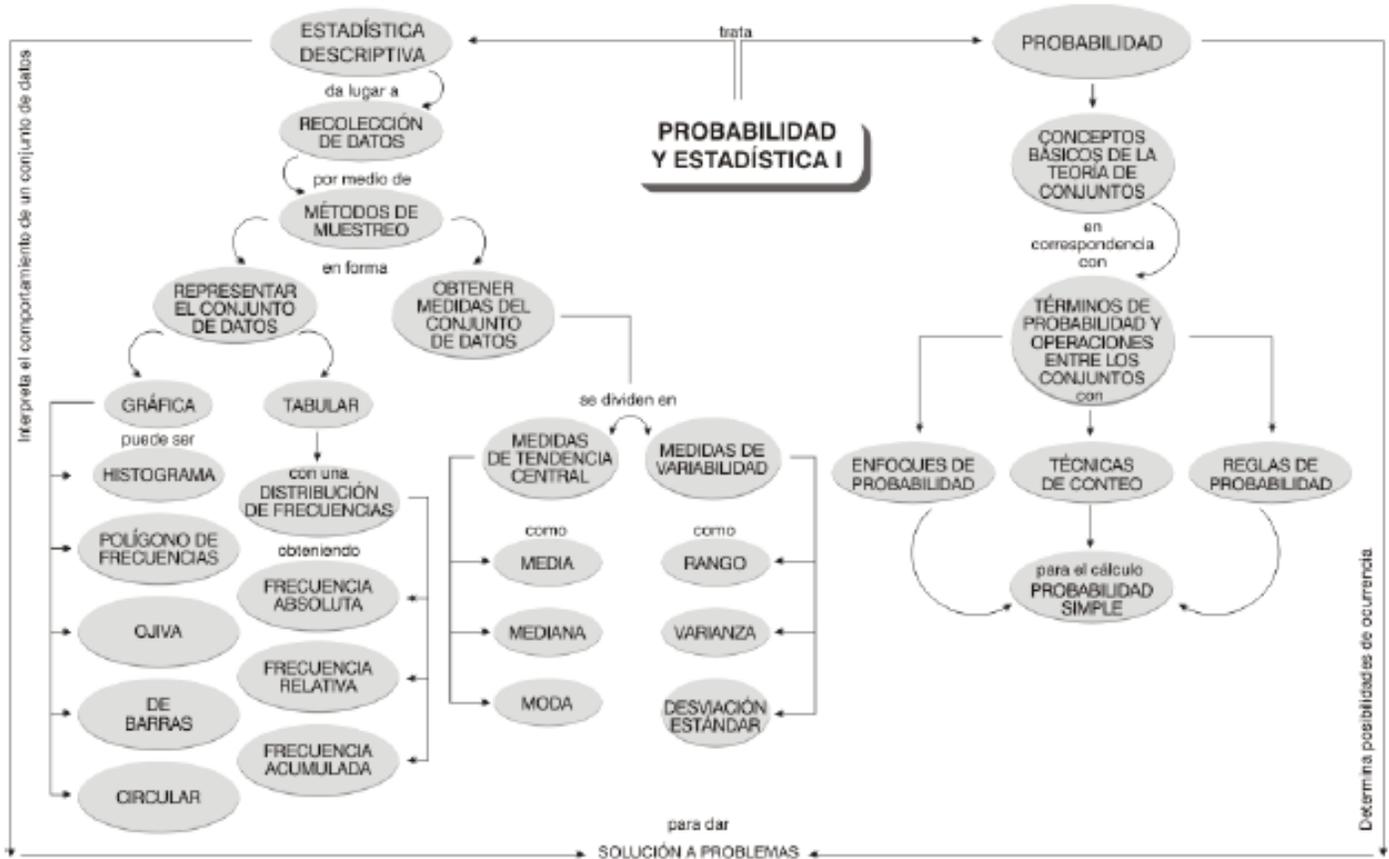


ESTADÍSTICA DESCRIPTIVA

MAPA CONCEPTUAL



Definiciones

Estadística: Es la rama de las matemáticas que se ocupa de recolectar, organizar, presentar, analizar e interpretar **datos** para ayudar a una toma de decisiones más efectiva.

La estadística se clasifica en dos grandes ramas la estadística descriptiva y la estadística inferencial.

Estadística Descriptiva

Conjunto de métodos para organizar, resumir y presentar los datos de manera informativa.

Ejemplo 1: El censo del INEGI reveló que la edad promedio de los mexicanos es de x años.

Ejemplo 2: La producción representativa durante el mes es de 500 toneladas diarias.

Estadística inferencial

Conjunto de métodos probabilísticos y estadísticos utilizados para saber algo acerca de un todo, basándose en una pequeña parte de este.

Ejemplo 1: La empresa de consulta Mitofsky concluyo que el partido político X es el referido por los mexicanos después de realizar una encuesta con algunos votantes.

Ejemplo 2: La empresa Nissan decidió rechazar toda la producción de motores del día de hoy al encontrar que un determinado número de ellos se encontraba defectuoso.

Población

Se le llama Población a la cantidad total de cualquier conjunto completo de datos, objetos, individuos o resultados que tengan alguna característica en común que se va a observar o analizar en un problema o experimento. Denotaremos al tamaño de la población por “N”.

En nuestro ejemplo 1 se considera como población a todos los conductores de automóviles. Así:

$N = 750,000$

Muestra

Se le llama Muestra a cualquier subconjunto de elementos de la población. El interés de la Estadística es proporcionar métodos que permitan elegir una muestra de datos representativos destinado a suministrar información a cerca de una población, será fundamental que los elementos deben tener todas las características de la población.

Denotamos al tamaño de la muestra por “n”. En nuestro ejemplo una muestra podría ser: 500 conductores elegidos al azar, en este caso quedará

Muestra $n = 500$ y

Población $N = 750,000$

Variable

Se le llama Variable a la cualidad o cantidad medible de cualquier suceso o acción que presente o experimente un cambio, la podemos representar mediante un símbolo ($X, Y, Z, \alpha, \beta, \gamma, \delta$) y al cual se le puede asignar un valor cualquiera de un conjunto determinado de datos.

Variable Aleatoria

Le llamamos Variable Aleatoria a aquella variable cuyos cambios no pueden ser determinados antes de que estos se presenten; es decir, están destinados a la suerte. También se le conoce como Variable Probabilista, Cabalística, de Azar o a la Suerte.

Tipos de variable

Para su estudio, las variables aleatorias se han clasificado según la naturaleza de los valores que toman en:

1. Variables Numéricas:

- a) Variables Numéricas Discretas

b) Variable Numérica Continua

2. Variables Categóricas:

a) Variables Categóricas Nominales

b) Variables Categóricas Ordinales

Variables Numéricas o Cuantitativas

Son aquellas que se identifican o se les puede asignar un valor numérico o que corresponden a aspectos que son medibles.

Ejemplo: Tiempo de uso, precio, tamaño, velocidades, número de hijos de una familia, número de carros que circulan por determinada calle, alturas, pesos, tallas, temperaturas, tiempo de vida de una persona, cantidad de azúcar para endulzar un café, medida de sombreros, etcétera.

Las variables numéricas se dividen en:

A) *Variables Numéricas Discretas:* son aquellas que solamente toman valores enteros con rango finito.

Ejemplo: Número de hijos en cada familia de una colonia de la ciudad, talla de calzado de cada alumno de un grupo escolar, la cantidad de alumnos por grupo, etc.

B) *Variable Numérica Continua:* son aquellas que pueden tomar cualquier valor entre dos valores dados. Es decir, el rango contiene no sólo valores enteros sino un intervalo (finito o infinito) de valores reales (esto es, que puede ser fraccionario, decimal o irracional).

Ejemplo: El tiempo de vida de una persona, la cantidad de azúcar para endulzar un café, el nivel de hemoglobina de los habitantes de una colonia, la temperatura ambiental durante un día, etcétera.

Variables Categóricas o Cualitativas

Son aquellas a las que no se les puede asignar o identificar con un valor numérico, sino con un aspecto, cualidad o característica que las distinga y que no se pueden medir sino solo observar, a ese aspecto, cualidad o característica se le llama categoría.

Ejemplos: Marca, tipo de sangre, deporte preferido, el estado en general de cualquier cosa, idioma, nacionalidad, colores, cabello o piel, himnos nacionales, sexo, estado de ánimo, clima, etcétera.

En las variables categóricas, un elemento no puede estar en dos o más categorías a la vez, lo cual las hace excluyentes y además no puede haber elementos de la población que no pertenezcan a alguna categoría, lo cual las hace exhaustivas.

Las variables categóricas se dividen en:

A) *Variables Categóricas Nominales:* son aquellas a las que no se les puede asignar un orden, es decir que sólo permite clasificación en categorías por mención de ésta.

Ejemplo: La nacionalidad de una persona, idioma, sexo, himnos nacionales.

- B) Variables Categóricas Ordinales:** son aquellas que además de clasificar a los elementos en distintas categorías les podemos asignar un orden o que podemos ordenar de acuerdo a cierta característica.

Ejemplo: El estado de salud de una persona; que podemos ordenarla según la urgencia del caso, el color de algún objeto según la tonalidad desde muy clara a más oscuro; que podemos ordenarlo de acuerdo a la intensidad del color, el grado militar, puesto en la empresa, día de la mamá, meses del año, etcétera.

Datos

Se le llama Datos a las agrupaciones de cualquier número de observaciones relacionadas. Para que se considere un dato estadístico debe tener dos características:

- a) Que sean comparables entre sí.
- b) Que tengan alguna relación.

La recolección de información o recopilación de datos estadísticos se divide en:

- A) Datos Internos:** son aquellos datos que no necesitan de observaciones adicionales al experimento; es decir, no es necesario buscar características que proporcionen información adicional acerca del experimento.

Ejemplo: Las calificaciones de un grupo, un experimento químico, etcétera.

- B) Datos Externos:** estos datos pueden ser de dos tipos:

- a) Datos Bibliográficos: son aquellos ya conocidos y que podemos encontrar fácilmente utilizando bibliografía, registros, actas, etcétera, como los datos históricos, censos y otros.
- b) Datos Originales: son aquellos que podemos obtener mediante métodos de recolección, como las encuestas, plebiscitos, referéndum, y nos proporcionan datos reales y certeros.

Para Organizar los datos: existen muchas formas de clasificarlos, en general pueden ser determinados de acuerdo a cuatro elementos que son: Tiempo, lugar, cantidad y cualidad.

Presentación de Datos

Después de la organización de los datos, la información se resume en Tablas Estadísticas con base en arreglos formados de renglones y columnas, adecuados según cronología, geografía, análisis cuantitativo o cualitativo.

Los principales elementos de una tabla estadística son: Título, unidades, encabezado, cuerpo o contenido, nota de pie y referencias; la información contenida en una tabla estadística también se puede presentar mediante graficas, siendo las más comunes las de líneas, barras, pictográficas, cronogramas, circulares o de pastel, histograma y polígono de frecuencias.

Experimento

Se le llama Experimento a toda acción o prueba que se realiza con el fin de observar su resultado. Existen dos tipos de experimentos, que son:

- a) Experimento Determinista:** son aquéllos en los que se puede predecir con certeza su resultado antes de que éste se presente.

