana de la Ciudad de México

Mes	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
ene-21	296,328	28.46%	296,328	28.46%
feb-21	208,885	20.06%	505,213	48.52%
mar-21	151,977	14.59%	657,190	63.11%
ago-21	129,619	12.45%	786,809	75.56%
abr-21	90,793	8.72%	877,602	84.28%
jul-21	74,630	7.17%	952,232	91.45%
may-21	44,256	4.25%	996,488	95.70%
jun-21	25,989	2.50%	1,022,477	98.20%
sep-21	18,877	1.80%	1,041,354	100.00%
<b>Frecuencia total</b>	1,041,354	100.00%		

Como se ha comentado en entregas anteriores, la columna de FRECUENCIA ABSOLUTA muestra el número de veces que un valor se repite dentro del conjunto de datos; la FRECUENCIA RELATIVA muestra las veces que se repite un valor expresado en porcentaje (296, 328 casos, representan el 28.46 % del total). La columna de FRECUENCIA ACUMULADA muestra el valor de la clase actual, sumado a las clases anteriores, de esta forma la FRECUENCIA ABSOLUTA ACUMULADA de la primera clase es 296,328, mientras que el de la segunda es 505,213 (resultado de sumar 296,328 + 208,885), el de la tercera es 657,190 ( 296,328 + 208,885+151,977) y así sucesivamente.

La cuarta columna, de manera análoga, muestra el valor porcentual de cada clase, sumado a los de las clases anteriores, así pues, tenemos que la FRECUENCIA RELATIVA ACUMULADA de la primera clase es 28.46 %, el de la segunda es 48.52 % (28.46 % + 20.06 %), etc. Note que los valores de FRECUENCIA ABSOLUTA ACUMULADA y de FRECUENCIA RELATIVA ACUMULADA corresponden al número total de datos y al 100.0 % respectivamente.

Documentos Educativos

## Tablas de contingencia

Al revisar la tabla simple del ejemplo anterior, se determina que el Valle de México incluye pacientes que fueron hospitalizados en la Ciudad de México y pacientes que fueron hospitalizados nel el estado de México, incluyendo esta variable en la tabla original, tenemos:

### Coronavirus: Hospitalizados totales en la Zona Metropolitana del Valle de México

Mes	Hospitalizados CdMx	Hospitalizados EdoMex	Hospitalizados Totales
Ene-2021	216576	79752	296328
Feb-2021	155073	53812	208885
Mar-2021	114477	37500	151977
Abr-2021	68147	22646	90793
May-2021	34049	10207	44256
Jun-2021	19714	6275	25989
Jul-2021	53456	21174	74630
Ago-2021	84448	45171	129619
Sep-2021	12369	6508	18877
<b>Frecuencia total</b>	1670635	623246	2294976

A diferencia de las TABLAS SIMPLES o DE FRECUENCIA, donde sólo se presenta información de una variable y sus respectivas categorías, la Tabla 4 organiza información correspondiente a dos variables con sus respectivas categorías, por lo que constituye una tabla de contingencia.

Las TABLAS DE CONTINGENCIA (también llamadas TABLAS DE DOBLE ENTRADA, TABLAS CRUZADAS O CROSS TABS) son herramientas que permiten observar la forma en que se asocian dos variables de tipo nominal, categórico u ordinal; constituyen una de las herramientas de análisis causal más utilizadas en ESTADÍSTICA INFERENCIAL, por lo que son de especial utilidad en el estudio de variables cualitativas.

Están organizadas por filas (horizontales) para la información de una variable y columnas (verticales) para la información de una segunda variable. El cruce da cada fila y columna define una celda, en cada celda se registran las frecuencias de los elementos que presentan cada combinación de variables analizadas, es decir, aquellos elementos o participantes que presentan los valores correspondientes a las dos variables bajo estudio.

En su forma más simple, las tablas de contingencia tienen solo 2 filas y 2 columnas (tablas de 2x2); en su estructura, estas tablas presentan la siguiente estructura tipo:

	Variable B		
Variable A	Valor 1	Valor 2	Total
Valor 1			
Valor 2			
Total			

Un ejemplo de una tabla de contingencia utilizada comúnmente en estudios de salud en el trabajo sería la siguiente:

	Enfermedad		
Factor de riesgo	Si	No	Total
Expuesto			
No expuesto			
Total			

En este ejemplo, las filas registran la exposición a un determinado factor de riesgo, mientras que las columnas registran la existencia de una determinada enfermedad

## Documentos Educativos

(incidencia o prevalencia, según el estudio del que se trate)<sup>1</sup>.

