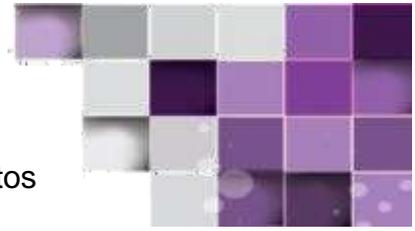




Estadística básica

Unidad 2. Representación numérica y gráfica de datos



Índice

Presentación de la Unidad	3
2.1. Recolección de datos	6
2.2. Organización de datos y distribución de frecuencias.....	8
2.2.1. Datos no agrupados	8
2.2.2. Datos agrupados	9
2.3. Representación gráfica de datos.....	21
2.4. Análisis de datos	25
Cierre de la Unidad	26
Fuentes de consulta.....	26

Presentación de la Unidad



Representación de datos

Fuente: patrisyu, 2014. freedigitalphotos.net

En la unidad anterior se dijo que la estadística es un conjunto de técnicas para describir grupos de datos y para tomar decisiones en ausencia de una información completa. Para realizar estas funciones se divide en dos ramas, la estadística **descriptiva** y la estadística **inferencial**.

La **estadística descriptiva** comprende la tabulación, representación y descripción de una serie de datos que pueden ser cuantitativos, como la medida de la estatura y el peso, o cualitativas, como el sexo o el nivel socioeconómico.

La **estadística inferencial** consiste en estimar las propiedades (variables) de una población a partir del conocimiento de sólo una muestra de ella. Está basada en la estadística descriptiva y la teoría de la probabilidad.

En esta unidad se estudiará a la estadística descriptiva, aprenderás a cómo organizar, presentar e interpretar los datos que se obtienen de las muestras tomadas de las poblaciones. Antes de comenzar con los temas, se verá de dónde y cómo se obtienen los datos que se van a organizar.

Propósitos

Mediante el estudio de esta unidad se busca que logres los siguientes propósitos:

- **Identificar** algunos conceptos que se utilizan en la estadística descriptiva.
- **Analizar** datos cuantitativos y cualitativos mediante la interpretación de tablas y gráficas.

Competencia específica

Con la integración de los elementos declarativos, procedimentales, actitudinales y contextuales de esta unidad, lograrás analizar bases de datos para describir fenómenos mediante la interpretación de tablas de frecuencias y gráficas.



Actividades

Conforme vayas avanzando en el estudio de esta unidad, puedes ir realizando las actividades correspondientes a esta unidad. La descripción de las mismas puedes encontrarla en el documento *Unidad 2. Actividades*; puedes encontrar este documento en la pestaña de la unidad.



Para saber más

Se ha seleccionado una serie de recursos en línea con el fin de ofrecerte un panorama general de la unidad y alternativas en caso de que se te dificulte la comprensión de algún concepto o proceso. Estos recursos tienen una extensión breve pero complementan lo expuesto en este material. Algunos de ellos te ayudarán a visualizar cómo puedes aplicar estos contenidos en las actividades de la unidad. Puedes revisarlos en este momento o volver cuando hayas concluido la revisión de esta unidad.

Todos menos el último de la lista están disponibles en la sección *Material de apoyo*.

Los primeros recursos son dos presentaciones que sintetizan los temas de esta unidad.

- Centro Universitario de Ciencias de la Salud (s.f.). *Distribución de frecuencias*. Recuperado de: http://www.cucs.udg.mx/movimientohumano/files/File/06-Distribucion_frecuencia.ppt
- UVEG (2009) *Distribución de frecuencias*. Recuperado de: <http://docplayer.es/41780815-Distribuciones-de-frecuencia.html>

Los siguientes materiales ofrecen una descripción y explicación breve de cada concepto y proceso que se revisa en la unidad.

- Cabrera, S. (s.f.). Distribuciones de frecuencias. En *Estadística descriptiva* (p. 2-32). Recuperado de: <https://vdocuments.mx/distribucion-de-frecuencias-fcpsuaq-muestral-dato-distribuciones-de-frecuencias.html>

- Morales, G. (2012) *Estadística y probabilidades*. Recuperado de: <https://hopelchen.tecnm.mx/principal/sylabus/fpdb/recursos/r116835.PDF>
- Villegas, A. (2012). Distribuciones de frecuencias de variables discretas. Recuperado de: <https://www.sistemamid.com/panel/uploads/biblioteca/1/349/1259/6572/6581/77555.pdf>

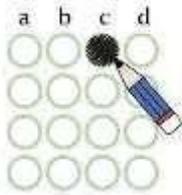
Este último recurso expone algunos de los temas de esta unidad en video.

- uned.fisicaymates.com (2014). *Tema 3: Distribución de frecuencias unidimensionales*. Disponible en: <https://fisicaymates.com/tema-3-distribucion-de-frecuencias-unidimensionales-introduccion-a-la-estadistica-grado-turismo-uned/>

Recolección de datos

Cuando se realiza un trabajo para analizar los datos acerca de un hecho o fenómeno, las personas que realizan el trabajo diseñan instrumentos para recolectar la información y obtener los datos que necesitan. Existen muchos métodos para recolectar información, pero los más frecuentes son:

Censos



Respuestas

Fuente: nongpimmy, 2013.
freedigitalphotos.net

Es una técnica de recolección de datos que se aplica a la totalidad de los elementos que componen la población o universo que se estudia. Un censo debe cumplir dos condiciones:

- **Universalidad.** Esto es, se debe tomar en cuenta a todos los elementos de la población.
- **Simultaneidad.** Debe realizarse dentro de un periodo de tiempo limitado.