Ejemplo: Al lanzar en un cuarto un libro al aire con el fin de determinar si flota, se queda unido al techo o cae al suelo, sabemos con certeza que el libro caerá al suelo, lo cual lo hace un experimento determinista.

- b) Experimento Aleatorio, Probabilista, casual o de azar: hablar de aleatorio, probabilista, casual o azar es hablar de algo que está determinado a la suerte. Así, decimos que un Experimento Aleatorio ocurre cuando no es posible asegurar el resultado que se va a presentar.

Ejemplo: Al lanzar una moneda al aire no sabemos si el resultado va a ser águila o sello, lanzar un dado, etcétera.

Muestreo

Se llama Muestreo al estudio que se hace de una población por medio de muestras representativas, debidamente elegidas de manera que posea todas las características de una población y de tamaño determinado según la precisión que de ella se quiere obtener en las decisiones y conclusiones estadísticas posteriores.

Se le llama valores estadísticos, estadísticos muestrales o simplemente estadísticos a los valores o cantidades desconocidas que son obtenidas de, o que hacen referencia a las características de una muestra.

Se le llama Parámetro o parámetros poblacionales o simplemente parámetros a los valores o cantidades desconocidas que son obtenidas de, o que hacen referencia a las características de una población

MÉTODOS DE MUESTREO

Censo

Llamamos *Censo* al método de recolección de datos mediante el cual la información se obtiene del estudio de todos los elementos que componen a la población o universo bajo estudio.

Un censo debe cumplir las condiciones de universalidad (censar a todos los elementos de la población) y simultaneidad (realizarse en un momento determinado).

El término censo no sólo se aplica a aquellos estudios que comprenden todas las unidades del país y que se realizan con frecuencia de recolección quinquenal o decenal, como es el caso de los censos de población, económicos, agropecuarios, etcétera, sino también a cualquiera independientemente de su cobertura geográfica, número de unidades de información, o frecuencia de su recolección, siempre que incluya todas las unidades que componen el universo que se investiga.

Población

Una Población es Finita cuando existe una cantidad determinada de elementos por analizar; esto es, una cantidad de elementos, numerable y que en determinado momento finaliza.

Ejemplo: Los habitantes del municipio de Cajeme.

Una Población es Infinita cuando existe una cantidad indeterminada de elementos por analizar; es decir, una cantidad de elementos que aunque los enumeráramos nunca terminaríamos de hacerlo.

Ejemplo:

1. Los valores de temperatura durante un día.
2. Todos los puntos de una línea.
3. Número de alumnos del Cobach del presente y en el futuro.

Métodos de muestreo

Fundamentalmente el muestreo es de los siguientes tipos:

- A) Probabilístico o aleatorios:** tipo de muestreo que se obtiene mediante sorteo de los individuos que la forman teniendo así, cada individuo la misma posibilidad de pertenecer a la muestra, permitiendo calcular el posible error de la muestra. De entre los que destacan, el muestreo aleatorio simple, el sistemático, el estratificado y el de conglomerados.
- B) No probabilística:** tipo de muestreo en el que no es posible estimar la probabilidad de que cada individuo o elemento estará incluido en la muestra, además no permite el cálculo del posible error de la muestra. Pueden ser de tres clases: Accidental o incidental, por cuotas, intencional por conveniencia o de juicio. Aunque este tipo de muestreo no será objeto de estudio en este curso.
- C) El muestreo Aleatorio simple** es el tipo de muestreo en el cual todos y cada uno de los elementos de la población se elige de tal forma que tengan la misma posibilidad de ser seleccionados y pertenecer a la muestra.
- D) El muestreo Sistemático** se utiliza cuando el universo es de gran tamaño o ha de extenderse en el tiempo y requiere de una selección aleatoria inicial de observaciones seguida de otra selección de observaciones, obtenida mediante una constante denominada constante de sistematización

$$Cs = N/n;$$

Donde N = es el tamaño de la población y n = el tamaño de la muestra.

Esta constante nos sirve para determinar cada cuántos elementos o cada cuánto tiempo se debe elegir el siguiente; para ello hay que elegir al azar un número entre 1 y Cs ; de ahí en adelante tomar uno de cada K a intervalos regulares. Es conveniente tener en cuenta la periodicidad del fenómeno.

Ejemplo 2: Para obtener una muestra de suscriptores telefónicos en una ciudad grande, puede obtenerse primero una muestra aleatoria de los números de las páginas del directorio telefónico; al elegir el vigésimo nombre de cada página obtendríamos un muestreo sistemático, también podemos escoger un nombre de la primera página del directorio y después seleccionar cada nombre del lugar número cien a partir del ya seleccionado. En este caso, podríamos seleccionar un número al azar entre los primeros 100; suponiendo que el elegido es el 40, entonces seleccionamos los nombres del directorio que corresponden a los números 40, 140, 240, 340 y así sucesivamente.

- E) El muestreo sistemático** suele ser más preciso que el aleatorio simple, ya que recorre la población de un modo más uniforme.
- F) En el tipo de muestreo Estratificado** se involucra la división previa de la población en subgrupos, clases o estratos que se suponen más homogéneos, y a los cuales se le asigna una cuota que determina el número de miembros del estrato que compondrán la muestra, estos son escogidos mediante muestreo aleatorio simple. Según la cantidad de elementos de la muestra que se han de elegir de cada uno de los estratos, existen dos técnicas de muestreo estratificado:
 - a. *Asignación proporcional:* el tamaño de cada estrato en la muestra es proporcional a su tamaño en la población.

- b. Asignación óptima: la muestra recogerá más individuos de aquellos estratos que tengan más variabilidad. Para ello es necesario un conocimiento previo de la población.

Ejemplo 3: Suponiendo un estudio sobre la población de estudiantes de cierto plantel del COBACH, en el que a través de una muestra de 10 de ellos queremos obtener información sobre el uso del lápiz labial. Pero reflexionando sobre que el comportamiento de la población con respecto a esta característica no es homogéneo, podemos dividir a la población en dos estratos:

Estudiantes masculinos 40%.

Estudiantes femeninos 60%.

De modo que la asignación proporcional a esta muestra es en función de sus respectivos tamaños (6 varones y 4 mujeres).

También se puede observar que el comportamiento de los varones con respecto a la característica en estudio es muy homogéneo y diferenciado del grupo de las mujeres que es muy variable. De modo que la *asignación óptima* de una muestra de 10 alumnos, nos indica que es más conveniente elegir más individuos en los grupos de mayor variabilidad.

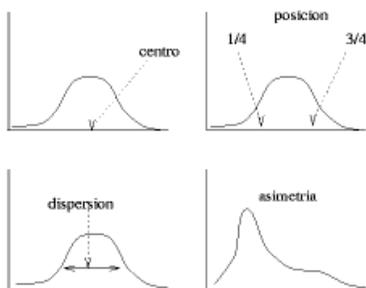
De la cual obtendríamos mejores resultados estudiando una muestra de 1 varón y 9 mujeres.

- G)** Se le llama muestreo **Por conglomerados** al dividir primero la población en grupos o conglomerados convenientes para el muestreo, seleccionando de cada uno de ellos una porción, al azar o por un método sistemático. Bajo este método, aunque no todos los grupos son muestreados, cada grupo tiene una igual probabilidad de ser seleccionado.

Por lo tanto, la muestra es aleatoria. Una muestra por conglomerados, usualmente produce un mayor error muestral que una muestra aleatoria simple del mismo tamaño; sin embargo, puede ser obtenida dentro de un corto período de tiempo y a bajo costo.

Además, una muestra por conglomerados ofrece la misma precisión en la estimación que una muestra aleatoria simple, si la variación de los elementos individuales dentro de cada conglomerado es proporcionalmente tan grande como la de la población.

DISTRIBUCIÓN DE FRECUENCIAS



En los gráficos podemos observar la característica natural de la distribución de datos a concentrarse hacia el centro y la variabilidad de los datos en los extremos.

La Distribución o Tabla de Frecuencias: Es la representación conjunta de los datos en forma de tabla o subgrupo de datos correspondientes a un fenómeno en estudio y su ordenamiento en base al número de observaciones que corresponden a cada dato o a cada grupo de datos, adecuados según cronología, geografía, análisis cuantitativo o cualitativo.

Los principales elementos de una tabla estadística son: Título, unidades, encabezado, cuerpo o contenido, nota de pie valores y referencias.

Se elabora colocando en la primera columna los datos diferentes o subgrupos de datos (llamados clases o intervalos de clase) y en la columna siguiente el número de observaciones que corresponden a cada dato o a cada grupo de datos llamada frecuencia).

Una tabla de este tipo dará, en forma abreviada, una información completa acerca de la distribución de los valores observados.

Estas tablas facilitan el uso de los métodos gráficos y aritméticos.

La presentación de los datos en forma ordenada, por medio de una tabla, dependerá de los datos de que se trate, y si estos son cualitativos o cuantitativos como se muestra a continuación:

Ordenamiento de Datos

Cualitativos:

- Alfabético A – Z
- Alfabético Z – A
- Del más al menos repetido
- Del menos al más repetido

Cuantitativos:

- Creciente (menor al mayor)
- Decreciente (mayor al menor)

EJEMPLO 1: Cierta universidad realizó un experimento sobre el coeficiente intelectual (C.I.) de sus alumnos, para lo cual aplicó un examen de C.I. a un grupo de 20 alumnos escogidos al azar, obteniendo los siguientes resultados:

119, 109, 124, 119, 106, 112, 112, 112, 112, 109, 112,
124, 109, 109, 109, 106, 124, 112, 112, 106.

Toda vez que se tienen los datos, se ordenan de menor a mayor o viceversa.

106, 106, 106, 109, 109, 109, 109, 109, 112, 112, 112,
112, 112, 112, 112, 119, 119, 124, 124, 124

Datos	Repeticiones
106	3
109	5
112	7
119	2
124	3

EJEMPLO 2. Se preguntó a un grupo de alumnos de primer año del Plantel “X”, por la asignatura de su preferencia, arrojándose los siguientes resultados:

Asignaturas

Mate Social Taller Quím. Infor Mate Inglés
 Mate Quím. Infor Inglés Ética Inglés Social
 Inglés Ética Mate Taller Quím. Mate Taller
 Social Mate Inglés Infor Inglés Ética Infor
 Mate Inglés Infor Ética Quím. Taller Inglés
 Social Inglés Ética Taller Infor Quím. Taller
 Taller Infor Mate Quím. Infor Mate Infor
 Inglés

Asignatura	Repeticiones (frecuencias)
Ética y valores	5
Informática	9
Inglés	10
Matemáticas	9
Química	6
Sociales	4
Taller de lectura	7
Total	50

Distribución de frecuencias

1. Frecuencia Absoluta de un dato: Es el número de veces que se repite ese dato, también se presenta la frecuencia absoluta de un intervalo que se refiere al número de datos que pertenecen a ese intervalo. La denotaremos por ***f***.
2. Frecuencia Absoluta Acumulada: Hasta un dato específico, es la suma de las frecuencias absolutas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia acumulada. De un intervalo es la suma de las frecuencias absolutas de todos los intervalos de clase anteriores.
3. Frecuencia acumulada. La última frecuencia absoluta acumulada deberá ser igual al número total de datos. La denotaremos por ***fa***.
4. Frecuencia Relativa: De un dato, se obtiene al dividir la frecuencia absoluta de cada dato entre el número total de datos. De un intervalo se obtiene al dividir la frecuencia absoluta de cada intervalo entre el número total de datos. La denotamos por ***fr***.
5. Frecuencia Relativa Acumulada: Hasta un dato específico de la observación, es la suma de las frecuencias relativas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia relativa acumulada de un intervalo es la suma de las frecuencias relativas de todos los intervalos de clase anteriores incluyendo la frecuencia del intervalo mismo del cual se desea su frecuencia relativa acumulada. La última frecuencia relativa acumulada deberá ser igual a la unidad. La denotaremos por ***fra***.