En general, una tabla de contingencia puede estar constituida por varias filas (M) y varias columnas (N), lo que da lugar a una tabla de MxN. Cuando una variable tiene dos valores (una variable de tipo dicotómico) y se cruzan sus valores con otra variable de tipo dicotómico, se obtiene una tabla de contingencia 2x2:

Obesidad	Sexo		Total
	Masculino	Femenino	
Si	45	20	65
No	10	30	40
Total	55	50	105

Note que los totales, tanto de filas como de columnas en el ejemplo, dan un total de 105, cantidad que se registra en la celda donde se cruzan ambos totales y se le conoce como GRAN TOTAL.

Otro ejemplo sería:

Tabaquismo	Enfisema		Total
	Si	No	
Si	28	11	39
No	3	9	12
Total	31	20	51

Las variables en una tabla de contingencia pueden tener más de dos valores, de forma que se pueden obtener tablas de 2x3, 3x3, 4x3, etc.

Un ejemplo de una tabla de estos tipos serían las siguientes:

<sup>1</sup> La **incidencia** de una enfermedad mide la velocidad a la que se producen casos nuevos durante un periodo de tiempo determinado en una población específica, mientras que la **prevalencia** es la frecuencia de casos de

### Tipo de tumor cerebral y localización

Localización	Tipo de tumor			Total
	Benigno	Maligno	Otros	
Lóbulo frontal	23	9	6	38
Lóbulo temporal	21	4	3	28
Otros	34	24	17	75
Total	78	37	26	141

Fuente: Davis, 1991

### Tratamiento clínico de un medicamento

Tratamiento	Mejora			Total
	Alta	Escasa	Ninguna	
Activo	16	5	6	27
Placebo	6	7	19	32
Total	22	12	25	59

El análisis del comportamiento de una variable con respecto de otra u otras es una tarea común en el ámbito de la Salud en el Trabajo, vinculado, la mayoría de las veces, a la búsqueda o identificación de un patrón de comportamiento que sugiera la relación o asociación, (o la falta de ellas) entre las variables analizadas.

Este es un proceso clave en la identificación de los posibles factores que provocan los problemas de salud, y también de aquellos factores que, aun cuando no puedan ser finalmente considerados causales, resulten estar asociados en cierta medida a estos daños o alteraciones de salud y que permitan la identificación de mayores riesgos de padecer un determinado trastorno de salud.

El estudio de la influencia de una variable (variable independiente) sobre la forma en que se modifica otra (variable dependiente) es conocido como análisis bivariado (considera el comportamiento, relación o asociación de dos variables); mientras que, si el estudio en cuestión involucra de manera simultánea el efecto de dos o más variables sobre la variable que

una enfermedad en una población y en un momento determinados.

## Documentos Educativos

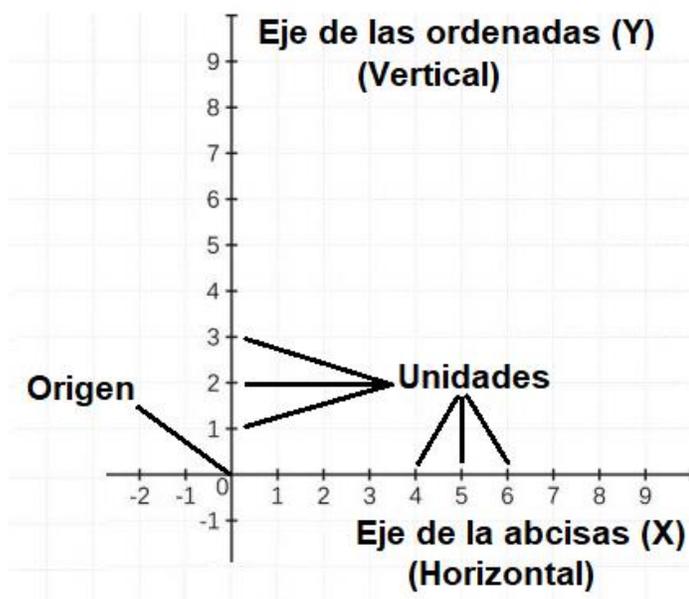
consideramos dependiente, el análisis se considera multivariado (una variable dependiente y dos o más variables independientes).

