SEGUNDA EDICIÓN DEL CURSO DE CAPACITACION
EN MATEMATICA
PARA PROFESORES DE PRIMARIA

MODULO IV
ESTADISTICA DESCRIPITVA

ENCUENTRO NÚMERO CUATRO Y CINCO

19 DE OCTUBRE DE 2014
MANAGUA
FINANCIADO POR: FUNDACIÓN UNO

Tipos de frecuencias:

Frecuencia absoluta

La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa por f_i.

La suma de las frecuencias absolutas es igual al número total de datos, que se representa por N.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

Frecuencia relativa:

La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento y se representa por f_{ri}.

$$fr_i = \frac{f_i}{n}$$

La suma de las frecuencias relativas es igual a 1.

Frecuencia acumulada y relativa acumulada

La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por F_i.

La frecuencia relativa acumulada es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento.

Ejemplo

Durante el mes de Agosto, en Managua se han registrado las siguientes temperaturas máximas: 32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

x_i	Recuento	f_i	F_i	fr_i	Fr_i
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097
29	~~IIII~~ I	6	9	0.194	0.290
30	~~IIII~~ II	7	16	0.226	0.0516
31	~~IIII~~ III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Este tipo de tablas de frecuencias se utiliza con variables discretas.

Las tablas de frecuencia para una variable cualitativa son similares a las tablas de frecuencia definidas para las variables cuantitativas.

Ejemplo: Se pidió a 40 profesores su peso en libras. Se clasificaron los datos registrados en tres rangos; S:sobrepeso, N:normal y B:por debajo del peso esperado. Los resultados fueron:

S,S,N,B,N,B,B,S,S,N,N,B,S,N,N,B,B,S,S,N,
S,N,B,B,S,N,N,B,S,S,S,N,S,S,N,N,B,N,N,B

Para construir una tabla de frecuencia se consideran tres intervalos o rangos: sobrepeso, peso normal y por debajo del peso esperado. Luego, es necesario determinar el numero de profesores que están están en cada uno de los rangos con el fin de determinar la frecuencia correspondiente. Finalmente se construye la columna de la frecuencia relativa teniendo en cuenta la frecuencia acumulada y el total de estudiantes, este proceso es igual que para variables cuantitativas.

Intervalo	fi	Fi	fri	Fri
Sobrepeso	14	14	0.35=35%	0.35
Normal	15	29	0.375=37.5%	0.725
Por debajo	11	40	0.275=27.5%	1.00
Total	40		1.00	

Ejercicios:

1. Las puntuaciones obtenidas por un grupo en una prueba han sido: 15, 20, 15, 18, 20, 13, 13, 16, 15, 19, 18, 15, 16, 20, 16, 15, 18, 16, 14, 13.
   Construir la **tabla de distribución de frecuencias**.

2. Las asignatura preferida de 20 estudiantes han sido las siguientes: EF, EF, E, E, CN, M, M, E, E, EF, M, CN, CN, EF, E, M, E, CN, M, EF. Construir la tabla de distribución de frecuencias

Distribución de frecuencias agrupadas

La distribución de frecuencias agrupadas o tabla con datos agrupados se emplea si las variables toman un número grande de valores o la variable es continua.
Se agrupan los valores en intervalos que tengan la misma amplitud denominados clases.
A cada clase se le asigna su frecuencia correspondiente.

Límites de la clase
Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase.

Amplitud de la clase
La amplitud de la clase es la diferencia entre el límite superior e inferior de la clase.
Marca de clase
La marca de clase es el punto medio de cada intervalo y es el valor que representa a todo el intervalo para el cálculo de algunos parámetros.

Construcción de una Tabla de Distribución de Frecuencias (TDF)

Para la construcción de una TDF, existen varias técnicas, en general se recomienda que el número de clases deban oscilar entre cinco y quince; y esto dependerá del número de elementos que tenga la información que se agrupará en ella.
La técnica que mostraremos a continuación, es una técnica que de antemano se debe dar el número de clases, los pasos son los siguientes:

1º) Hallar el rango; $R = V_{mayor} - V_{menor}$

2º) Establecer el cociente R / K ; donde K es el número de intervalos

3º) Seleccionar la amplitud de las clases, la denotaremos con la letra: C; para la selección de la amplitud de los intervalos, debe tomarse en cuenta el número de decimales en que vienen dado los datos. En este caso se encuentra la unidad de los datos;

Ejemplo:
u = 1, si los datos son enteros
u = 0.1, si al menos un dato tiene como máximo un decimal
u = 0.01, si al menos un dato tiene como máximo dos decimales.
Etc. Luego $C = (R / K)_{truncado} + u$; el truncamiento de este cociente se debe hacer hasta el número de decimales que tenga la unidad de los datos.