Construcción de distribución o tabla de frecuencias para datos no agrupados y agrupados.

Datos no agrupados

Datos diferentes: Consideraremos como un dato diferente, a cada uno de los distintos datos que se presentan en la muestra, los denotaremos por x_i , y al número total de datos diferentes lo denotaremos por m .

Cuando el tamaño de la muestra (n) es finito y el número de datos diferentes es pequeño (consideraremos pequeño $k \leq 10$), es fácil hacer un análisis de los datos tomando cada uno de los datos diferentes y ordenándolos cualitativa o cuantitativamente.

Ejemplo:

Utilicemos los datos de los EJEMPLOS 1 y 2.

Coeficiente Intelectual				
x_i	f	f_a	f_r	f_{ra}
106	3	3	0.15	0.15
109	5	8	0.25	0.40
112	7	15	0.35	0.75
119	2	17	0.10	0.85
124	3	20	0.15	1.00
Total	20		1.00	

Asignatura de Preferencia				
x_i	f	f_a	f_r	f_{ra}
Ética y valores	5	5	0.1	0.1
Informática	9	14	0.18	0.28
Inglés	10	24	0.2	0.48
Matemáticas	9	33	0.18	0.66
Química	6	39	.12	0.78
Sociales	4	43	0.08	0.86
Taller de lectura	7	50	0.14	1.00
Total	50		1.00	

Datos agrupados

Cuando el tamaño de la muestra es considerable o grande y los datos numéricos son muy diversos ($n > 15$), conviene agrupar los datos de tal manera que permita establecer patrones, tendencias o regularidades de los valores observados. De esta manera podemos condensar y ordenar los datos tabulando las frecuencias asociadas a ciertos intervalos de los valores observados.

Intervalos de Clase: Son los intervalos en los que se agrupan y ordenan los valores observados. Cada uno de estos intervalos está delimitado (acotado) por dos valores extremos que les llamamos límites.

Pasos a seguir para construir intervalos de frecuencia.

1. Determinar la cantidad de intervalos apropiada

La selección del número adecuado de intervalos y los límites entre ellos dependen del criterio o experiencia de quien realiza el estudio. Sin embargo, existen reglas empíricas para calcular el número de intervalos; la más empleada es la **Regla de Sturges**, cuya expresión es:

$$K = 1 + 3.3 \log n$$

Dónde: K = Número de intervalos el cual siempre debe ser un número entero. Razón por la cual se deberá redondear el resultado al entero más cercano.

n = Número de datos.

\log = logaritmo en base 10.

Otra regla utilizada es la de **Vellemán** que establece que el número de Intervalos se obtiene de la raíz cuadrada del número de datos; es decir $K = \sqrt{n}$, recomendable para tamaños de muestra pequeños ($n < 50$)

El número de *intervalos* determinado mediante cualquier regla se aproxima al valor entero más cercano pero deberá ser responsabilidad de quien realiza el estudio, pudiendo utilizar éste en ocasiones uno menor o mayor al obtenido por cualquier regla, si esto le permite tener intervalos con la misma amplitud. Sin embargo, la mayoría de las reglas subestiman el número de intervalos.

2.- Calcular el rango de los datos.

Llamamos rango al número de unidades de variación presente en los datos recopilados y se obtiene de la diferencia entre el dato mayor y el dato menor. Se representa con la letra R.

$$R = \text{Dato mayor} - \text{dato menor}$$

3.- Obtención de la amplitud o anchura que tendrá cada intervalo.

Se encuentra dividiendo el rango por el número de intervalos regularmente es de 5 a 6. Se representa con la letra A de tal manera que $A = \frac{R}{K}$.

4.- Construcción de los intervalos.

Los intervalos de clase son conjuntos numéricos y deben ser excluyentes y exhaustivos; es decir, si un dato pertenece a un intervalo determinado, ya no podrá pertenecer a otro, esto quiere decir excluyentes y además todos y cada uno de los datos deberá estar contenido en alguno de los intervalos, esto les da el valor de exhaustivos.

Las dos caracteres mencionadas anteriormente se logran construyendo intervalos cerrados por la izquierda y abiertos por la derecha; esto se simboliza a través del uso de corchetes y paréntesis respectivamente. Por razones naturales, el último intervalo será cerrado por ambos extremos.

El primer intervalo se construye de la siguiente manera: Habrá de iniciar con el dato menor, el cual será el extremo inferior del intervalo; el otro extremo se obtiene de la suma del dato menor y la amplitud, con este mismo valor iniciamos el segundo intervalo, del cual el segundo extremo se encuentra sumando al valor anterior la amplitud y este proceso se repite sistemáticamente hasta completar el total de intervalos indicado por la regla elegida, por ejemplo la de Sturges.

Los valores extremos o límites de intervalo.

Los intervalos de clase deben estar definidos por límites que permitan identificar plenamente si un dato pertenece a uno u otro intervalo. Estos límites son los valores extremos de cada intervalo.

Límite inferior: Es el valor menor de cada intervalo, se denota por L_i

Límite superior: Es el número mayor de cada intervalo, se denota por L_s .

También será muy útil conocer y calcular la **Marca de Clase** (MC) de cada intervalo: Se refiere al Punto Medio del intervalo y a través de él representaremos a todo el intervalo y una de las maneras de calcularla es promediando los valores límite de cada intervalo, su fórmula es:

$$MC = \frac{L_i + L_s}{2}$$

EJEMPLO 2.1.4: Un grupo de investigadores pertenecientes a la secretaría de seguridad pública, tomó una muestra aleatoria de las velocidades (km/h) registradas por 30 vehículos en el trayecto Hermosillo a Ures, con el fin de establecer nuevos límites máximos de velocidad para una carretera. La muestra arrojó los datos siguientes:

90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118,
100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

Toda vez que se tienen los datos, se recomienda ordenarlos de menor a mayor o viceversa

90, 90, 95, 95, 96, 97, 98, 98, 99, 99, 100, 100, 100, 104,
105, 106, 108, 111, 112, 112, 114, 114, 115, 116, 118, 119, 120

Ahora llevamos a la práctica los pasos descritos anteriormente para la construcción de los intervalos.

1º obtendremos el número de intervalos que vamos a utilizar, para lo cual empleamos la Regla de Sturges:

$$K = 1 + 3.3 \log(30) = 1 + 3.3 (1.4771212547) = 1 + 4.87 = 5.87 \approx 6$$

2º calculamos el rango de variación.

$$R = 120 - 90 = 30$$

3º obtenemos la amplitud de cada intervalo de clase como sigue:

$$Ac = \frac{30}{6} = 5$$

4º construimos los intervalos: el primero de ellos inicia con 90 que es el extremo inferior que, sumado a 5 obtenemos 95, que será el extremo superior; este extremo será el inferior del segundo intervalo; y al sumar nuevamente la amplitud tendremos 100 que será el extremo superior y así sucesivamente hasta completar los 6 intervalos., que se muestran enseguida:

[90 - 95), [95 - 100), [100 - 105), [105 - 110), [110 - 115) y [115 - 120]

Los corchetes expresan que el valor extremo se incluye en el intervalo y los paréntesis dan a entender que el valor extremo del intervalo no se incluye en el.

Para la construcción de distribuciones de frecuencias, contamos el número de datos que le corresponden a cada intervalo; es decir obtenemos las frecuencias absolutas y de estas podemos generar los demás tipos de frecuencias y presentarlas en una tabla de resumen como la que a continuación se muestra:

Distribuciones de frecuencias para las velocidades

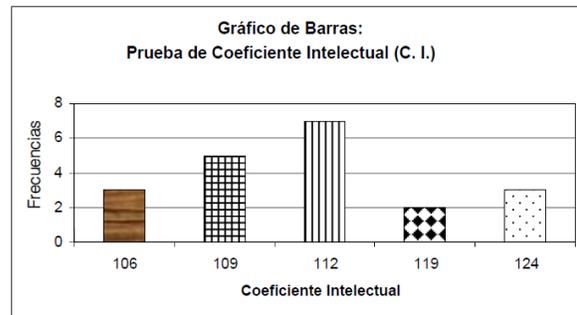
x_i Intervalos de Clase	f	f_a	f_r	f_{ra}	m_c
[90 – 95)	2	2	0.07	0.07	92.5
[95 – 100)	8	10	0.27	0.34	97.5
[100 – 105)	5	15	0.17	0.51	102.5
[105 – 110)	4	19	0.13	0.64	107.5
[110 – 115)	6	25	0.20	0.84	112.5
[115 – 120]	5	30	0.16	1.00	117.5
Total	30		1.00		

REPRESENTACIÓN GRÁFICA

Gráfica de Barras:

Es un método gráfico que consta de dos ejes: Uno horizontal, en el que se representan los valores (Eje de los datos) utilizando barras verticales en forma rectangular y de la misma amplitud, y un eje vertical, en el cual la frecuencia representa la altitud que tendrá la barra rectangular (Eje de las frecuencias), las barras van separadas la misma distancia unas de otras y para distinguirlas puede utilizarse distintos colores o entramados según se considere.

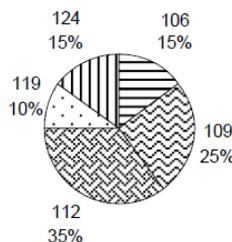
Coeficiente Intelectual				
x_i	f	f_a	f_r	f_{ra}
106	3	3	0.15	0.15
109	5	8	0.25	0.40
112	7	15	0.35	0.75
119	2	17	0.10	0.85
124	3	20	0.15	1.00
Total	20		1.00	



Gráfica Circular de Pastel o también llamada del 100%:

Utilicemos las formulas correspondientes para calcular el porcentaje con la regla de tres, como anteriormente lo realizamos.

Gráfica Circular:
Prueba de Coeficiente Intelectual

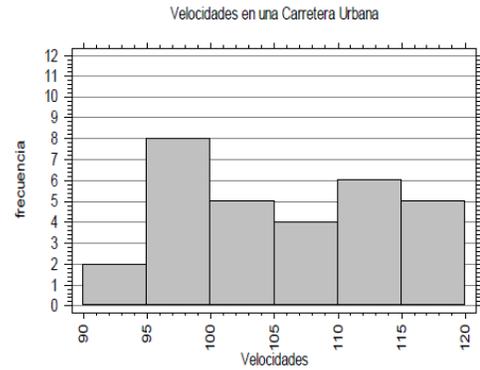


Histograma:

Las barras no van separadas, y que se rotula con los límites inferiores de cada clase o intervalo excepto el último que deberá llevar también el límite superior, centradas en la marca de clase.