Encuesta



Opciones

Fuente: Jeroen van Oostrom, 2011.
freedigitalphotos.net

Esta técnica se utiliza para recolectar información de una muestra de la población. Consiste en presentar un conjunto de preguntas abiertas (preguntas que no tienen respuestas predeterminadas) o cerradas (preguntas que cuentan con una serie de respuestas establecidas).

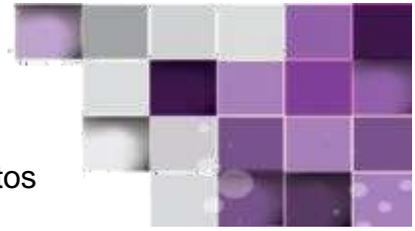
Experimento



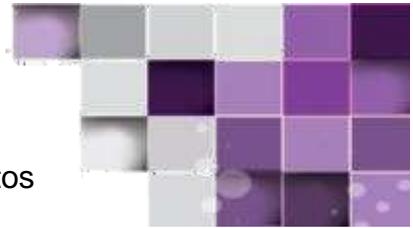
Llenado

Fuente: nongpimmy, 2011.
freedigitalphotos.net

Otra de las técnicas más usadas en estadística para recolectar información es el experimento. Un **experimento** es una prueba que se realiza para determinar las características o comportamiento de un objeto o sujeto. Por ejemplo, experimentar con el sentido del gusto para conocer qué alimentos nos parecen más salados.

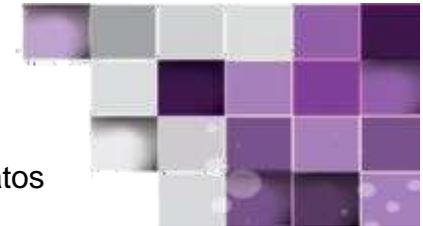


Universidad Abierta y a Distancia de
México

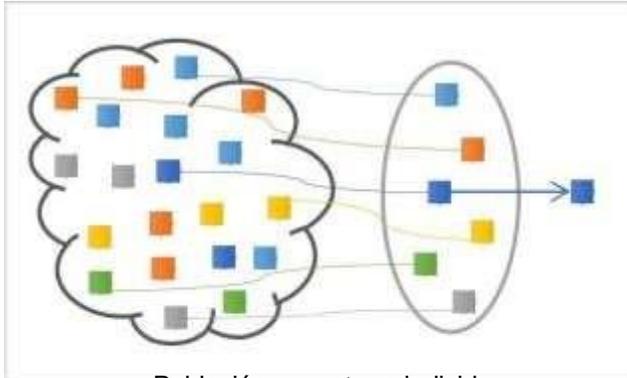


Un experimento también se define como **el proceso que se realiza para verificar una serie de hipótesis relacionadas con un determinado fenómeno**, en el cual se determinan las características o comportamientos del fenómeno que se analiza. Por ejemplo, un experimento para determinar la velocidad de la luz en el vacío, donde se está determinando la velocidad de la luz.

La diferencia entre la primera y la segunda definición está en que en la última se parte de una hipótesis. En el primer ejemplo, se experimentan los sabores de los alimentos sin antes predecir cuál sabrá más salado. En el segundo ejemplo, la hipótesis, a partir de estudios anteriores, es que la velocidad de la luz en el vacío es de 300 000 km/seg. El experimento verificará si esta hipótesis es cierta o no y en éste cabe un margen de error experimental.



Organización de datos y distribución de frecuencias



Población, muestra e individuo

Adaptado de: Sosci (1998). <http://simon.cs.vt.edu/>

Como se ha dicho anteriormente, la estadística descriptiva organiza, representa, describe y resume los datos obtenidos de una población o de una muestra de ésta, sin elaborar inferencias ni obtener conclusiones.

Con el propósito de que los datos obtenidos de una muestra o población sean más significativos, es común realizar una distribución de frecuencias y dibujar gráficas de varios tipos para representar dichos datos. De esta forma

se pueden tener datos agrupados y no agrupados. Y si se tienen datos agrupados, se tienen que considerar conceptos como frecuencia e intervalo.

Datos no agrupados

Se denominan datos no agrupados al conjunto de datos obtenidos, que por ser muy pocos, no requieren una agrupación bajo ciertas especificaciones. En este caso, se considera que el número de datos no debe sobrepasar a 30.

¿Qué procedimiento se utiliza para organizar y presentar estos datos? En ocasiones es útil ordenar los valores de los datos en orden creciente o decreciente pero aún esto no resulta una labor sencilla. Recientemente, se ha encontrado una técnica para ordenarlos denominada **gráfica de tronco y hoja**.

Para ilustrar la técnica gráfica de tronco y hoja, observa las siguientes calificaciones en una prueba de coordinación física aplicada a 20 personas que habían ingerido una cantidad de alcohol equivalente a 0.1% de su peso.