4º) Construir el primer intervalo: las clases contienen dos columnas:
$Li = V_{menor}$ (límite inferior)
$Ls = V_{menor} + C - u$ (límite superior)

5º) Para los demás intervalos solamente se suma la amplitud al límite anterior. El último intervalo deberá contener al dato mayor.
Posteriormente con un ejemplo, se construirá una TDF completa, ya que los pasos enumerados anteriormente solamente son para obtener las clases, que es una parte importante en esta forma de representar información, pero hay otros elementos que también son de interés en una TDF.

Ejemplo:Un taxista anota en 20 días la cantidad de servicios que hizo. Estos datos se presentan abajo . Hacer una tabla de frecuencias con 4 intervalos:

23	45	23	23
24	35	26	25
13	46	33	31
43	15	43	34
28	32	19	50

Ordenar los datos en forma ascendente (no es necesario ordenarlos, pero ayuda a organizar)

13	23	31	43
15	24	32	43
19	25	33	45
23	26	34	46
23	28	35	50

1. Calcular el rango:

R=50-13=37.

2. Establecer el cociente R/k:

$$\frac{Rango}{N^{o}\ de\ int\ ervalos} = \frac{37}{4} = 9{,}25$$

3. Seleccionar la amplitud:

C = 10, lo redondeamos a 10 (si redondeamos al inferior, pasa frecuentemente que quedan datos sin incluir en la tabla). Entonces el tamaño del intervalo será de C-u= 10-1=9 unidades.

u=1 porque los datos no tienen decimales, es decir son cantidades enteras.
4. Construir los intervalos.

Iniciando con el dato menor 13 los números, 14, 15,16, 17, 18, 19, 20, 22, formarían parte del primer intervalo.
Los siguientes serán: [23-32], [33-42] y [43-52]. Estos serían los cuatro intervalos pedidos en el ejercicio.

Construir la tabla de frecuencias.

Intervalo	*Frecuencia*
13 - 22	3
23 -32	9
33 -42	3
43 -52	5
Total	*20*

La frecuencia corresponde al número de repeticiones que se da en el intervalo. Por ejemplo en el intervalo 23 – 32 están los siguientes números: 23, 23, 23, 24, 25 ,26 ,28 ,31 y 32.Los intervalos de clase tienen un límite inferior y uno superior. En el intervalo 33-42 teóricamente se incluyen las anotaciones desde 33 hasta 42 servicios. Estos números los cuales podemos escribir como 32,5 y 42,5 se llaman límites de clase

REPRESENTACION GRAFICA DE DISTRIBUCIONES DE FRECUENCIAS

Otra forma de presentar la información de manera comprensible y útil es a través de gráficos adecuados, lo cual nos permitirá visualizar una descripción de los datos, así también como lograr obtener una idea del "modelo teórico" que se ajusta mejor a los mismos a partir de la comparación de gráficos empíricos con teóricos. Las tablas de frecuencia a menudo son presentadas en forma gráfica debido a que así se obtiene una mejor imagen visual de la distribución que en la tabla. Las gráficas son usadas para mostrar las fluctuaciones en los datos.

Las formas más comunes de representar gráficamente una tabla de frecuencias son: Diagrama de sectores (o pastel o gráfico circular), Diagrama de Barras, Polígonos de frecuencias, Histograma y Ojiva.

Los tipos de gráficos que se deben usar están en dependencia del tipo de variable, así

TIPO DE VARIABLE	TIPO DE GRÁFICO
Variable Cualitativa	Diagrama de sector. Gráfico de barras
Variable cuantitativa discreta	Diagrama de sector. Gráfico de barras. Polígono de frecuencias. Ojiva.
Variable cuantitativa continua	Diagrama de sector. Gráfico de barras. Polígono de frecuencias. Ojiva

Regla de los tres cuartos de altura

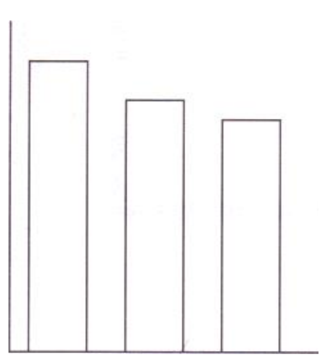
La mayoría de las representaciones gráficas en la Estadística se hacen con referencia a un sistema de coordenadas rectangulares en el que en el eje horizontal se sitúan los valores de la variable y en el eje vertical, las frecuencias con que se observan dichos valores, correspondiendo la mayor altura a la mayor de las frecuencias.