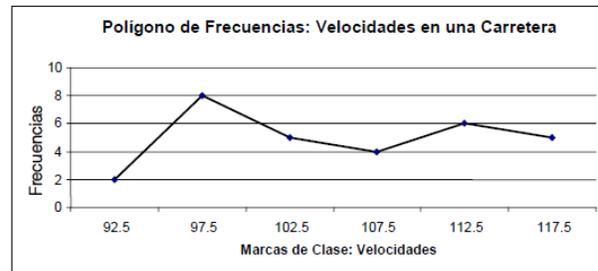
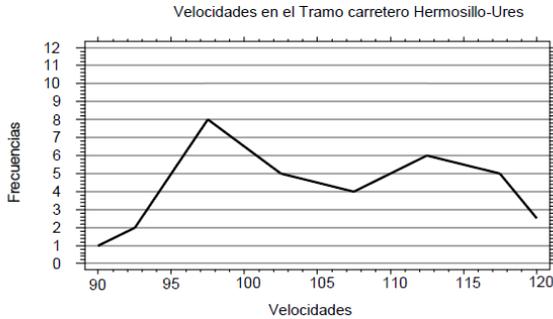
Distribuciones de frecuencias para las velocidades

x_i Intervalos de Clase	f	f_a	f_r	f_{ra}	m_c
[90 – 95)	2	2	0.07	0.07	92.5
[95 – 100)	8	10	0.27	0.34	97.5
[100 – 105)	5	15	0.17	0.51	102.5
[105 – 110)	4	19	0.13	0.64	107.5
[110 – 115)	6	25	0.20	0.84	112.5
[115 – 120]	5	30	0.16	1.00	117.5
Total	30		1.00		



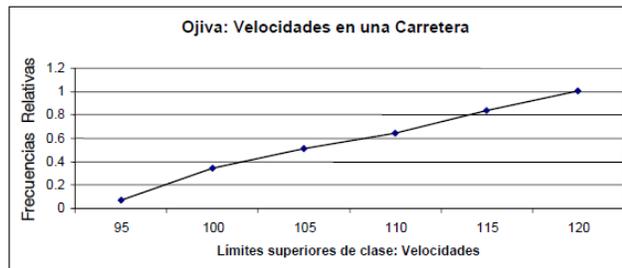
Polígono de Frecuencias:

es una gráfica del tipo de las gráficas de líneas trazadas sobre las marcas de clase, (de ahí el nombre de polígono), y se traza uniendo con segmentos de recta, de izquierda a derecha, las parejas ordenadas que se forman, al considerar como abscisa la marca de clase (eje horizontal) y como ordenada la frecuencia del intervalo representado (eje vertical).



Gráfica de Frecuencias Acumuladas u Ojiva:

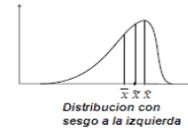
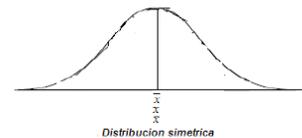
Es un gráfico que igual al histograma y polígono de frecuencias se utiliza para el análisis y representación de variables continuas, sólo que en vez de utilizar las frecuencias absolutas, por sus características se construye uniendo con segmentos de recta, de izquierda a derecha, las parejas ordenadas que se forman, al considerar como abscisa los límites superiores de cada intervalo (eje horizontal) y como ordenada las frecuencias relativas acumuladas hasta cada intervalo representado (eje vertical)



Medidas de Centralización y Variabilidad

De los muchos aspectos de los datos, que intentamos representar numéricamente con estadísticas, dos son los más importantes:

Las Medidas de Centralización



Medidas de centralización para datos no agrupados

a) Media Aritmética. La medida más evidente que podemos calcular para describir un conjunto de observaciones numéricas es su valor medio. La media no es más que la suma de todos los valores de una variable dividida entre el número total de datos de los que se dispone. Siendo su fórmula la siguiente:

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

Donde:

Σ Símbolo de sumatoria que indica que se deberá sumar lo que aparezca a derecha, es decir, x.

X Cada uno de los datos obtenidos de la muestra.

n Número total de datos.

Como ejemplo, consideremos 10 pacientes de edades 21 años, 32, 15, 59, 60, 61, 64, 60, 71, y 80. La media de edad de estas personas será de:

$$\bar{x} = \frac{21+32+15+59+60+61+64+60+71+80}{10}$$

$$\bar{x} = 52.3 \text{ años}$$

VENTAJAS Y DESVENTAJAS DE LA MEDIA ARITMÉTICA

Ventajas

Es de fácil cálculo e interpretación sencilla. Es la más utilizada y es útil en muchos desarrollos matemáticos.

Desventajas

La principal desventaja se presenta cuando alguno o los dos valores extremos de la muestra son desproporcionados respecto al resto de los datos, sobre todo cuando éstos son poco numerosos. En este caso la media se aleja de la realidad; es decir, deja de ser representativa de los datos.

b) Mediana

En esta medida, los datos u observaciones equidistantes o que se encuentran más en medio de todo el conjunto de datos.

La mediana del ejemplo anterior, sería el valor que deja a la mitad de los datos por encima de dicho valor, y a la otra mitad por debajo, es decir el 50 % por arriba y el 50% por debajo del conjunto de datos.

Para obtener la mediana para datos agrupados, primeramente deberemos ordenar los datos en forma ascendente o descendente observando la siguiente secuencia de datos: 5, 21, 32, 59, 60, 60, 61, 64, 71, 80. (Datos ordenados) Como quiera que sea, en este ejemplo, el número de datos u observaciones es par (10 personas), los dos valores que se encuentran en medio son 60 y 60. Si realizamos el cálculo para la media nos dará:

$$5, 21, 32, 59, \textcircled{60, 60}, 61, 64, 71, 80. \text{ (Número de observaciones par)}$$

$$\bar{x} = \frac{60 + 60}{2} \qquad \bar{x} = \bar{x} = 60 \text{ años}$$

Si al ejemplo anterior le agregamos un paciente más de 55 años entonces la mediana se determinará como el dato u observación que se encuentra más en el medio es decir:

$$5, 21, 32, 55, 59, \textcircled{60}, \textcircled{60}, 61, 64, 71, 80. \text{ Entonces la mediana } (\bar{x}) = 60 \text{ años}$$

Si la media y la mediana son iguales, la distribución o conjunto de datos de la variable es simétrica. Sin embargo, la media es muy sensible a la variación de las puntuaciones, y la mediana es menos sensible a dichos cambios.

Geoméricamente la mediana es el valor de la variable que corresponde a la vertical que divide al histograma en dos áreas iguales. Cuando determinados valores de un conjunto de datos u observaciones son muy grandes o muy pequeños con respecto a los demás, entonces la media aritmética se puede distorsionar y perder su carácter representativo, en esos casos, es conveniente utilizar la mediana como medida de tenencia central.

VENTAJAS Y DESVENTAJAS DE LA MEDIANA	
Ventajas	a) La Mediana no se ve afectada por los valores extremos, por lo que la podemos utilizar en aquellos casos en que la media aritmética no es útil.
Desventajas	a) La más importante, es que no podemos hacer cálculos adicionales con la mediana. b) No utiliza mucha información de un conjunto de datos. c) Finalmente, al menos que dispongamos de una computadora o unos esclavos, no es fácil ordenar un conjunto grande de números. En este caso, la mediana no es fácil de calcular.

c) **La moda** (x) se suele definir como el valor más frecuente. En el caso de una variable no agrupada, es el valor de la variable que más se repite.

Ejemplo 1: En el caso del ejemplo anterior, 5, 21, 32, 59, 60, 60, 61, 64, 71, 80. La moda será:

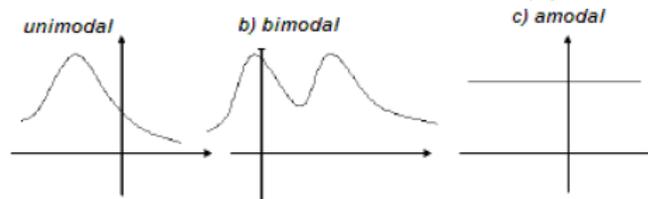
$$\hat{x} = 60 \text{ años. (Unimodal)}$$

Ejemplo 2: Determinar la moda del siguiente conjunto de datos 1, 2, 3, 4, 4, 5, 2, 1, 3, 4, 2, 3, 4, 6, 3, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 6.

$$\hat{x} = 3 \text{ y } 4 \quad \text{Bimodal}$$

Ejemplo 3: Determinar la moda del siguiente conjunto de datos 1, 2, 3, 4, 5, 6, 7, 8, 9. En este caso, como ningún dato se repite será **amodal**.

Gráficamente eso se puede reflejar mediante el análisis de un histograma de frecuencias.



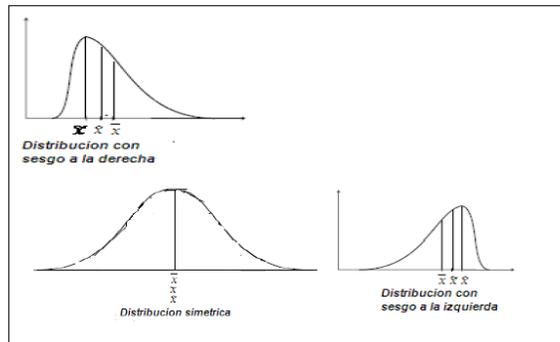
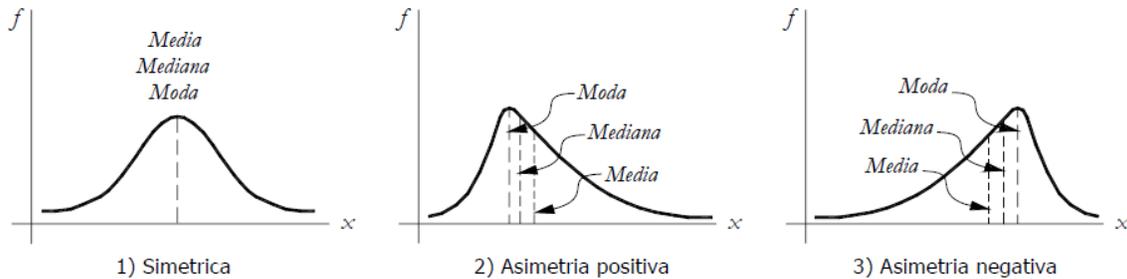
Grafica 1

En el caso de que la distribución o conjunto de datos tenga una moda, se dirá que el conjunto de datos es unimodal; si tiene dos modas, se llamará bimodal; más de dos modas, se le llamará polimodal; y en caso que no tenga ninguna moda se denominará amodal.

VENTAJAS Es la que más fácilmente se determina, puesto que la obtenemos por inspección y no por cómputo.
--

Posiciones relativas de la Media, la Mediana y la Moda.