69, 84, 52, 93, 61, 74, 79,
65, 88, 63, 57, 64, 67, 72,
74, 55, 82, 61, 68, 77

Ahora se separan las cifras de cada número en sus decenas y unidades, disponiendo juntos los valores que comparten las decenas. Esto es, pensaremos en el número 69 como 6/9. Entonces las decenas se dispondrán en forma vertical con unidades dispuestas al lado. Para el conjunto de las 20 calificaciones de coordinación física, la gráfica sería la ubicada al

5	2 7 5
6	9 1 5 3 4 7 1 8
7	4 9 2 4 7
8	4 8 2
9	3

lado de este texto. El primer renglón de la gráfica 5 2 7 5 indica que la lista contiene los valores de 52, 57 y 55. El segundo renglón indica que la lista contiene ocho valores de la decena de los 60. Esta tabla se conoce como una representación gráfica de tronco y hoja porque cada renglón representa una posición de tronco y cada dígito a la derecha de la línea vertical se puede considerar como una hoja.



Para saber más

La revisión del siguiente material te ayudará a visualizar otras formas de utilización de la técnica tronco y hoja, para que puedas aprovecharla al máximo (disponible en *Material de apoyo*).

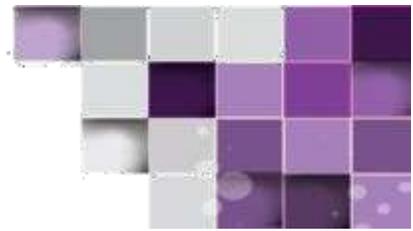
- Estadística para todos (s/a). *Diagrama de tallos y hojas*. Recuperado de: http://www.estadisticaparatodos.es/taller/graficas/tallos_hojas.html

Datos agrupados

Se denominan datos agrupados cuando las observaciones de una muestra se agrupan en clases o intervalos de clase. El hecho de agrupar los datos, cuando el número de observaciones es muy grande, permite sintetizar la información para una mejor descripción de la muestra. Para sintetizar la información, en estadística se utilizan las frecuencias para poder condensar los datos y entender mejor su comportamiento como a continuación se describe.

Frecuencia

- **Frecuencia:** es el número de veces que se repite un dato, también se le conoce como frecuencia absoluta.
- **Frecuencia acumulada:** es la suma de las frecuencias absolutas de las variables hasta el renglón. También es conocida como frecuencia absoluta acumulada.
- **Frecuencia relativa:** es el resultado de dividir la frecuencia entre el número total de datos. Este dato también puede verse como un porcentaje.
- **Frecuencia relativa acumulada:** es la suma de las frecuencias relativas hasta el renglón.



Supongamos que se tiene la siguiente distribución de datos de la edad de personas que aprenden a nadar en una alberca pública en un horario de 16 a 17 horas:

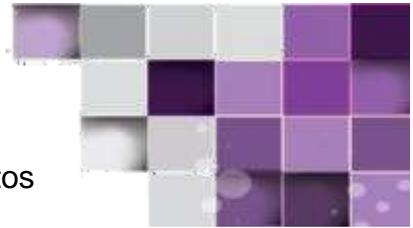
18, 41, 23, 47, 18, 23, 23, 41, 41, 47, 47, 52, 23, 47, 23, 47, 18, 47, 7, 23, 18, 47, 52, 41, 52, 18, 23, 52, 7, 18, 52, 23



En la siguiente tabla puedes ver los datos anteriores organizados en una tabla para que puedas identificar los tipos de frecuencia mencionados (los datos siempre se ordenan de manera creciente).

Tipos de frecuencia						
Número de renglón ()	Datos obtenidos de la variable	Frecuencia	Frecuencia acumulada	Otra forma para obtener	Frecuencia relativa	Frecuencia relativa acumulada
1	7	$f_1 = 2$	$f_1 = F_1 = 2$	$f_1 = F_1 = 2$	$h_1 = f_1 / N = 0.0625$	$h_1 = H_1 = 0.0625$
2	18	$f_2 = 6$	$f_1 + f_2 = F_2 = 8$	$F_1 + f_2 = F_2 = 8$	$h_2 = f_2 / N = 0.1875$	$h_1 + h_2 = H_2 = 0.2500$
3	23	$f_3 = 8$	$f_1 + f_2 + f_3 = F_3 = 16$	$F_2 + f_3 = F_3 = 16$	$h_3 = f_3 / N = 0.2500$	$h_1 + h_2 + h_3 = H_3 = 0.5000$
4	41	$f_4 = 4$	$f_1 + f_2 + f_3 + f_4 = F_4 = 20$	$F_3 + f_4 = F_4 = 20$	$h_4 = f_4 / N = 0.1250$	$h_1 + h_2 + h_3 + h_4 = H_4 = 0.6250$
5	47	$f_5 = 7$	$f_1 + f_2 + f_3 + f_4 + f_5 = F_5 = 27$	$F_4 + f_5 = F_5 = 27$	$h_5 = f_5 / N = 0.2187$	$h_1 + h_2 + h_3 + h_4 + h_5 = H_5 = 0.8437$
6	52	$f_6 = 5$	$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = F_6 = 32$	$F_5 + f_6 = F_6 = 32$	$h_6 = f_6 / N = 0.1563$	$h_1 + h_2 + h_3 + h_4 + h_5 + h_6 = H_6 = 1.0000$
Total		$N = 32$			1.0000	

Significado de símbolos		
i Renglón	f Frecuencia	h Frecuencia relativa
N Número total de datos	F Frecuencia acumulada	H Frecuencia relativa acumulada



Distribución de frecuencias

Como se dijo anteriormente, el hecho de agrupar los datos cuando el número de observaciones es muy grande, permite sintetizar la información para una mejor descripción de la muestra. Para lograrlo, las observaciones se agrupan en clases o intervalos de clase en una tabla de distribución de frecuencias.