### Herramientas gráficas para el análisis de frecuencia

Antes de entrar de lleno con el tema de los gráficos, es necesario tener en cuenta una serie de conceptos que serán de uso cotidiano.

Para realizar representaciones gráficas de conjuntos de datos, se utiliza el PLANO CARTESIANO, que permite la representación de distribuciones de datos y funciones matemáticas en DOS dimensiones.

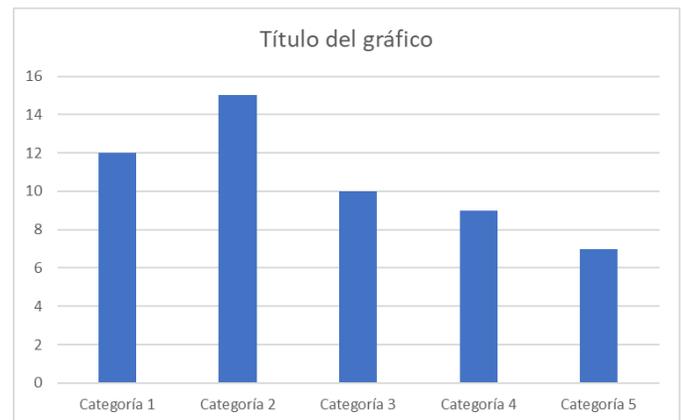
Recibe su nombre por René Descartes, que es el primero que lo utilizó y constituye El sistema de coordenadas de ubicación de puntos en el espacio más utilizados. Un plano cartesiano está conformado por dos rectas perpendiculares entre sí, una horizontal designada por la letra X conocida como eje de las ABCISAS y una línea vertical designada con la letra Y conocida como eje de las ORDENADAS. El plano cartesiano tiene la siguiente apariencia:



### Gráficos de barras.

Un gráfico de barras es una herramienta que permite representar el comportamiento de variables de tipo cualitativo ordinal o de variables cuantitativas discretas.

En el eje horizontal X se representan las diferentes categorías y sobre él se levantan unas columnas o barras cuya altura es proporcional a la frecuencia de cada categoría, la cual aparece representada en el eje vertical Y. Cabe señalar que los diagramas de este tipo NO se emplean para variables cualitativas nominales.



El diagrama de barras es especialmente útil para expresar y mostrar las diferencias de valor entre las modalidades que presenta la variable.

### Histograma

Cuando se elabora una representación gráfica de variables cuantitativas discretas, es posible utilizar un gráfico de barras; pero cuando las variables que se desea representar son variables continuas, en las que entre dos valores puede existir un número infinito de fracciones de los mismos, se debe utilizar un tipo particular de gráfico de barras: el histograma.

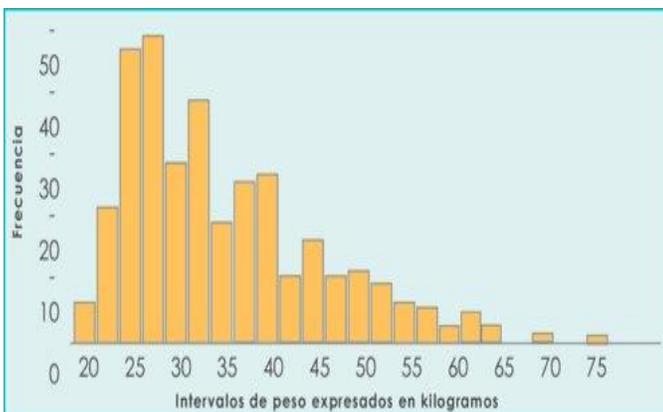
El histograma es un gráfico con un significado mucho más profundo que el gráfico de barras. Un histograma se utiliza para representar la frecuencia de variables

## Documentos Educativos

cuantitativas continuas. Considerando lo anterior, no es la altura sino el área de la barra lo que es proporcional a la frecuencia de cada conjunto de datos. Estos conjuntos de datos que dividen de manera homogénea la distribución de valores se conocen como INTERVALO. A diferencia del diagrama de barras, todas las áreas están juntas y el punto medio es el que da el nombre al intervalo. Los intervalos no tienen por qué ser todos iguales (aunque es lo más habitual), pero siempre tendrán un área mayor aquellos intervalos con mayor frecuencia.