Si la *Media*, *Mediana* y *Moda* se localizan en el centro y son siempre iguales, la distribución es simétrica. Ello significa que si se doblara por la mitad al polígono de frecuencia, ambos lados tienen la misma forma. El punto más alto de la curva, corresponde a la moda. Como la curva es simétrica, la mediana corresponde al punto en que la distribución se parte a la mitad. Las frecuencias más altas se compensan con las más bajas; y así, la Media, Mediana y Moda coinciden, lo que significa que cualquiera de las tres medidas *es adecuada* para representar una distribución. Conforme la distribución se hace menos simétrica o sesgada, la relación entre los tres promedios cambia.



Gráfica 2 Tipos de distribuciones

Relación entre la media, la mediana y la moda

Las diferencias entre los valores de la media, la mediana y la moda permiten saber la forma de la curva de frecuencias en términos de asimetría.

- a) Para una distribución unimodal simétrica, el valor de la media, la mediana y la moda es igual.
- b) Para una distribución asimétrica positiva, la media es el mayor valor de los tres y la mediana es mayor que la moda, pero menor que la media.
- c) Para una distribución asimétrica negativa, la media es el menor valor de los tres y la mediana es inferior a la moda, pero mayor que la media.
- d) El coeficiente de asimetría de Pearson, es una medida conocida de asimetría que utiliza la diferencia observada entre la media y la mediana de un grupo de valores.

Medias de tendencia central para datos agrupados

Ejemplo: La siguiente distribución de datos representan las calificaciones de 30 alumnos.

	88	77	49	38	100
	95	60	75	100	80
	63	69	50	90	82
	65	75	100	95	50
Academia de Matemáticas 20	80	70	60	100	75
	80	100	90	85	75

Para resumir la información del número de estudiantes que obtuvieron una determinada calificación, se hace por medio de una tabla con dos encabezados, lo cual permite exhibir, en forma concisa, el número de veces que se presenta una determinada cantidad en un conjunto de datos.

Los datos anteriores se pueden presentar por medio de una tabla de frecuencias como sigue:

Calificación	38	49	50	60	63	65	69	70	75	77	80	82	85	88	90	95	100
Número de Estudiantes	1	1	2	2	1	1	1	1	4	1	3	0	1	1	2	2	5

Ejemplo. El siguiente conjunto de datos nos representan los pesos en kilogramos de 40 pacientes sometidos a una dieta.

49.0	50,5	53.5	56.0	60.0	67.0	68.6	71.0
49.8	50.6	54.0	57.0	60.0	67.5	69.0	71.5
49.8	50.6	54.3	57.5	63.5	68.0	69.5	72.0
50.0	51.0	55.0	58.3	64.0	68.4	69.6	72.5
50.3	52.0	55.0	59.0	64.0	68.6	70.0	73.0

Los datos se deberán ordenar en forma ascendente o descendente como se prefiera; en nuestro caso se ordenaron en forma ascendente.

Con la información del presente ejemplo, primero hay que decidir en cuantas clases deberá dividirse el intervalo y después su amplitud

Para el caso del ejemplo anterior se aplicara primero la regla de Sturges. Al tomar el ejemplo de los pesos en kilogramos, donde $n=40$, el cálculo de Sturges quedará:

$$K = 1 + 3.3 \log 40 = 6.28 \approx 6$$

En caso de aplicar la regla empírica, tendremos:

$$K = \sqrt{n} = \sqrt{40} = 6.3 \approx 6$$

Para determinar el valor de la amplitud(A) es decir el ancho de cada intervalo, lo calcularemos con la siguiente fórmula empírica:

$$A = \frac{R}{K}, \text{ donde:}$$

R será el rango de la distribución; que se calculara con la siguiente fórmula:

$$R = DM - dm$$

R= Rango de la distribución

DM= dato mayor de la distribución o conjunto de datos

Dm= Dato menor de la distribución o conjunto de datos.

Para nuestro ejemplo tendremos:

DM= 73 ; dm=49 por lo tanto el rango será:

$$R = 73 - 49 = 24$$

R=24

K= Número de clases o intervalos que llevará la distribución.

Por lo tanto, la amplitud para nuestro ejemplo será:

$$A = \frac{R}{K} = \frac{24}{6} = 4 \Rightarrow A = 4$$

Con la Amplitud o ancho del intervalo y a partir del dato menor que es:

49 le sumaremos 4 unidades que representa la amplitud del intervalo y así sucesivamente hasta tener las seis clases que representa K.

Geométricamente la amplitud representa la base del rectángulo de cada barra del histograma; por tanto tendremos:

49+4=53; 49 y 53 representan los límites superior e inferior de la primera clase
 53+4=57; 53 y 57 representan los límites superior e inferior de la segunda clase
 57+4=61; 57 y 61 representan los límites superior e inferior de la tercera clase
 61+4=65; 61 y 65 representan los límites superior e inferior de la cuarta clase
 65+4=69; 65 y 69 representan los límites superior e inferior de la quinta clase
 69+4=73; 69 y 73 representan los límites superior e inferior de la sexta clase

Al construir la tabla con los límites inferiores y superiores, tendremos los límites de clase de la distribución de la siguiente manera:

INTERVALOS DE CLASE
[49 , 53)
[53 , 57)
[57 , 61)
[61 , 65)
[65 , 69)
[69 , 73]

[= Significa intervalo cerrado, es decir, que el intervalo contiene al dato o número; en cambio, (= Significa intervalo abierto, indicando lo contrario; es decir, que no lo contiene.

Esta información nos servirá para conocer cuantos datos están contenidos dentro de cada intervalo de clase, por ejemplo:

Para en el primer intervalo de clase: [49 , 53) tendremos los siguientes datos que los llamaremos frecuencias absolutas de la primera clase: [49, 49.8, 49.8 , 50.0, 50.3, 50.5, 50.5, 50.6, 51.0, 52.0); el 53.5 lo tomaremos en cuenta hasta el siguiente intervalo, así sucesivamente, hasta completar toda la distribución o conjunto de datos.

El siguiente paso para construir la Tabla de Frecuencias, es contar el número de observaciones que pertenecen a cada clase. Este número es llamado Frecuencia Absoluta de clase (fa); Quedando de la siguiente manera:

Intervalos De Clase	Conteo	Frecuencia Absoluta(f _a)	Marca de Clase (Mc)
[49 , 53)	//// //	10	51
[53 , 57)	//// //	7	55
[57 , 61)	//// /	5	59
[61 , 65)	///	3	63
[65 , 69)	//// //	7	67
[69 , 73]	//// //	8	71
Total		n=Σ40	

DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS

Ahora calcularemos el punto medio o marca de clase de cada intervalo; sumando el límite de clase superior e inferior de cada clase y luego lo dividiremos entre 2, originándose la siguiente fórmula:

$$Mc = \frac{Li + Ls}{2}$$

Donde:

- Mc= Marca de clase o punto medio
- Li= Límite inferior
- Ls= Límite Superior

Entonces la marca de clase o punto medio de la primera clase será:

$$Mc = \frac{49 + 53}{2} = 51 ; \text{ Y así sucesivamente con todas las clases, las marcas de}$$

clase se muestran en la distribución de frecuencias absolutas

MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

Para calcular esta medida de centralización o tendencia central se tomaran en cuenta las frecuencias absolutas y la marca de clase de cada clase; mediante la siguiente fórmula:

$$\bar{X} = \frac{\sum Mc fa}{n}$$

Donde:

- \bar{X} = Media aritmética
- \sum = Sumatoria
- Mc= Marca de clase
- f_a= Frecuencia absoluta
- n= Número de datos (frecuencias absolutas) de la distribución.

Ejemplo: De la

tabla anterior

calcularemos; la Media aritmética para datos agrupados; Aplicando la fórmula tendremos lo siguiente

Intervalos De Clase	Conteo	Frecuencia Absoluta(f_a)	Marca de Clase (Mc)
[49 , 53)	//// //	10	51
[53 , 57)	//// //	7	55
[57 , 61)	//// /	5	59
[61 , 65)	///	3	63
[65 , 69)	//// //	7	67
[69 , 73]	//// //	8	71
Total		$n=\sum 40$	

DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS

$$\bar{X} = \frac{\sum Mcfa}{n}$$

$$\bar{X} = \frac{(10)(51) + (7)(55) + (5)(59) + (3)(63) + (7)(67) + (8)(71)}{40}$$

$$\bar{X} = \frac{510 + 385 + 295 + 189 + 469 + 568}{40} = \frac{2416}{40} \Rightarrow \boxed{\bar{X} = 60.4}$$

MEDIANA PARA DATOS AGRUPADOS.

Para determinar la mediana nos apoyaremos en la siguiente fórmula:

$$\tilde{x} = \tilde{L}_i + \left(\frac{\frac{\sum n}{2} - \sum f_a \text{ anteriores}}{f_{\text{mediana}}} \right) \cdot A$$

Donde:

\hat{L}_i = Límite inferior de la mediana

$\sum n$ = Suma total de frecuencias absolutas

$\sum f_a \text{ anteriores}$ = Suma de todas las frecuencias absolutas que anteceden a la mediana

f_{mediana} = Frecuencia de la mediana

A = Amplitud del intervalo de clase

Ejemplo: De la tabla que se muestra a continuación calcularemos la mediana para esta distribución.

DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS

Intervalos De Clase	Frecuencia Absoluta(f_a)
[49 , 53)	10
[53 , 57)	7
[57 , 61)	5 \tilde{x}
[61 , 65)	3
[65 , 69)	7
[69 , 73]	8
Total	$n=\sum 40$

En el caso en que el

número de clases de una

distribución de frecuencias sea impar como la siguiente distribución de frecuencias, la mediana caerá en la clase que se encuentra a la mitad o en medio de la distribución

Determinaremos primero $\frac{\sum n}{2}$ que será en nuestro caso $\frac{40}{2} = 20$, después contaremos el total de frecuencias absolutas de la segunda columna hasta llegar a 20, sin exceder de esta cantidad ($(n \leq 20)$), es decir:

$$\sum f_a \text{ anteriores} = 10+7=17; \text{ ya que } 17 \text{ es menor que } 20 \text{ entonces}$$

$$f_{\text{mediana}} = 5$$

$$\hat{L}_i = 5$$

$$A=61-57=4;$$

Sustituyendo los valores encontrados en la fórmula para la mediana, tendremos:

$$\tilde{x} = 57 + \left(\frac{17-5}{5}\right) \cdot 4$$

$\tilde{x} = 66.6$

$$\tilde{x} = 57 + \left(\frac{12}{5}\right) \cdot 4 \Rightarrow \tilde{x} = 57+9.6$$

Limites De Clase	Frecuencias absolutas
0,40	4
0,50	5
0,60	8
0,70	2
70,80	1
total	20

$\hat{x} \rightarrow$

Esto significa que la clase que contiene a la mediana será la tercera clase, por lo tanto la mediana será:

$$\hat{x} = 50 + \left(\frac{10-9}{8}\right)(10) \Rightarrow \hat{x} = 51.25$$

MODA PARA DATOS AGRUPADOS.