Una vez que se han tabulado o representado los datos, se pueden calcular medidas de tendencia central y dispersión (que se verán más adelante), las cuales describen con mayor precisión la muestra o población de interés.

Se iniciará con la revisión de algunos conceptos sobre el intervalo, para continuar con los pasos a seguir para elaborar una distribución de frecuencias para una muestra de datos.

Definiciones

Intervalo o rango

Conjunto de valores comprendidos entre otros dos números dados, conocidos estos últimos como límites del intervalo.

Intervalo de clase

Es la expresión que se utiliza para nombrar a un intervalo. Tiene un límite superior y un límite inferior .

Amplitud del intervalo

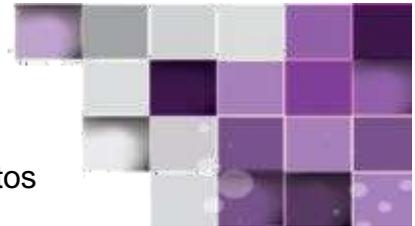
Es la diferencia del límite superior menos el límite inferior (-).

Fronteras de clase

Son los puntos medios entre los límites de intervalos consecutivos. Las fronteras de clase se utilizan para recuperar los datos entre el límite superior de un intervalo y el límite inferior del siguiente.

Marca de clase

Es el punto medio del intervalo y es el resultado de la suma de los límites inferior y superior del intervalo dividido entre 2. A la marca de clase también se le denomina punto medio de clase.



Ejemplo

Intervalo o rango

Para ejemplificar los conceptos se utilizarán los números 15 y 25
El intervalo corresponde a todos los números que se encuentran entre el 15 y el 25.

Intervalo de clase

El intervalo de clase sería: 15-25. Los límites del intervalo son:

- Límite inferior = 15
- Límite superior = 25

Amplitud del intervalo

La amplitud del intervalo 15-25 sería: 25 menos 15, es decir 10. Es recomendable que todos los intervalos tengan la misma amplitud. Para ello podemos restar el dato menor del dato mayor y dividir este resultado entre el número de intervalos que se deseen.

Fronteras de clase

Si se toman los intervalos 4-14, 15-25 y 26-36, las fronteras de clase serían: 3.5 y 14.5 para el primer intervalo, 14.5 y 25.5 para el segundo intervalo, y por último, 25.5 y 36.5 para el tercer intervalo.

Intervalos	Fronteras de clase
4-14	3.5 y 14.5
15-25	14.5 y 25.5
26-36	25.5 y 36.5

La frontera de clase no debe coincidir con los datos límite del intervalo, porque sería complicado identificar el intervalo al que pertenece dicho dato.

Nuevos intervalos
3.5-14.5
14.5-25.5
25.5-36.5

Ejemplo: Con base en las fronteras dadas se construyen los nuevos intervalos 3.5-14.5, 14.5-25.5 y 25.5-36.5. Si se tiene el dato 25.5 no se sabría si ponerlo en el segundo o en el tercer intervalo. Si esta coincidencia sucede deberá moverse el intervalo. Siguiendo con el ejemplo, si se mueve un punto a la izquierda, se tendrían los intervalos 2.5-13.5, 13.5-24.5 y 24.5-35.5.

Ajuste de intervalos
2.5-13.5
13.5-24.5
24.5-35.5

Marca de clase

La marca de clase del intervalo 15 – 25 es igual a:

$$\frac{15 + 25}{2} = \frac{40}{2} = 20$$

Es recomendable que la marca del intervalo coincida con alguno de los datos. Esto no es estrictamente necesario y no siempre se logra, sobre todo cuando los intervalos tienen la misma amplitud.

Elaboración de una distribución de frecuencias

Los pasos a seguir para determinar una distribución de frecuencias para una muestra o población se presentan a continuación.

1. Calcular el rango

La formación de clases o intervalos de clase, que se representan con (), dependen, generalmente, del tamaño del rango de la población o muestra

Para calcular el rango se identifica el número mayor () y el número menor () en los datos. El rango es el resultado de la resta, esto es:

Por ejemplo:

Si en una serie de datos que van desde el 18 hasta el 56, se tendría lo siguiente:
 $R = 56$ y $r_1 = 18$, por lo tanto:

2. Determinar el número de intervalos que se desea tener

No existe una regla para determinar el número de intervalos, pero generalmente se suelen crear entre 5 y 20 intervalos. La decisión la toma el investigador. Siguiendo con el ejemplo, se diría que vamos a construir 7 intervalos. Entonces se dice que $k = 7$.

3. Dividir el rango entre el número de intervalos que se desea tener

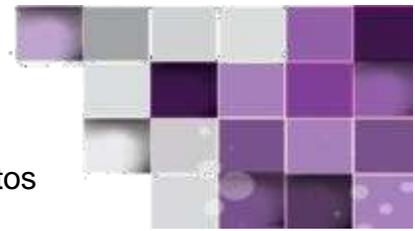
Recuerda que lo recomendable es elegir un número entre 5 y 20 para los intervalos. Se divide entre uno menos de los intervalos deseados porque con el número de datos se acumula un intervalo más.

Siguiendo con el ejemplo, se quieren 7, entonces:

$$\frac{38}{7} \approx 5.428$$

Ésta será la amplitud de los intervalos. Cuando no es un número entero, se escoge el entero más cercano, como en este caso, se toma el rango igual a 5.