Un convenio utilizado sobre la unidad de medida que debemos utilizar en cada eje, a fin de que la información que queremos representar se deforme lo menos posible, es la llamada

Regla de los tres cuartos:
El eje vertical tendrá como altura *aproximadamente* tres cuartos de la longitud del eje horizontal. Por ejemplo, si el eje horizontal mide 8 cm, entonces el eje vertical deberá medir aproximadamente:

$\frac{3}{4}$ x (8 cm) = 6 cm

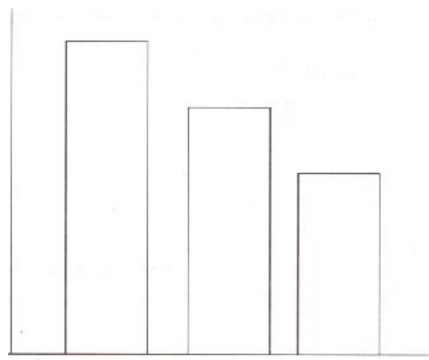
Veamos dos gráficas de barras que representan los mismos datos, en las cuales no se sigue la regla de los tres cuartos, el efecto visual de ambas es diferente:



Gráfica de barras con una proporción inadecuada entre los ejes

Como podemos notar, el efecto visual que producen ambos gráficos es diferente. El primero indica una diferencia mayor entre las barras que el segundo y sin embargo representan la misma información. En el primer gráfico la unidad de medida en el eje vertical es mayor en relación con la unidad de medida utilizada en el eje horizontal; en el segundo, al contrario, la relación es la inversa de la del primero.

Veamos ahora la representación gráfica de los mismos datos pero aplicando la regla de los tres cuartos:



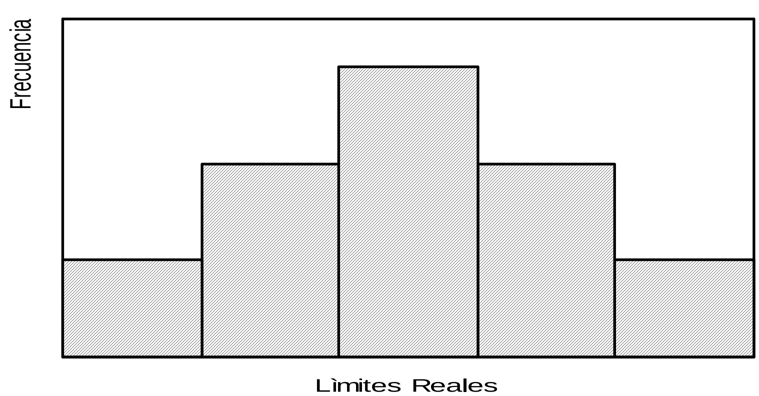
Gráfica de barras con la proporción adecuada entre los ejes

Características de los gráficos

(a) Deben llevar título y todas indicaciones necesarias para su fiel interpretación.
(b) Ajustarse a la realidad de los datos que representen.
(b) Deben ser claros, fácil de leer y entender.
(c) Pueden llevar un número que indique la secuencia de varios gráficos y que sirva como frecuencia.

Histograma

Un histograma de frecuencia consiste en una serie de rectángulos que tiene sus bases sobre el eje horizontal (x), con centros en las marcas de clases y longitud igual al tamaño de los intervalos de clase. Las alturas sobre el eje vertical (y) son proporcionales a las frecuencias de las clases.