Para calcular la moda, en una distribución de frecuencias absolutas, observaremos la columna de las frecuencias absolutas, después escogeremos la frecuencia mayor de todas ellas. Ejemplo. La siguiente distribución de frecuencias nos muestra las estaturas de 35 alumnos elegidos aleatoriamente.

Limites de clase	Frecuencias absolutas
1.50 – 1.55	4
1.55 – 1.60	6
1.60 – 1.65	8
1.65 – 1.70	10
1.70 – 1.75	5
1.75 – 1.80	2
Totales	35

En este caso específico será 10 la frecuencia mayor de todas las frecuencias absolutas. Después procederemos a determinarla con la siguiente fórmula:

$$\bar{x} = \hat{L}_i + \left(\frac{d_a}{d_a + d_p} \right) A$$

Donde:

Luego entonces:

$$\begin{aligned} L_i &= 1.65 \\ d_a &= 10 - 8 = 2 \\ d_p &= 10 - 5 = 5 \\ A &= 0.05 \end{aligned}$$

\bar{x} = Moda para datos agrupados

\hat{L}_i = Límite inferior modal

d_a = Diferencia anterior

d_p = Diferencia posterior

Sustituyendo los datos se tiene:

$$\bar{x} = 1.65 + \left(\frac{2}{2+5} \right) (0.05) \Rightarrow \bar{x} = 1.65 + \left(\frac{2}{7} \right) (0.05)$$

$$\bar{x} \cong 1.66$$

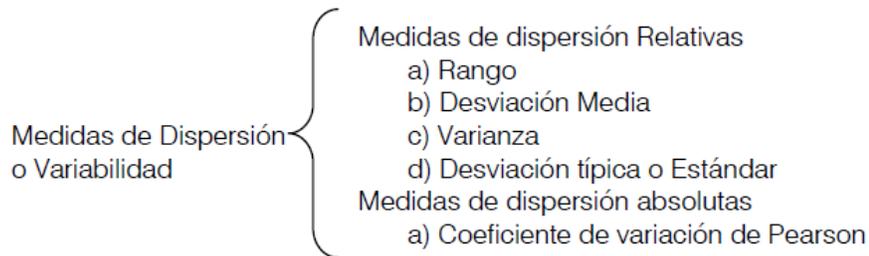
MEDIDAS DE VARIABILIDAD O DISPERSIÓN

Se llaman medidas de dispersión o variabilidad aquellas que permiten retratar la distancia de los valores de la variable a un cierto valor central, o que permiten identificar la concentración de los datos en un cierto sector del rango de la variable.

Las medidas de tendencia central tienen como objetivo el sintetizar los datos en un valor representativo, las medidas de dispersión nos dicen hasta que punto estas medidas de tendencia central son

representativas como síntesis de la información. Las medidas de dispersión cuantifican la separación, la dispersión, y la variabilidad de los valores de la distribución respecto al valor central.

Distinguimos entre medidas de dispersión absolutas, que no son comparables entre diferentes muestras y las relativas que nos permitirán comparar varias muestras.



Medidas de Variabilidad o Dispersión Absolutas

a) **Rango (R)** es una medida razonable de Variabilidad llamada también en algunas ocasiones amplitud y que se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto. Es fácil de calcular y sus unidades son las mismas que las de la variable, aunque posee varios inconvenientes o desventajas tales como las siguientes:

Ventajas y desventajas del Rango

<p>Ventajas</p> <p>a) Es fácil de calcular y es comúnmente usado como una medida burda pero eficaz de variabilidad</p> <p>b) Es comprensible para cualquier persona; aun cuando no conozca de Estadística</p> <p>Desventajas</p> <p>a) No utiliza todas las observaciones (sólo dos de ellas) los extremos, ignorando los valores intermedios</p> <p>b) Se puede ver muy afectado por algunas observaciones o datos extremos</p> <p>c) El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso nunca disminuye.</p> <p>d) No es aconsejable para muestras grandes sólo para muestras pequeñas</p>

b) **Desviación Media** Se define como la media de las diferencias en valor absoluto de los valores de la variable a la media (D:M); es decir, que se define como desvió que es la diferencia que se observa entre la variable y la media aritmética.

Cada valor individual x se desvía de la media por una cantidad igual a $|x - \bar{x}|$

Esta desviación es cero cuando X es igual a \bar{x} , $(x - \bar{x})$

c) **Varianza:** cuyo símbolo es (S^2), es la media de las desviaciones al cuadrado, calculada usando n o $n-1$ como divisor.

d) **Desviación típica o Standard:** cuyo símbolo es (S), es simplemente la raíz cuadrada de la varianza. La varianza y la desviación miden la dispersión promedio alrededor de la media; es decir, como las

observaciones mayores fluctúan por encima de ésta y como las observaciones menores se distribuyen por debajo de ésta como medidas de variabilidad más importantes.

Características de la varianza y la desviación estándar o típica.

- a) Son índices que describen la variabilidad o dispersión y por tanto cuando los datos están muy alejados de la media, el numerador de sus fórmulas será grande y la varianza y la desviación estándar o típica también lo serán.
- b) Al aumentar el tamaño de la muestra, disminuye la varianza y la desviación estándar.
- c) Cuando todos los datos de la distribución son iguales, la varianza y el desvío estándar son iguales a 0.
- d) Para su cálculo se utilizan todos los datos de la distribución; por tanto, cualquier cambio de valor será detectado.
- e) En el caso de la varianza esta será siempre positiva

Medidas de Variabilidad o Dispersión Relativas

Coefficiente de variación de Pearson

Cuando se quiere comparar el grado de dispersión de dos distribuciones que no vienen dadas en las mismas unidades o que las medias no son iguales se utiliza el llamado "**Coefficiente de Variación de Pearson**" y que se define como la relación por cociente entre la desviación estándar y la media aritmética; o en otras palabras, es la desviación estándar expresada como porcentaje de la media aritmética

Medidas de dispersión o Variabilidad para datos agrupados

Recuerda que el rango es una medida de dispersión o variabilidad que se obtiene restando el dato mayor del menor; y en algunas ocasiones recibe el nombre de recorrido y que no lo deberás confundir con el rango visto en Matemáticas 4

a) *Rango o recorrido.* La fórmula para calcular el rango o recorrido es la siguiente

Donde:

$$R = D_m - d_m$$

R= Rango
 D_m= Dato Mayor de la distribución
 d_m= dato menor de la distribución

Ejemplo: Tenemos los siguientes datos, que representan los montos de 20 préstamos personales, en dólares, en una compañía financiera de consumidores:

- 900, 500, 450, 1900, 1200, 1250, 2500, 550, 1650, 1200,
 1000, 550, 650, 600, 750, 1300, 850, 350, 1400, 700,

El rango de estos 20 préstamos será:

R=2500-350

R=2150

La desviación media se calcula utilizando primordialmente utilizando el valor de la media aritmética mediante la siguiente fórmula:



En el caso de la desviación media el valor absoluto significa la distancia de cada uno de los datos con respecto a la media aritmética (Desvío)

$$D.M = \frac{\sum |x - \bar{x}|}{n}$$

Donde: \sum = sumatoria

$| |$ = Valor absoluto

x = cada uno de los datos de la distribución
 \bar{x} = Media aritmética de los datos de la distribución

c) **Varianza (S^2)** Para calcular la varianza para datos no agrupados nos apoyaremos en la siguiente fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n}$$

$$S^2 = \frac{\sum_{i=1}^n (x - \bar{x})^2}{n - 1}$$

donde:

σ^2 = Varianza Poblacional

S^2 = Varianza Muestral

\bar{x} = media aritmética.

x_i = suma total de datos que componen la población

n = número de datos de la población o tamaño de la muestra



Esta varianza se obtiene como la suma de las diferencias de los cuadrados y por tanto tiene unidades de medida el cuadrado de las unidades de medida en que se mide la variable estudiada

Como ejemplo: Las calderas de una planta de energía de vapor a alta presión tuvieron las siguientes eficiencias en porcentajes: 90.3, 91.6, 90.9, 90.4, 90.3, 91.0, 87.9, 89.4.

$$\bar{x} = \frac{90.3 + 91.6 + 90.9 + 90.4 + 90.3 + 91.0 + 87.9 + 89.4}{8} = \frac{721.8}{8} = 90.225$$

La media aritmética para este conjunto de datos será por tanto, el promedio de la eficiencia de las calderas será 90.23

La varianza será:

$$s^2 = \frac{(90.3 - 90.23)^2 + (91.6 - 90.23)^2 + \dots + (89.4 - 90.23)^2}{8 - 1} \cong 8.65$$

d) **Desviación Típica o estándar (S)** La fórmula para determinar esta medida estadística será la siguiente

$$s = +\sqrt{s^2}$$

Donde

s = desviación típica o estándar
 s^2 = varianza de la distribución



Recuerda que la desviación típica o estándar sólo tomará los valores positivos del radical

La desviación típica o estándar del ejemplo anterior será:

$$s = \sqrt{8.65} = 2.94$$

Esto significa que las eficiencias de la caldera de la planta de energía se dispersan en promedio 2.94 unidades con respecto a la media aritmética

CUARTILES, DECILES Y PERCENTILES

Los cuartiles, deciles y percentiles se parecen mucho a la media porque también subdividen una distribución de mediciones de acuerdo con la proporción de frecuencias observadas. Mientras que la mediana divide a la distribución en dos mitades, los cuartiles la dividen en cuatro cuartos, los deciles en diez décimos y los puntos percentiles la dividen en cien partes. Matemáticamente, a manera de ejemplo, se pueden expresar:

$$Q_1 \text{ (Primer cuartil)} = x_{\left[\left(\frac{n+1}{4}\right)\right]}$$

$$D_3 \text{ (Tercer decil)} = x_{\left[\left(\frac{3n+1}{10}\right)\right]}$$

$$P_{70} \text{ (Percentil 70)} = x_{\left[\left(\frac{70n+1}{100}\right)\right]}$$

EJEMPLO

En el restaurante “Nueva Asia” de la zona centro de Mexicali, se obtuvieron las siguientes cifras por el consumo de 15 personas de diversos platillos a la carta.

Sopa de pollo chica	25	Choap Suey de res	52	Pollo asado con hongos	50
Platillo A	29	Pollo asado	48	Sopa aleta de Tiburón	50
Platillo B	32	Pato asado	56	Arroz frito	44
Platillo C	35	Codorniz	62	Tofu y res	50
Platillo D	40	Bocón al vapor	55	Chow main	66

Determine la media (\bar{x}), mediana (\tilde{x}) y moda (\hat{x}), para el total de precios por cada uno de los platillos.

(a) La media aritmética (\bar{x})

$$\bar{x} = \frac{25 + 29 + 32 + 35 + 40 + 52 + 48 + 56 + 62 + 55 + 50 + 50 + 44 + 50 + 66}{15} = \frac{694}{15}$$

$$\bar{x} = 46.27$$

(b) La mediana (\tilde{x})

Ordenando de menor a mayor: 25, 29, 32, 35, 40, 44, 48, 50, 50, 50, 52, 55, 56, 62, 66

$$\tilde{x} = 50$$

(c) La moda (\hat{x}).