Ejemplo:

Histograma de peso de niños de 5 a 11 años



Es importante señalar que mientras el diagrama de barras representa los valores de las variables que se han medido durante el estudio, el histograma es una herramienta de análisis que va mucho más allá, ya que representa todos los valores posibles que existen dentro de los intervalos, aunque no hayan sido observados o medidos de forma directa. Es por esta característica que los histogramas permiten calcular la frecuencia esperada de un valor cualquiera de la distribución, lo que es de gran importancia si se requiere hacer inferencia y estimar valores de la población a partir de los resultados de una muestra.

## Pasos para elaborar un histograma

**Paso 1:** Realizar el acopio (colecta) de datos que se van a analizar, ya sea a través de cuestionarios, mediciones directas o consulta de otras fuentes (bases de datos, historias clínicas, estudios previos, etc.).

**Paso 2:** Determinar el rango. Recordemos que el rango se define como la resta entre el valor más alto con el valor más bajo de los existentes en el conjunto de datos obtenido.

**Paso 3:** Determinar el número de clases que ha de considerar el histograma. Una CLASE se define como un subconjunto de elementos (generalmente del mismo tamaño) en los que se dividen los datos ordenados provenientes de la población o muestra y que presentan características comunes.

El número de clases se identifica con la letra K; existen diversas formas para definir cuántas clases se pueden considerar en la elaboración de un histograma; recordemos tres de ellas.

**Primera:** Considerar una tabla guía. Diversos autores han propuesto tablas para la elaboración de histogramas y selección de números de clases, por ejemplo, la propuesta por Roberto Behar y Pere Grima, la cual propone:

Cantidad de datos	Número de clases
20-50	7
50-75	10
75-100	12
Más de 100	15

Otros autores sugieren 4 clases si tenemos entre 20 y 50 datos, 7 clases si tenemos entre 50 y 100 datos, 10 clases para más de 100 pero menos de 150 datos, 12 clases para más de 150 y menos de 200 datos y 14 clases para más de 200 datos.

*Documentos Educativos*

**Segunda:** En ocasiones se recomienda determinar el número de clases a través de obtener la raíz cuadrada de la cantidad de datos. El resultado redondeado será el número de clases. La fórmula para resolver será:

$$K = \sqrt{\text{número de datos } (N)}$$

Dónde:

K = Número de clases

N = Número de datos

**Tercera:** La opción matemáticamente más consistente es la conocida como regla de Sturges, propuesta en el año de 1926 por el matemático Hebert Sturges. La solución de esta ecuación nos proporciona una regla práctica para obtener el número de clases:

$$K = 1 + 3.322\log(N)$$

Dónde

K = Número de clases

Log(N) = Logaritmo del número de datos

N = Número de datos

Cabe señalar que existen otras formas de determinar el número de clases a utilizar, algunas más complejas, otras más simples. Independientemente de la forma de cálculo seleccionada, lo realmente importante es la forma que va a tomar el histograma y que la información que en él se refleja permita comprender cómo se presentan los datos en la población.

**Paso 4:** Determinar la amplitud de clase o ancho del intervalo. Se define como intervalo la distancia que existe entre los límites superior e inferior de una clase; se identifica con la con la letra h.

Para determinar la amplitud de clase, debemos dividir el RANGO entre el número de clases (K) definido en

el paso anterior. Si se obtiene un número decimal, se debe redondear redondea al entero superior más cercano.

$$h = \frac{\text{Rango}}{k}$$

**Paso 5:** Definir las clases. Una vez que se tiene el Rango, el número de clases(K), el límite inferior de clase y la amplitud (h) se procede a determinar las clases.

Se toma el valor más pequeño de la distribución y se le suma la amplitud de clase (h). Se repite la operación tantas veces como número de clases se tenga.

**Paso 6:** Elaborar una tabla de frecuencia con base en los intervalos de clase definidos, esto es, se deben agrupar los datos en la clase que les corresponda de acuerdo con su valor. Esta operación dará como resultado la frecuencia de cada clase.

**Paso 7:** Construir el histograma:

En el eje x ubica los intervalos de clase.

En el eje y ubica la frecuencia.

Traza un rectángulo cuyo lado izquierdo sea el límite inferior de clase, el lado derecho el límite superior de clase y la altura esté dada por la frecuencia de clase.