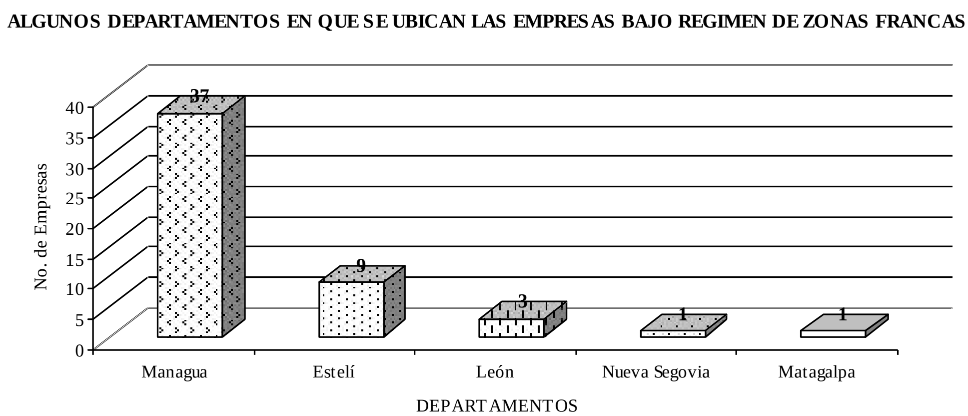
Diagramas de barras

Son representaciones gráficas de colecciones de observaciones de variables cualitativas o cuantitativas, consistentes en barras o rectángulos separados (Histograma rectángulos unidos), ubicados en un sistema de ejes coordenados igual que el Histograma

Ejemplo
Representar en un diagrama de barras la siguiente información
Algunos Departamentos en que se ubican las Empresas bajo régimen de zonas francas

Departamentos	Número de Empresas
Managua	37
Estelí	9
León	3
Nueva Segovia	1
Matagalpa	1

Fuente: Comisión Nacional de Zonas Francas



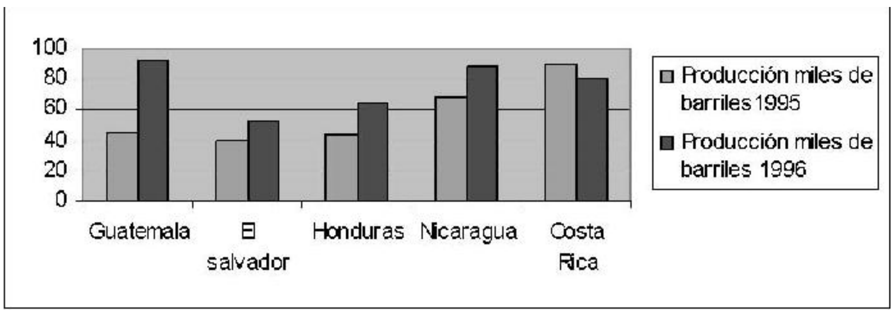
Fuente: Comisión Nacional de Zonas Franca
De 51 Empresas zonas francas, 37 se encuentran ubicadas en el Departamento de Managua, 9 Empresas zonas francas hay en Estelí; Nueva Segovia y Matagalpa tienen una empresa de zona franca respectivamente.

Ejemplo

La producción de aceite vegetal en Centroamérica en los años 1995 y 1996 se muestran en el siguiente cuadro .

País	*Producción miles de barriles 1995*	*Producción miles de barriles 1996*
Guatemala	45	92
El salvador	40	53
Honduras	43	65
Nicaragua	69	88
Costa Rica	90	80

Un diagrama de barras para la información dada es:



Basado en la información, responda las siguientes preguntas:

1 El país que tuvo la mayor producción de aceite vegetal en 1995 fue: ____________________.

2 El país que experimento un crecimiento mayor en su producción de aceite vegetal en 1996 fue: ____.

3. ¿Cuál de las aseveraciones siguientes es posible hacer con base en el gráfico?

a. La producción de aceite vegetal se incrementó en todos los países centroamericanos en 1996.

b. El porcentaje de crecimiento en la producción de aceite vegetal fue el mismo en todos los países centroamericanos.

c. Nicaragua ocupó el segundo lugar en porcentajes de crecimiento en la producción e aceite vegetal en Centroamérica.

d. El país que experimentó el menor porcentaje de crecimiento en la producción de aceite vegetal fue El Salvador.

2. Gráfica de sectores o de pastel

Otra forma muy difundida de gráficos son aquellos que utilizan un círculo y sectores circulares para representar los datos. Se divide el círculo por medio de radios en sectores representados de las proporciones de la frecuencia de cada clase respecto al total, expresado en %; esto es, el total de los datos corresponde al número total de grados en el círculo; es decir, 360^0.

Ejemplo

Representar en un gráfico de sector la siguiente información

Algunos Departamentos en que se ubican las Empresas bajo régimen de zonas francas

Departamentos	Número de Empresas
Managua	37
Estelí	9
León	3
Nueva Segovia	1
Matagalpa	1
Total	51

Fuente: Comisión Nacional de Zonas Francas

Para determinar el ángulo que le corresponde a cada sector se debe establecer una regla de tres en la siguiente forma:

Al total de frecuencia 34 + 9 + 3 + 1 + 1 = 51 le corresponde 360 grados, que es el ángulo central subtendido por el perímetro del círculo.