Es el dato que más veces se repite:

$$\hat{x} = 50$$

Hallar los valores de:

El segundo cuartil, el segundo decil y el punto percentil 40

(a) El segundo cuartil

$$Q_2 = x_{\left[\frac{2n+1}{4}\right]}, \text{ donde } n = 15$$

$$Q_2 = x_{\left[\frac{2(15)+1}{4}\right]} = x_{[7.5+0.5]} = x_8$$

$$Q_2 = 50$$

(b) El segundo decil

$$D_2 = x_{\left[\frac{2n+1}{10}\right]} = x_{\left[\frac{2(15)+1}{10}\right]} = x_{[3+0.5]} = x_{3.5}$$

$$x_{3.5} = \frac{x_3 + x_4}{2} = \frac{32 + 35}{2}$$

$$D_2 = 33.5$$

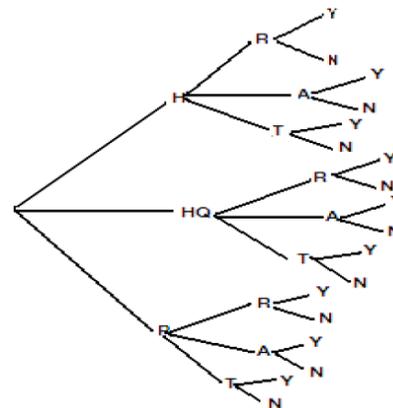
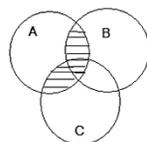
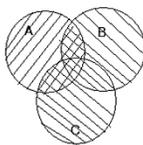
(c) El punto percentil 40

$$P_{40} = x_{\left[\frac{40n+1}{100}\right]} = x_{\left[\frac{40(15)+1}{100}\right]} = x_{[6+0.5]} = x_{6.5}$$

$$x_{6.5} = \frac{x_6 + x_7}{2} = \frac{44 + 48}{2}$$

$$P_{40} = 46$$

TEORÍA DE CONJUNTOS



Conceptos básicos

El concepto de **Conjunto** aparece en todas las ramas de las Matemáticas. De manera intuitiva, un conjunto (1) es cualquier lista bien definida o cualquier colección de objetos, y será representado por las letras mayúsculas A, B, Y, X, etcétera.

Los objetos que componen al conjunto se llaman sus elementos (2) o miembros, y se escriben con letras minúsculas a, b, x, y, etcétera.

Algunos conjuntos básicos de la matemática, llamados también Campos Numéricos:

Ejemplos de conjuntos son:

- ∅: Conjunto vacío, que carece de elementos.
- N: Conjunto de los números naturales.
- Z: Conjunto de los números enteros.
- Q: Conjunto de los números racionales.
- R: Conjunto de los números reales.
- C: Conjunto de los números complejos.

Un conjunto: es la reunión en un todo de objetos bien definidos y diferenciables entre sí, que se llaman elementos del mismo. O también puede ser una colección de objetos.

Si a es un elemento del conjunto A , se denota con la **relación de pertenencia** $a \in A$. En caso contrario, si a , no es un elemento de A , se denota $a \notin A$.

Se puede decir que el símbolo se utiliza para comparar o relacionar un conjunto respecto de un elemento y nos permite relacionar la pertenencia o no, de un elemento en un conjunto. **No es correcto utilizar este símbolo para comparar dos conjuntos si no que exclusivamente para relacionar elementos respecto de un conjunto.**

Ejemplo:

CONJUNTOS	ELEMENTOS	PERTENENCIA
D= Un día de la semana	l = lunes	$l \in D$
M= Un mes del año	m = Mayo	$m \in M$
Z= Número entero	n = 2	$n \in Z$

Los integrantes del grupo “the Beatles”, representados por B: {John, Paul, Ringo, George}

Otros ejemplos de conjuntos son:

- (a) un ejército es un conjunto de soldados.
- (b) una parvada es un conjunto de aves como palomas, huilotas, faisanes, pericos, tucanes, chachalacas, etc.
- (c) un enjambre es un conjunto de insectos como moscas, alimañas, abejas, cigarrones, etc.
- (d) una zahúrda es un conjunto de cerdos, puercos o lechones exclusivamente.

(e) una manada es un conjunto de bestias.

(f) una jauría es un conjunto de animales peligrosos como perros, lobos, dingos, entre otros depredadores.

ESPACIO MUESTRA (S)

Son todas las formas posibles en que pueda suceder un experimento. Se representa por S o Ω .

EVENTO

Un evento se define como la posibilidad que ocurra un suceso de interés.

EVENTOS MUTUAMENTE EXCLUYENTES

Son los eventos en los cuales en la aparición de uno evita la aparición de otro evento

EVENTO COLECTIVAMENTE EXHAUSTIVO

Con la unión de estos eventos se genera el espacio muestra

NOTACION DE CONJUNTOS

$\{ \}$ conjunto $<$ menor que

\emptyset conjunto vacío \geq mayor o igual que

C subconjunto \leq menor o igual que

C subconjunto propio \cup unión de

| tal que \cap intersección de

\in pertenece a - diferencia de

\notin no pertenece a \bar{A} conjunto complemento de A

$=$ igual a $A \cup B$ unión de conjunto A y B

\neq diferente a $C \cap D$ intersección de C y D

\approx aproximado a $A - B$ diferencia de A y B

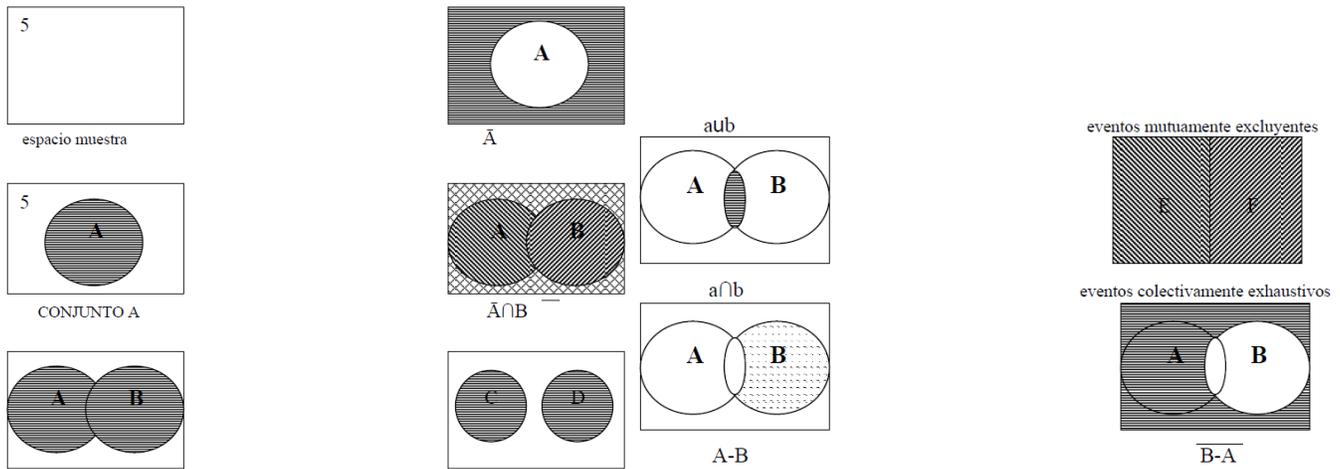
$>$ mayor que B conjunto complemento de B

DIAGRAMAS DE VENN

Los diagramas de Venn son figuras geométricas cerradas que se utilizan para representar operaciones entre conjuntos.

Esta representación de un conjunto se llama gráfica del conjunto o el diagrama de Venn, en honor del matemático inglés de siglo xix, John Venn. En otros países les llaman diagramas de Euler en honor al eminente matemático suizo-alemán Leonard Euler (1707-1783), quien 100 años antes ya utilizaba estos diagramas.

Diagramas de Venn



1.- LEYES DE IDEMPOTENCIA

$$A \cup A = A$$

$$A \cap A = A$$

2.- LEYES ASOCIATIVAS

$$(A \cup A) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

3.- LEYES CONMUTATIVAS

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

4.- LEYES DISTRIBUTIVAS

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

5.- LEYES DE IDENTIDAD

$$A \cup \emptyset = A$$

$$A \cap U = A$$

$$A \cup U = U$$

$$A \cap \emptyset = \emptyset$$

6.- LEYES DE COMPLEMENTO

$$A \cup \bar{A} = U$$

$$A \cap \bar{A} = \emptyset$$

$$(\bar{\bar{A}}) = A$$

$$\bar{U} = \emptyset, \emptyset = U$$

7.- LEYES DE MORGAN

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

Tipos fundamentales de conjuntos

1.- Conjunto universal es el conjunto formado por la totalidad de los elementos considerados para una determinada operación. Se representa por U.

Ejemplo: el conjunto de números de dígitos: $U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

2.- Conjunto vacío conjunto que no tiene elementos, al cual también se le llama conjunto nulo. Se representa por $\{\emptyset\}$.

3.- Conjuntos equivalentes

Son aquellos que poseen la misma cardinalidad u orden, aunque sus elementos sean diferentes.

4.- Conjuntos iguales

Dos conjuntos son iguales, si son equivalentes y además, los elementos de uno son también los elementos del otro.

5.- Subconjunto

Dados dos conjuntos a y b en que todos los elementos de a pertenecen al conjunto b , entonces decimos que el conjunto a es subconjunto de b .

6.- Subconjunto propio

Dados dos conjuntos a y b decimos que a es subconjunto propio de b , si a es subconjunto de b y existe al menos un elemento de b que no pertenece al conjunto a .

7.- Conjunto complemento

Sea u el conjunto universal y s un subconjunto cualquiera de u . el conjunto de los elementos que faltan a s para completar u , es el "complemento de s ", (s).

8.- Conjuntos disjuntos

A 2 conjuntos a y b se les denomina "disjuntos" si no tienen elemento en común, es decir, $a \cap b = \emptyset$

9.- conjuntos finitos

Son aquellos conjuntos, en los cuales sus elementos terminan de enumerarse completamente, no obstante la dificultad que pueda presentarse.

10.- conjunto infinito

Es aquel en que no es posible de enumerar a el total de sus elementos, por ejemplo los números, el tiempo.