**Paso 8:** Interpretar el histograma. Analiza aspectos como la tendencia, la variabilidad y la forma de distribución de los datos.

Apliquemos los pasos anteriores con un ejemplo:

**Paso 1:** Partiendo de un estudio de peso en 100 alumnos de nivel primaria, se obtuvieron los siguientes datos:

No.	Peso (kg)	No.	Peso (kg)	No.	Peso (kg)	No	Peso (kg)
-----	-----------	-----	-----------	-----	-----------	----	-----------

Documentos Educativos

1	23	26	24	51	41	76	36
2	54	27	43	52	17	77	46
3	55	28	43	53	36	78	35
4	27	29	34	54	39	79	48
5	54	30	20	55	32	80	39
6	41	31	29	56	29	81	44
7	38	32	41	57	32	82	31
8	30	33	36	58	36	83	21
9	52	34	47	59	59	84	35
10	32	35	18	60	46	85	20
11	41	36	32	61	26	86	40
12	39	37	54	62	59	87	54
13	34	38	54	63	23	88	53
14	54	39	53	64	38	89	20
15	39	40	52	65	24	90	33
16	16	41	53	66	34	91	30
17	59	42	35	67	41	92	25
18	43	43	29	68	44	93	43
19	50	44	54	69	23	94	27
20	22	45	47	70	42	95	58
21	31	46	30	71	32	96	29
22	19	47	28	72	57	97	33
23	47	48	59	73	33	98	41
24	43	49	54	74	59	99	45
25	42	50	42	75	36	100	32

**Paso 2:** Para determinar el rango, se ubica el valor más bajo (16) y el valor más alto (59), una vez hecho esto, se procede a obtener el rango resolviendo la ecuación:

$$\text{Rango} = \text{Límite superior} - \text{Límite Inferior}$$

$$\text{Rango} = 59 - 16$$

$$\text{Rango} = 43$$

**Paso 3:** Determinar el número de clases en que se dividirá el conjunto de datos.

Considerando la primera técnica, se tendrían 12 clases.

Cantidad de datos	Número de clases
20-50	7
50-75	10
75-100	12
Más de 100	15

Considerando la segunda técnica se tiene:

$$K = \sqrt{\text{número de datos } (N)}$$

$$K = \sqrt{100}$$

$$K = 10 \text{ clases}$$

Considerando la Regla de Sturges, se obtiene:

$$K = 1 + 3.322 \log(N)$$

$$K = 1 + 3.322 \log(100)$$

$$K = 1 + 3.322 * 2$$

$$K = 1 + 6.644$$

$$K = 7.644$$

Redondeando:  $K = 8$  clases

Para fines de este ejemplo, se tomará el resultado de la Regla de Sturges, ya que es matemáticamente más consistente.

**Paso 4:** Determinar la amplitud de clase, para lo cual se debe resolver la siguiente ecuación:

$$h = \frac{\text{Rango}}{k}$$

$$h = \frac{43}{8}$$

$$h = 5.375 \approx 6$$

**Paso 5:** Determinar las clases, iniciando con el límite inferior de clase:

$$\text{Primera clase} = 16 + 6 = 22$$

$$\text{Segunda clase} = 22 + 6 = 28$$

$$\text{Tercera clase} = 28 + 6 = 34$$

$$\text{Cuarta Clase} = 34 + 6 = 40$$

$$\text{Quinta clase} = 40 + 6 = 46$$

$$\text{Sexta clase} = 46 + 6 = 52$$

Documentos Educativos

Séptima clase =  $52 + 6 = 58$

Octava clase =  $58 + 6 = 64$

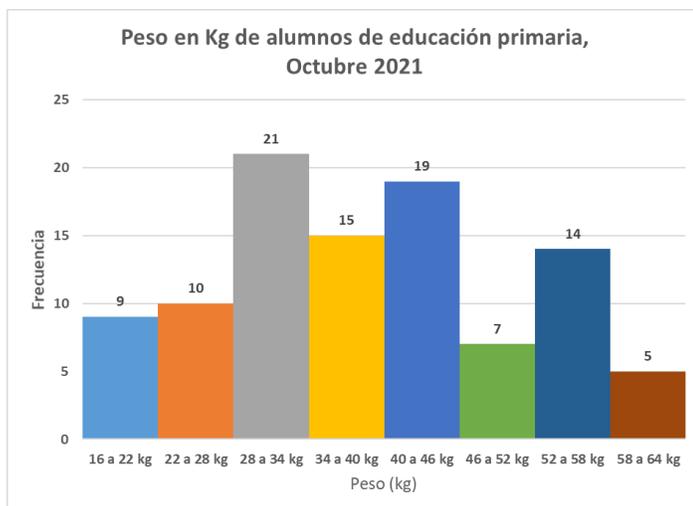
Las clases definidas son:

Clase	Límites de clase
1	16 a 22 kg
2	22 a 28 kg
3	28 a 34 kg
4	34 a 40 kg
5	40 a 46 kg
6	46 a 52 kg
7	52 a 58 kg
8	58 a 64 kg

**Paso 6:** Elaborar la tabla de frecuencia agrupando cada dato en su clase correspondiente.

Clase	Frecuencia
16 a 22 kg	9 datos en la clase
22 a 28 kg	10 datos en la clase
28 a 34 kg	21 datos en la clase
34 a 40 kg	15 datos en la clase
40 a 46 kg	19 datos en la clase
46 a 52 kg	7 datos en la clase
52 a 58 kg	14 datos en la clase
58 a 64 kg	5 datos en la clase

**Paso 7:** Se procede a trazar el histograma con los datos obtenidos en la tabla:



El histograma permite la exposición clara, ordenada, resumida y precisa de un conjunto de datos. En este tipo de gráficos se muestra a través de la anchura de las barras los intervalos que reflejan si una variable es continua o discreta. Además, la altura de cada barra, representa la frecuencia relativa de cada categoría.

En este caso, como la variable Peso (kg) es una variable continua, las barras del histograma deben tocarse entre sí; si se tratara de una variable discreta, cuyas modalidades sólo pudieran adquirir valores enteros, las barras deben presentar espacios en blanco entre ellas.

**Paso 8:** Interprete los datos:

1. ¿Cuál es la clase que tiene un mayor número de representantes?
2. ¿Cuál es la clase que tiene menos?
3. ¿Cuáles serían los pesos menos frecuentes en la muestra?
4. ¿y los más frecuentes?

Intente responder estas preguntas y de encontrar mayor cantidad de información analizando detenidamente el histograma

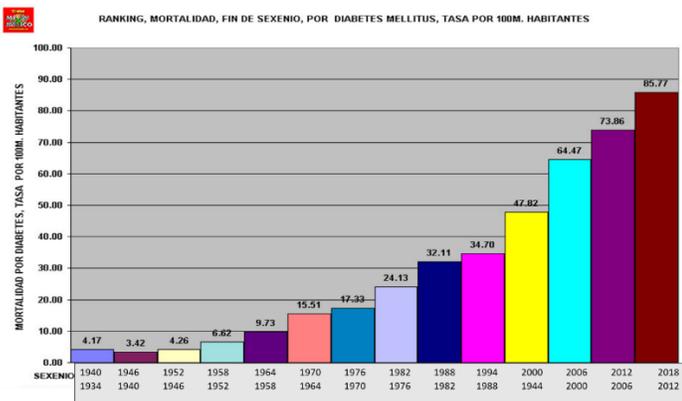
Para mostrar la importancia de esta herramienta en el estudio y resolución de problemas de salud pública, a continuación, se muestra un ejemplo de su aplicación en el estudio por la OMS de una afección muchas veces emanada del sobrepeso y la obesidad: **diabetes tipo II**.

La diabetes, según la OMS, es una enfermedad crónica que se desarrolla cuando el páncreas no produce insulina suficiente o cuando el organismo no es capaz de utilizar eficazmente la que produce.

El sobrepeso y la obesidad son dos padecimientos que facilitan el desarrollo de la diabetes tipo II, agravando el problema de salud pública. Según proyecciones de la OMS, para 2030 la diabetes será la séptima causa de mortalidad a nivel mundial.

## Documentos Educativos

En el informe “Principales causas de Mortalidad en México, 1938-2017” se publica información histórica correspondiente a todas las causas de muerte en México, de ahí se toma la siguiente gráfica sobre la causa de muerte por diabetes durante los sexenios presidenciales de 1935 hasta 2018, donde el último sexenio corresponde a cifras estimadas.