A la frecuencia 37. ¿Cuántos grados le corresponde?

Frecuencia	Grados
51	360
37	x

Entonces,

$$x = \frac{37 \text{ x } 360}{51} = 261^o$$

A la frecuencia 9:

Frecuencia	Grados
51	360
9	x

Entonces,

$$x = \frac{9 \text{ x } 360}{51} = 64^o$$

A la frecuencia 3:

Frecuencia	Grados
51	360
3	x

Entonces,

$$x = \frac{3 \text{ x } 360}{51} = 21^o$$

A la frecuencia 1:

Frecuencia	Grados
51	360
1	x

Entonces,

$$x = \frac{1 \text{ x } 360}{51} = 7^{o}$$

Debemos observar que la suma de los ángulos $261^{o} + 64^{o} + 21^{o} + 7^{o} + 7^{o} = 360^{o}$; de lo contrario, hemos incurrido en algún error en los cálculos. Para medir el ángulo que le corresponde a cada sector debemos usar un transportador.

Además, debemos calcular el porcentaje que representa cada categoría.

Frecuencia	porcentaje
51	100
37	x

Entonces,

$$x = \frac{100 \text{ x } 37}{51} = 72.5 \ \%$$

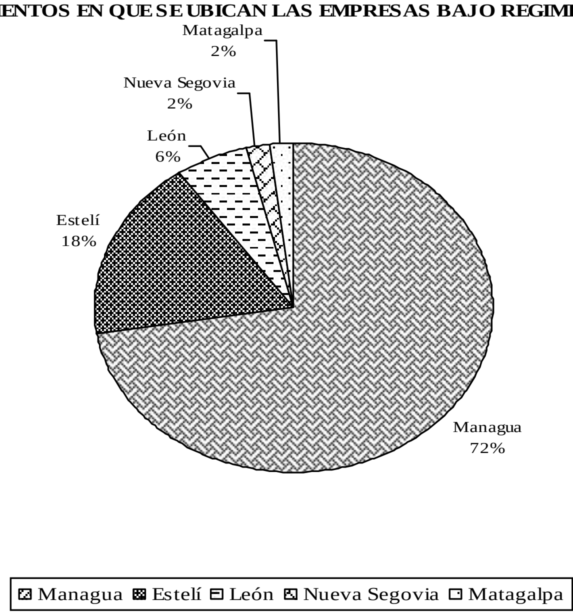
Frecuencia	porcentaje
51	100
9	x

Entonces,

$$x = \frac{100 \text{ x } 9}{51} = 18 \ \%$$

De manera análoga se hace para las otras frecuencias.

ALGUNOS DEPARTAMENTOS EN QUE SE UBICAN LAS EMPRESAS BAJO REGIMEN DE ZONAS FRANCAS



Fuente: Comisión Nacional de Zonas Francas

El 72 % de las Empresas de Zonas Francas se encuentran ubicadas en el Departamento de Managua; 18 % en el Departamento de Estelí; 6 % en el Departamento de León; 2 % en el Departamento de Nueva Segovia y 2 % en el Departamento de Matagalpa.

Cuando las variables que queremos representar son continuas podemos representar gráficamente su tabla de frecuencias en un histograma o un polígono de frecuencias.

Ejercicios:

1. En un concurso de "Bebidas típicas nicaragüenses" en las que participan 40 personas se encontraron los siguientes resultados de calidad:

Categoría	*Frecuencia*
Excelente	12
Muy Bueno	7
Bueno	10
No calificado	11
Total	***40***

a. Haga un pastel.
b. Haga un histograma.

2. Analiza el siguiente gráfico y responde. El gráfico muestra las ventas de arroz y azúcar de un almacén, en cuatro días de la semana:

De acuerdo al gráfico, a medida que pasan los días:

A. la venta de arroz y de azúcar aumenta.
B. la venta de arroz y de azúcar disminuye.
C. la venta de arroz aumenta y la de azúcar disminuye.
D. la venta de arroz disminuye y la de azúcar aumenta.