OPERACIONES CON CONJUNTOS

Ejemplo 1.- sea la proposición abierta: $a = \{x/ x \text{ es un entero y } 10 < x < 14\}$

A la respuesta de la proposición abierta se le conoce como conjunto de reemplazamiento, y a consiste en los números mayores a 10 pero menores a 14, luego entonces:

$$a = \{11, 12, 13\}$$

Ejemplo 2.- dados los conjuntos, hallar $A \cap B$ $a = \{x / x^2 + x - 12 = 0\}$ y $b = \{x / 2x - 6 = 0\}$

Los elementos de a son las raíces de la ecuación cuadrática:

$$x^2 + x - 12 = 0, \text{ factorizando:}$$

$$(x + 4)(x - 3) = 0$$

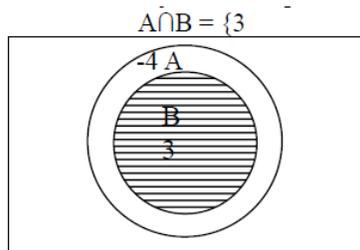
$$x_1 = -4 \quad x_2 = 3$$

para el conjunto b:

$$2x - 6 = 0, \text{ despejando } x,$$

$$x = 6/2 \quad x = 3$$

El término que se repite en ambos conjuntos es 3, por lo tanto:



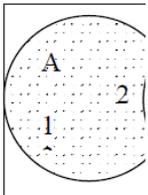
Ejemplo 3.-

Sea el conjunto $u = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ como el conjunto universo y $a = \{1, 2, 3, 4, 5, 6\}$ $c = \{4, 5, 6, 7\}$ y $b = \{2, 3, 4, 5\}$ $d = \{7, 8, 9, 10\}$

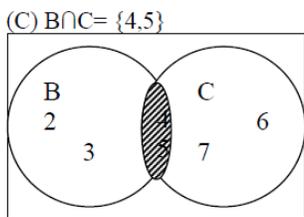
Determinar los conjuntos que se indican y representar la operación gráficamente, sombreando el resultado.

- | | | |
|----------------|------------------------|----------------------------------|
| (A) $A \cup C$ | (E) $A \cup B$ | (I) $C \cup D$ |
| (B) $D \cup C$ | (F) $B \cup \emptyset$ | (J) $C \cup (A \cap D)$ |
| (C) $B \cap C$ | (G) $C \cap D$ | (K) $(B \cap D) \cup (B \cap C)$ |
| (D) $A \cap D$ | (H) $A \cap B$ | (L) $(C \cup D) \cap (A \cup B)$ |

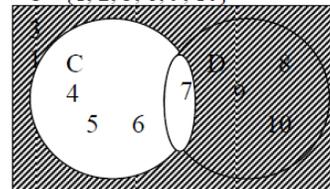
(A) $A \cup C = \{1, 2, 3, 4, 5, 6, 7\}$



(B) $D \cup B = \{2, 3, 4, 5, 7, 8, 9, 10\}$

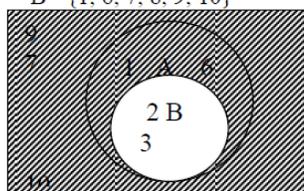


(G) $C \cap D = \{7, 8, 9, 10\}$
 $\bar{C} = \{\text{ELEMENTOS DE } U \text{ EXCEPTO LOS DE } C\}$
 $C = \{1, 2, 3, 4, 5, 6, 7\}$

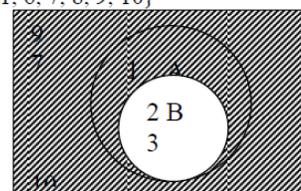


(D) $A \cap D = \{7, 8, 9, 10\}$

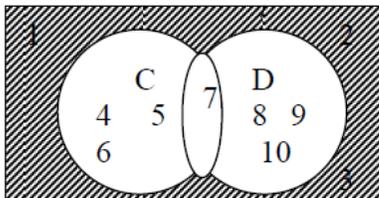
(E) $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 $\bar{B} = \{\text{ELEMENTOS DE } U \text{ EXCEPTO } B\}$
 $B = \{1, 6, 7, 8, 9, 10\}$



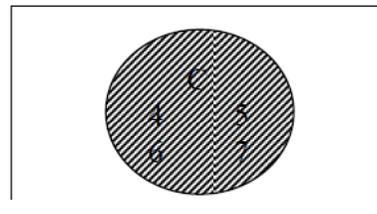
(H) $\overline{A \cap B} = \{\text{ELEMENTOS DE } U \text{ EXCEPTO } A \cap B\}$
 $A \cap B = \{2, 3, 4, 5\}$
 $\overline{A \cap B} = \{1, 6, 7, 8, 9, 10\}$



(i) $\overline{C \cup D} = \{\text{ELEMENTOS DE U, EXCEPTO C \cup D}\}$
 $C \cup D = \{4, 5, 6, 7, 8, 9, 10\}$
 $\overline{C \cup D} = \{1, 2, 3\}$

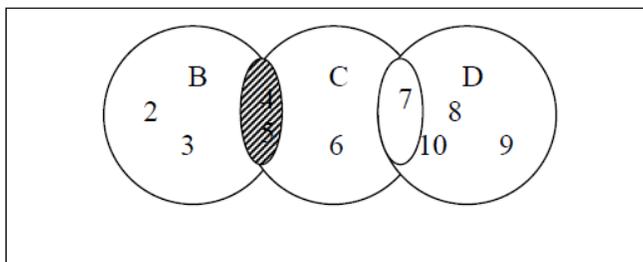


(j) $(U \setminus (A \cap D)) = \{C\}$
 $A \cap D = \{\emptyset\}$

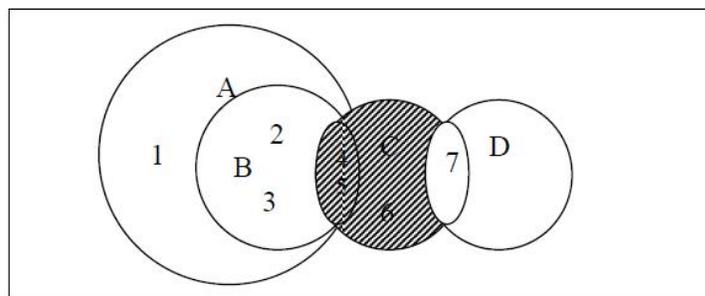


(k) $(B \cap D) \cup (B \cap C) = \{4, 5\}$
 $B \cap D = \{\emptyset\}$

$B \cap C = \{4, 5\}$



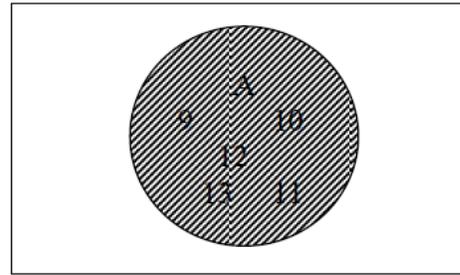
(l) $(C \cup D) \cap (A \cup B) = \{4, 5, 6\}$
 $C \cup D = \{4, 5, 6, 7, 8, 9, 10\}$
 $A \cup B = \{1, 2, 3, 4, 5, 6\}$



Ejemplo 4.- Enumerar los elementos del conjunto siguiente y representarlo gráficamente.

$$a = \{x \mid x \text{ es un entero y } 9 \leq x \leq 13\}$$

$$a = \{9, 10, 11, 12, 13\}$$



Ejemplo 5. Dados los conjuntos: $A = \{x \mid x^2 - 7x = 8\}$, $B = \{x \mid 2x - 16 = 0\}$. Hallar $A \cap B$, y representar gráficamente.

Simplificando: $x^2 - 7x - 8 = 0$,

$$2x - 16 = 0$$

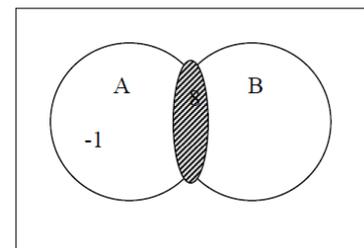
Factorizando: $(x - 8)(x + 1) = 0$

$$2x = 16 \quad x = 16/2$$

$$x_1 = 8 \quad x_2 = -1$$

$$x = 8$$

$$A \cap B = \{8\}$$



Luego entonces: $A = \{8, -1\}$ $B = \{8\}$

Ejemplo 6. Dados los conjuntos: $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ $A = \{1, 2, 3\}$ $B = \{2, 4, 6, 8\}$ $C = \{1, 2, 4, 8\}$

Verificar las siguientes propiedades del algebra de eventos,

(a) cerradura. $(A \cup B) \subset U$

$A \cup B = \{1, 2, 3, 4, 6, 8\}$, se comprueba $(A \cup B) \subset U$

(b) asociativa $(A \cup B) \cup C = A \cup (B \cup C)$

$(A \cup B) \cup C = \{1, 2, 3, 4, 6, 8\}$ $A \cup (B \cup C) = \{1, 2, 3, 4, 6, 8\}$

Comprobación $\{1, 2, 3, 4, 6, 8\} = \{1, 2, 3, 4, 6, 8\}$

(c) conmutativa $A \cup B = B \cup A$

$A \cup B = \{1, 2, 3, 4, 6, 8\}$ $B \cup A = \{1, 2, 3, 4, 6, 8\}$

Comprobación $\{1, 2, 3, 4, 6, 8\} = \{1, 2, 3, 4, 6, 8\}$

(d) existencial de neutro $A \cup \emptyset = A$

$A \cup \emptyset = \{1, 2, 3\}$ $A = \{1, 2, 3\}$, comprobación: $\{1, 2, 3\} = \{1, 2, 3\}$



(e) existencial de complemento: $A \cup \bar{A} = U$

$A = \{1, 2, 3\}$ $\bar{A} = \{\text{elementos de } U, \text{ excepto los de } A\}$ $\bar{A} = \{4, 5, 6, 7, 8, 9\}$

$A \cup \bar{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

comprobación: $\{1, 2, 3, 4, 5, 6, 7, 8, 9\} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

(f) distributiva $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

$(A \cup B) \cap C = \{1, 2, 4, 8\}$ $A \cap C = \{1, 2\}$ $B \cap C = \{2, 4, 8\}$

Comprobación: $\{1, 2, 4, 8\} = \{1, 2, 4, 8\}$

TÉCNICAS DE CONTEO

PRINCIPIO ADITIVO

Para dos operaciones mutuamente excluyentes en donde la primera operación puede hacerse de m formas y la segunda operación de n formas. Entonces una o la otra pueden hacerse de:

$$M+N \text{ FORMAS}$$

PRINCIPIO MULTIPLICATIVO

Si un experimento se realiza en r pasos y si el primer paso se realiza en n_1 formas, el segundo de n_2 , el tercero de n_3 , y así sucesivamente hasta n_r , entonces el experimento se puede llevar a cabo de:

$$N_1 \cdot N_2 \cdot N_3 \cdot \dots \cdot N_r \text{ FORMAS}$$

NOTACION FACTORIAL

El producto de los enteros positivos desde 1 hasta n incluso, se emplea con mucha frecuencia en matemáticas y aquí lo representaremos por el símbolo especial que es $n!$, el cual se lee "numero factorial de".

PERMUTACIONES

Una permutación es todo arreglo de números posibles en un experimento, en los que importa el orden, matemáticamente se expresa:

$${}_n P_x = \frac{n!}{(n-x)!}$$

COMBINACIONES

Una combinación es todo arreglo de número posibles en un experimento en los que no importa el orden, matemáticamente se expresa:

Ejemplo 1.- hallar el valor de la notación factorial

(A) $4!$ (B) $6!$ (C) $2!$

(A) $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$

$$(B) 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$$

$$(C) 2! = 2 \cdot 1 = 2$$

$$(D) 9! = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{9 \cdot 8 \cdot 7 \cdot 6!}{6!} = 9 \cdot 8 \cdot 7 = 504$$

POR DEFINICION: $0! = 1$