Fuente: <http://mexicomaxico.org/Voto/RankPres.htm#mortdiabetes>

En este histograma se muestran intervalos correspondientes a cada sexenio presidencial donde la longitud de cada intervalo se obtiene mediante la diferencia, por ejemplo, sexenio 1934 a 1940 es de longitud 6 dada la diferencia  $1940-1934=6$ . Para cada intervalo tenemos una barra con la misma longitud (6 años), y su altura es la frecuencia relativa.

Analizando la gráfica, es posible identificar la siguiente información:

La tendencia de la mortalidad por diabetes mellitus tiene va en aumento de un sexenio a otro.

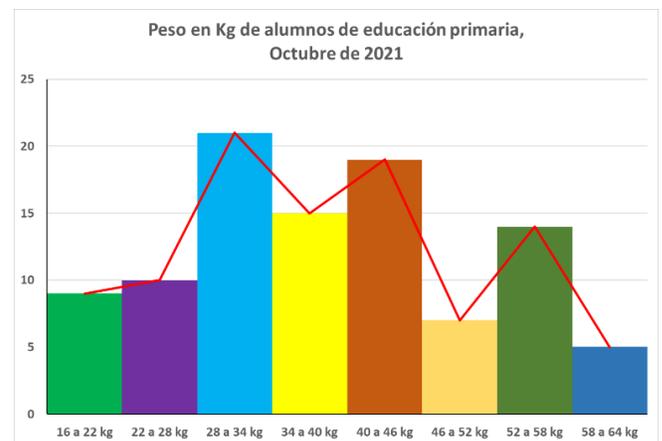
El caso menos crítico es para el sexenio 1940 a 1946 donde tan solo el 3.42% de la población moría a causa de la diabetes mellitus.

En el sexenio 2006 - 2012 la mortalidad por diabetes mellitus aumentó considerablemente, un 73.86%, es decir, aumentó en 70.44 puntos porcentuales.

## Polígono de frecuencia

Los polígonos de frecuencia son herramientas gráficas que se construyen a partir de un HISTOGRAMA.

Para obtener un POLÍGONO DE FRECUENCIA, se deben marcar los CENTROS DE CLASE de la parte superior de cada una de las barras que conforman el histograma, una vez hecho esto, se procede a unir con una línea cada uno de los centros, obteniendo lo siguiente:



Si se omiten las barras, tenemos la siguiente gráfica:



Note que en el POLÍGONO DE FRECUENCIA se considera un punto al inicio y otro punto al final, ambos con frecuencia 0.

La ventaja de presentar un polígono de frecuencia es que el investigador es capaz de inferir la frecuencia de

## Documentos Educativos

los pesos que se encuentran entre las marcas de clase y que no se consideraron en el estudio. Esta gráfica, que presenta segmentos de línea que unen los centros de clase, muestra de una manera más clara el patrón de comportamiento de los datos. Mientras más categorías y observaciones se consideren en la construcción del polígono, la línea se suavizará y presentará la apariencia de una curva, lo que se puede ver en la siguiente gráfica:



Fuente: <https://www.mexicosocial.org/un-ano-de-covid-19-en-mexico/>

## Ojiva de frecuencia

Una OJIVA DE FRECUENCIA es la gráfica que, a través de una línea, muestra las FRECUENCIAS ACUMULADAS de una serie de datos, en otras palabras, puede considerarse a la ojiva de frecuencia como un POLÍGONO DE FRECUENCIAS ACUMULADAS de un conjunto de datos.

Si la gráfica presenta las FRECUENCIAS ABSOLUTAS ACUMULADAS, recibe el nombre de OJIVA, por otra parte, si en su construcción se utilizan las frecuencias relativas acumuladas expresadas en porcentaje, la gráfica recibe el nombre de OJIVA PORCENTUAL.

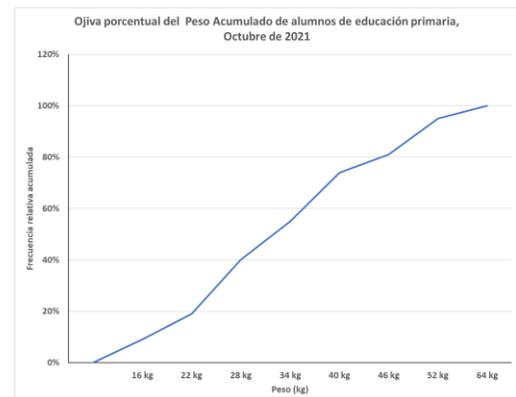
De manera similar a la construcción de un polígono de frecuencia, para construir una ojiva a partir de una tabla de frecuencia, se consideran en el eje de las abscisas los valores de la variable y en el eje de ordenadas las

frecuencias acumuladas o las frecuencias relativas acumuladas.