Cuando la cantidad de datos es tal que no alcanza para acumular un intervalo más, entonces se divide entre el número de intervalos que se quieren.



4. Se forman los intervalos

Los intervalos se forman comenzando un número antes del primer dato:

INTERVALOS:

17 a 22 (se cuenta 5 desde 18 hasta 22) 23 a

28

29 a 34

35 a 40

41 a 46

47 a 52

53 a 58

Nota: No importa que el último intervalo exceda el último dato.

Ejemplo de distribución de frecuencias

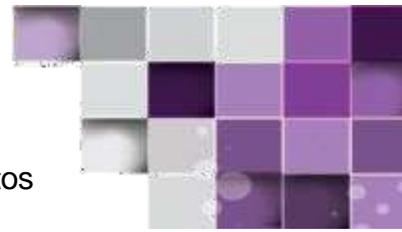
Para ilustrar la distribución de frecuencias de una muestra de datos se usará el siguiente ejemplo:

El director de una consultoría en desarrollo de software desea conocer el número de incidencias en sus desarrollos reportadas durante los meses de agosto y septiembre. Para ello pide a uno de sus empleados que le elabore un reporte; el empleado tiene los siguientes datos:

35, 24, 26, 23, 50, 20, 25, 56, 30, 30, 38, 36, 35, 29, 28, 30, 40, 39, 38, 40, 27, 24, 30, 32, 35, 27, 29, 22, 28, 27, 48, 40, 48, 31, 39, 28 46, 36, 37, 52, 44, 49, 52, 41, 31, 31, 56, 58, 38, 26, 25, 24, 60, 55, 48, 37, 31, 30, 22, 20.

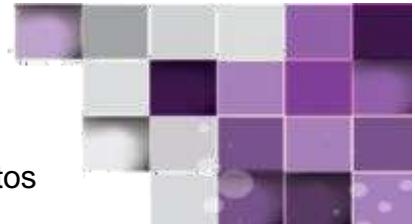
1. Calcular el rango	$= - 1 = 60 - 20 = 40$
2. Determinar el número de intervalos que se desea tener	Se eligieron 8 intervalos
3. Dividir el rango entre el número de intervalos que se desea tener	$\frac{40}{8} = 5$
4. Se forman los intervalos	Se comienza por un número anterior al límite inferior: 19 - 24, 25 - 29, 30 - 35, 36 - 40, 41 - 45, 46 - 50, 51 - 55, 56 - 60.

Finalmente se elabora la distribución de frecuencias. En la siguiente tabla puedes observar la forma en que quedan distribuidas las frecuencias.



Distribución de frecuencias

Clase	Intervalo de clase	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Distribución de porcentajes	Distribución de porcentajes acumulados	Marca de clase
1	19-23	5	5	0.083	8.3%	8.3%	21
2	24-28	13	18	0.216	21.6%	29.9%	26
3	29-33	12	30	0.20	20%	49.9%	31
4	34-38	10	40	0.166	16.6%	66.5%	36
5	39-43	5	45	0.083	8.3%	74.8%	41
6	44-48	6	51	0.1	10%	84.8%	46
7	49-53	4	55	0.066	6.6%	91.4%	51
8	54-58	4	59	0.066	6.6%	98%	56
9	59- 63	1	60	0.016	1.6%	99.6%	61
		60					



Tablas

Existen diferentes tipos de tablas para presentar los datos, las más utilizadas son: **Tabla de datos**, **Tabla de frecuencias**, **Tabla por intervalos de clase** y **Tablas de doble entrada**. A continuación se presenta en qué consiste cada una.

Tabla de datos

Una tabla de datos es la forma más sencilla de organizar un conjunto de datos y se utiliza cuando la información que se necesita son los datos mismos. Se organizan en columnas o renglones y se registran las mediciones o datos obtenidos.

Ejemplo:

Supón que la medición de temperatura a lo largo del día da como resultado los siguientes valores en grados Celsius: 20.4, 21.2, 22.1, 23.9, 25.3, 26.9, 27.7. Entonces se construye una tabla como la que se muestra.

Temperatura (Celsius)	20.4	21.2	22.1	23.9	25.3	26.9	27.7
--------------------------	------	------	------	------	------	------	------

Tabla de frecuencias

Ésta nos aporta mayor información pues está formada por categorías de la variable que se esté midiendo y su frecuencia (es decir, el número de ocurrencias de un valor dado).

Ejemplo:

Supón que un experimento da los siguientes valores medidos:

1,2,2,2,1,1,5,4,3,2,2,1,3,4,5,6,2,3,4,5,5,4,3,3,2

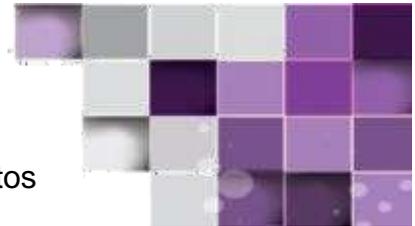
Se procede entonces a agrupar por categorías, según la frecuencia o número de veces que aparece cada medición:

Nota: Observa que aunque los datos son numéricos, la variable es cualitativa.