Las ojivas son ampliamente utilizadas para representar series de frecuencia en variables de tipo cuantitativo. Es posible construir una ojiva tanto con conjuntos de datos agrupados en clases o con los datos sin agrupar.



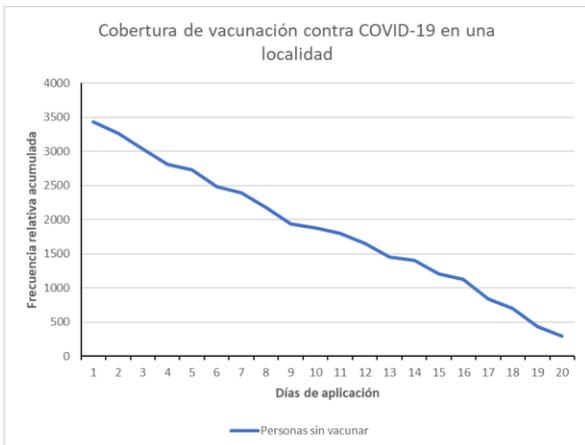
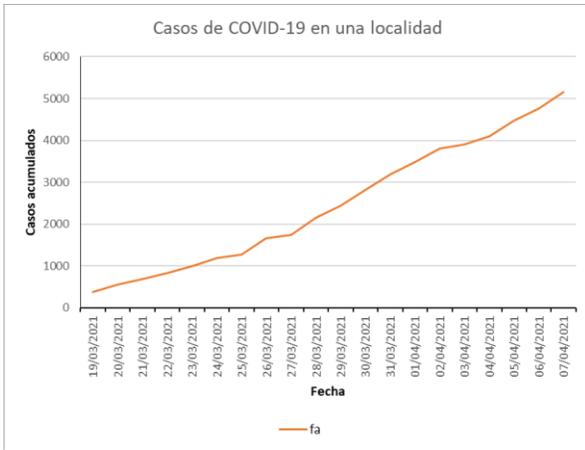
Si se utiliza la FRECUENCIA RELATIVA ACUMULADA, expresada en porcentaje, la ojiva permite identificar la distribución de los percentiles de cada una de las modalidades que presenta la variable de estudio.



Dependiendo del ordenamiento de los datos al construir la gráfica, podemos distinguir varios tipos de ojiva de frecuencia. Si la gráfica presenta los datos en orden creciente de izquierda a derecha, se le conoce como OJIVA DE FRECUENCIA MAYOR QUE  $F_i$ , mientras que si presenta los datos descendentes en el mismo sentido (el valor porcentual decrece a medida

## Documentos Educativos

que se desplaza a la derecha el eje x), se conoce como OJIVA DE FRECUENCIA MENOR QUE  $F_i$ .



## Referencias:

Anderson, D. R., Sweeney, D., & Williams, T. A. (1999). *Estadística para la Administración y Economía*. México DF, México: International Thompson Editores.

Departamento de Didáctica de la Matemática. (2011). *Estadística con proyectos*. (C. Batanero, & C. Díaz, Eds.) Granada, España: Facultad de Ciencias de la Educación, Universidad de Granada.

García Pérez, A. (2008). *Estadística aplicada: conceptos básicos (2a edición ed.)*. Madrid, España: Educación permanente / Universidad Nacional de Educación a Distancia.

Kazmier, L. J., Díaz Mata, A., & Eslava Gómez, G. (1991). *Estadística Aplicada a Administración y Economía*. Naucálpan, Estado de México México, Atlacomulco, México: McGraw Hill.

Pérez López, C. (1999). *Control estadístico de la calidad*. Madrid, España: Alfa Omega.

Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. (2010). *Estadística Matemática con aplicaciones*. México, D.F., México: Cengage Learning Editores, S.A.

## Declaración de conflicto de intereses

El autor expresa que no tuvo ningún conflicto de intereses durante la preparación de este documento ni para su publicación.

## Obra protegida con una licencia

**Creative Commons**



Atribución - No Comercial -  
No Derivadas