Valor de la Variable medida	Frecuencia
1	4
2	7
3	5
4	4
5	4
6	1

Tabla por intervalos de clase

Grupo de edad	Frecuencia
De 12 a 19	9600
De 20 a 24	7100
De 25 a 34	3900
De 35 a 44	1500



En este tipo de tablas los datos son presentados por intervalos de clase y no por los valores correspondientes a cada variable.

De 45 a 99	1600
------------	------

Ejemplo:

En una encuesta sobre el desempleo en el Área Metropolitana de la Ciudad de México, se organizan los datos por grupos de edades (intervalos de clase) y se presenta la frecuencia de cada intervalo, teniendo un total de 23,700 desempleados.

Tabla de doble entrada

Estas tablas proporcionan información referente a dos variables o eventos relacionados entre sí. Se forma poniendo en los renglones de la tabla la información de una de las variables y en las columnas la información de la otra variable.

Edades / No. de cirugías	Menos de 2 cirugías	Más de 2 cirugías
0-10	1	0
11-20	2	2
21-30	6	4
31-40	11	7
41-50	17	6
Más de 50	30	14

Ejemplo:

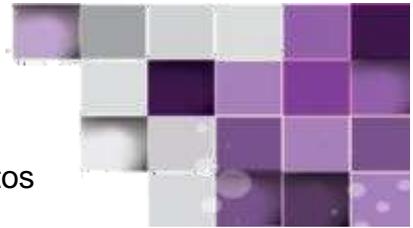
Supón que se mide el número de cirugías realizadas por edades en una muestra de 100 personas, encontrándose lo que se observa en la tabla.

Una tabla cualquiera puede ser vista como una tabla de doble entrada, en la cual las variables relacionadas son los rangos contra el valor de las variables en dicho rango.

Por ejemplo:

Supón que se mide la temperatura de un líquido con respecto al tiempo de calentamiento. En el renglón se colocan los tiempos y en las columnas la temperatura obtenida. Se podría considerar como una tabla de frecuencias o como una tabla de doble entrada.

Tiempo (min)	Temperatura (°C)
1-5	36
6-10	44
11-15	67



Representación gráfica de datos

En el tema anterior viste cómo tabular datos de una muestra y elaborar la distribución de frecuencias. Cuando las distribuciones se estructuran para condensar numerosos datos y representarlos en forma fácil de asimilar, es mejor presentarlos gráficamente, pues como se dice en el día a día, una fotografía dice más que mil palabras.

Ahora verás la representación gráfica de los datos. Las gráficas son representaciones visuales de los datos que se muestran en una tabla, existen diferentes tipos de gráficas, cada una de ellas se elabora con base en el tipo de información que se quiere representar.

Histograma

Histograma es la representación gráfica más común de una variable continua. Se elabora en un sistema de coordenadas rectangulares.

- El eje horizontal se utiliza para representar a la variable independiente, se pueden registrar los límites de clase o fronteras de clase.
- El eje vertical representa a las frecuencias.
- Si los intervalos de clase tienen el mismo ancho, las alturas de las barras serán proporcionales a las frecuencias.

El histograma también proporciona visualmente el aspecto de la distribución y dispersión de las mediciones.

Gráfica de barras

Este tipo de gráfica se utiliza para datos de tipo ordinal, nominal y discreto. En éstas se pueden representar la frecuencia, la frecuencia relativa y el porcentaje por medio de la altura de la barra y no por el área de la barra. Esta gráfica muestra las discontinuidades en las mediciones por medio de espacios vacíos entre las barras.

La gráfica de barras se traza sobre un eje de coordenadas y puede ser de dos formas:

Barras verticales:

- En el eje horizontal se representan los valores de la variable, pero no es necesario tener una escala horizontal continua.
- En el eje vertical se representa la frecuencia de cada clase.

Barras horizontales:

- En el eje horizontal se representan las frecuencias.
- En el eje vertical los valores de la variable.

Un histograma y una gráfica de barras son muy semejantes, la diferencia radica en que el histograma no presenta separación entre las barras.

Polígono de frecuencias

A menudo se usa el polígono de frecuencias en lugar del histograma. Difiere de éste en que sobre el eje de las x se registran las marcas de clase, que se completan con una marca en los extremos de la distribución cuya frecuencia es 0.

El polígono de frecuencias es la representación en un plano de los puntos (x_i, f_i) unidos por una línea quebrada o polígono. En los puntos obtenidos, las marcas de clase x son las abscisas, y la frecuencia f son las ordenadas.

Gráfica circular o de pastel

Para representar datos u observaciones de una variable cualitativa se usa una gráfica circular. Donde se divide un círculo en secciones, las cuales son proporcionales en tamaño con las frecuencias relativas o los porcentajes correspondientes.

Ojiva

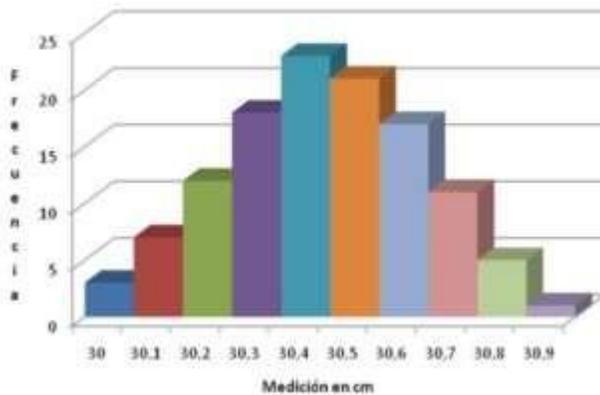
La representación gráfica de una distribución de frecuencias relativas acumuladas se denomina **ojiva**, se elabora sobre un plano de manera similar al polígono de frecuencias, pero en la ojiva el eje de las abscisas corresponde a los límites de clase y el de las ordenadas a los porcentajes acumulados.

Ejemplos de gráficas:

Para ilustrar los tipos de gráficas, antes mencionados, se ha considerado la siguiente tabla de datos.

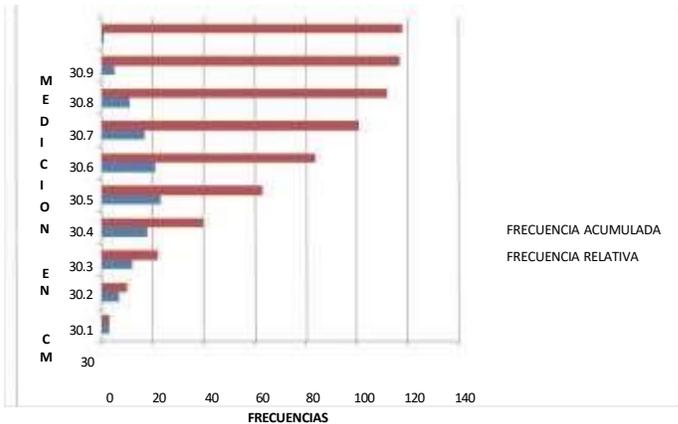
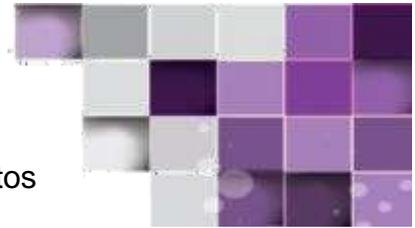
Medición en cm	Frecuencia	Frecuencia acumulada	Porcentaje
30	3	3	3%
30.1	7	10	6%
30.2	12	22	10%
30.3	18	40	15%
30.4	23	63	19%
30.5	21	84	18%
30.6	17	101	14%
30.7	11	112	9%
30.8	5	117	4%
30.9	1	118	1%

Histograma

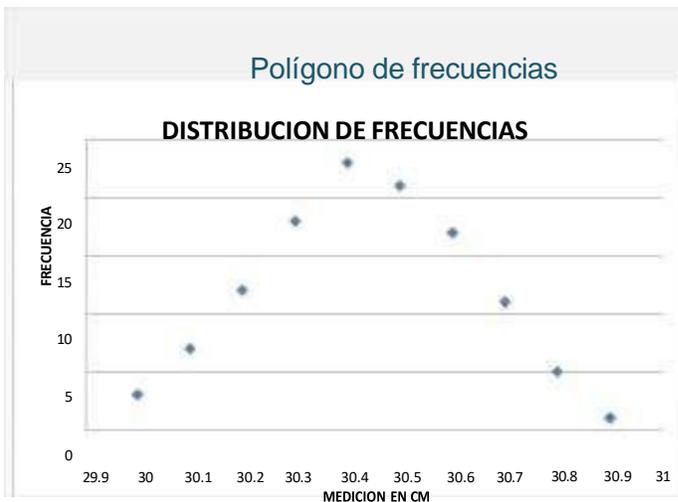


En esta figura se muestra el histograma de las mediciones en cm vs frecuencia, note cómo el ancho de las clases es el mismo.

Gráfica de barras



En esta gráfica de barras se puede observar la frecuencia relativa de cada evento y cómo se van acumulando las frecuencias hasta el total de los datos en la frecuencia absoluta.



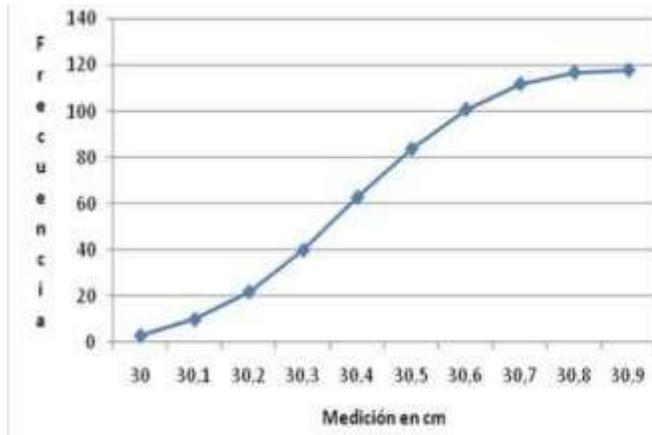
En esta gráfica se muestra la distribución acampanada de las frecuencias.

Gráfica circular o de pastel



En la gráfica de pastel se muestra dentro de cada "rebanada" la medición en cm y el porcentaje que corresponde a la frecuencia relativa.

Ojiva



En esta figura se muestra la frecuencia acumulada mediante una gráfica de línea.

Análisis de datos

La finalidad de construir distribuciones de frecuencias, ya sea con datos agrupados o no agrupados, consiste en que seas capaz de analizar e interpretar los datos, para ello, recurrirás en primera instancia al análisis de datos a partir de las frecuencias y posteriormente elaborarás representaciones gráficas que te permitan visualizar el comportamiento de los datos para obtener una primera aproximación a alguna conclusión.

Frecuencias relativas

La frecuencia relativa de una clase se obtiene dividiendo la frecuencia de cada clase entre el número total de observaciones de la muestra. Cuando estos resultados se multiplican por 100 el resultado se denomina **distribución de porcentajes**, la suma de las frecuencias relativas es igual a 1 (que representa al 100%). Por esta razón son muy útiles para elaborar una gráfica circular, para lo cual se requiere primero convertir la distribución de frecuencias relativas en una distribución porcentual.

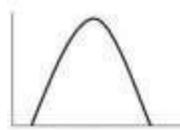
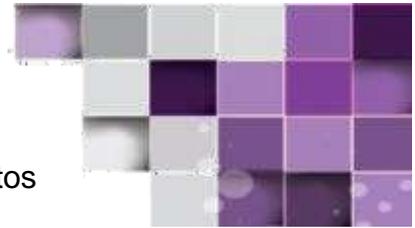
Frecuencias acumuladas

Cuando se quiere establecer el número de observaciones que están por debajo de determinada clase, se suman las frecuencias de una clase con la inmediata superior, a este tipo de frecuencia se le llama frecuencia acumulada de esa clase. Si ese resultado se expresa en porcentaje se denomina **distribución de porcentajes acumulados**.

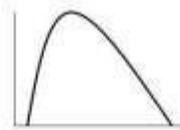
La ojiva, llamado gráfico de porcentajes acumulados, proporciona, hablando de estatura, el porcentaje de individuos cuya estatura es superior o inferior a determinado valor.

Sesgo de la distribución

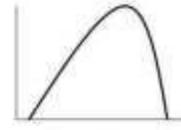
Aunque las distribuciones de frecuencias pueden tener casi cualquier contorno o forma, como lo viste en los histogramas y las gráficas de barras del ejemplo del punto anterior, la mayoría de las distribuciones que encontrarás en la práctica se pueden describir mediante alguno de los tipos siguientes.



Simétrica



Sesgada a la derecha



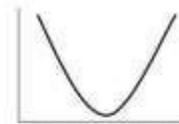
Sesgada a la izquierda



En forma de J



En forma de J invertida



En forma de U



Bimodal



Multimodal

Curvas de frecuencias

Fuente: Universidad Autónoma del Estado de Hidalgo, 2011. uaeh.edu.mx/

Cierre de la Unidad



Indicador

Fuente: Master isolated images, 2011. freedigitalphotos.net

Al concluir con el estudio de esta unidad y haber realizado las actividades correspondientes, has aprendido a identificar un problema y obtener datos para sintetizarlos a través de frecuencias y visualizarlos a partir de gráficas, lo cual te permite tener una primera aproximación al comportamiento de los datos en el contexto del problema abordado.

Fuentes de consulta

A continuación, se enlistan las referencias que fueron consultadas para construir y fortalecer el desarrollo de la segunda unidad.

- Bacchini, R. D., Vázquez, L. V., Bianco, M. J., y Llamas, J. (2018). Introducción a la Probabilidad y a la Estadística. Universidad de Buenos Aires: Facultad de Ciencias Económicas. http://bibliotecadigital.econ.uba.ar/download/libros/Bacchini_Introduccion-a-la-probabilidad-y-a-la-estadistica-2018.pdf
- Centro de Investigación y Desarrollo. (2009). Guía para la presentación de gráficos estadísticos. Instituto Nacional de Estadística e informática. <https://www.inei.gob.pe/media/MenuRecursivo/metodologias/libro.pdf>
- Grupo Compás. (2020). Apuntes de clase estadística básica. Estadística y probabilidad: Teoría y ejercicios (Primera edición). Recuperado de: <http://142.93.18.15:8080/jspui/bitstream/123456789/511/1/libro.pdf>
- Matemáticas profe Alex. (2022). Regla de Sturges | tabla de frecuencias (Video). YouTube. Recuperado de <https://www.youtube.com/watch?v=5fe9VJb85H8&t=2s>
- Mendenhall, W., Beaver, R. J. y Beaver, B. M. (2010). Introducción a la probabilidad y estadística. (Décima tercera edición). México: Editorial Cengage Learning. <https://www.fcfm.buap.mx/jzacias/cursos/estad2/libros/book5e2.pdf>
- Montgomery, D. C. y Runger, G. C. (1996). *Probabilidad y Estadística aplicadas a la ingeniería*. (Cuarta edición). México: McGraw-Hill.
- Porras Velázquez, A. (s.f). Recordando las medidas de tendencia central, de dispersión y de la forma. Recuperado de <https://centrogeo.repositorioinstitucional.mx/jspui/bitstream/1012/161/1/17-Recordando%20las%20Medidas%20de%20Tendencia%20Central%2C%20de%20Dispersi%C3%B3n%20y%20de%20la%20Forma%20-%20%20Diplomado%20en%20An%C3%A1lisis%20de%20Informaci%C3%B3n%20Geoespacial.pdf>
- Torres, M., Paz, K. y Salazar, F. (s.f.). *Métodos de recolección de datos para una investigación*. Disponible en: https://fgsalazar.net/LANDIVAR/ING-PRIMERO/boletin03/URL_03_BAS01.pdf
- Walpole, R. E., Myers, R. H. et al. (2007). *Probabilidad y Estadística para Ingeniería y ciencias*. (Octava edición). México: Pearson Educación.

Complementaria

- Vitutor. (2012). *Estadística y probabilidad*. Disponible en: <http://www.vitutor.com/estadistica.html>