3. En una cooperativa de taxis de se midió el consumo de gasolina, obteniendo los kilómetros que hicieron 40 vehículos por galón de consumo. Los resultados fueron los siguientes:

45.0	38.4	44.3	44.2	45.3	44.2	43.6	44.5	44.4	39.8
43.2	44.0	43.8	43.8	45.3	44.5	44.6	44.0	45.2	38.7
44.4	44.7	44.1	42.6	43.9	44.1	45.8	42.2	41.2	40.6
42.1	45.6	44.5	39.7	40.7	42.3	45.2	44.7	43.3	38.6

a) agrupe estos datos en una distribución que tenga 7 clases
b) Haga un Diagrama de barras que represente a los datos de la frecuencia absoluta de la TDF

4. Los datos son mediciones de intensidad solar directa (en watts/m^2) realizados en distintos días en una localidad.

562	869	708	775	775	704	809
856	655	806	878	909	918	558

768	870	918	940	946	661	820
898	935	952	957	693	835	905

a) Construya una T.D.F con 5 clases e interprete la cuarta.
b) Haga un histograma y diagrama de pastel con los datos de la TDF

5. En una empresa de zona franca se producen 5 tipos de prendas de vestir, la producción en una semana de éstas se muestra a continuación:

Prendas de vestir	Producción semanal
Camisa manga corta	1500
Camisa manga larga	1800
Pantalones de vestir	1200
Pantalones de azulon	1900
Otro tipo	1200

a) grafique un diagrama de barra
b) cual es la proporción de las dos prendas que más se produce en dicha empresa.

Construir la **tabla de distribución de frecuencias** y dibuja el **diagrama de barras**.

MEDIDAS DE TENDENCIA CENTRAL PARA DATOS NO AGRUPADOS

En estadística, además de recopilar, organizar y representar gráficamente los datos, se hace necesario hacer resúmenes de los mismos en un solo número que es la síntesis de las características básicas del total de datos; se puede decir de ellos, que son representativos del conjunto de datos.

Antes de iniciar el estudio de las medidas de tendencia central, señalaremos algunos ejemplos donde se utiliza el término promedio como sinónimo de medidas de tendencia central.

a. La edad promedio de los estudiantes de sexto grado de una escuela es de 12 años.
b. El salario promedio de los profesores de la una escuela es de C$3500.00 (tres mil córdobas netos).
c. La temperatura promedio en Managua en la última década fue de 31 grados centígrados.
d. El peso promedio de l@s niñ@s Nicaragüenses al nacer, es de 7.5 libras.
e. Los estudiantes de mi escuela estudian un promedio de 1 horas diaria.

Las medidas de tendencia central son valores numéricos que tienden a localizar en algún sentido, la parte central de un conjunto de datos.
Las medidas de tendencia central para datos no agrupados que vamos a estudiar con datos libres son: La Media, La Mediana y La Moda

La Media:

La media aritmética o simplemente media se denota así: $\bar{x}$, también se conoce como el promedio de un conjunto de números. El procedimiento para calcular la media de datos libres es sumar todos los valores y dividir por el número de esos valores. La media de una población y la media de una muestra se calculan de la misma manera.
La fórmula que aparece a continuación es para calcular la media de una muestra.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

donde:

$\bar{x}$: es la media muestral.

x_i: indica los valores específicos para i = 1, 2, 3, …, n.

n : el número total de datos.

Ejemplo

1. Calcule la media de los siguientes datos: 6, 8, 7, 5, 3, 7. Interprete el resultado obtenido

Solución

Como n = 6 y x_1 = 6, x_2 = 8, x_3 = 7, x_4 = 5, x_5 = 3, x_6 = 7; entonces,

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{6 + 8 + 7 + 5 + 3 + 7}{6} = \frac{36}{6} = 6$$

Por tanto, el promedio o media aritmética del conjunto de datos es 6.

2. Los años de servicios de los siguientes empleados que se jubilan, son los que aparecen en la siguiente tabla:

EMPLEADO	AÑOS DE SERVICIOS
Celia Zúñiga	13
Armando Cáceres	22
Juana Blanco Meza	27
María Inés Paredes	24
Rosa Montes	19

(a) Calcule la media aritmética de los años de servicios.

(b) Cuál de los empleados a jubilarse está más próximo al promedio de los años de servicios.

Solución:

(a) $$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{13 + 22 + 27 + 24 + 19}{5} = \frac{105}{5} = 21$$

La media de los años de servicios es de 21 años.

(b) El empleado que está más próximo al promedio (21 años) es Armando Cáceres con 22 años de servicio.

La media de un conjunto de datos se puede representar gráficamente en la recta numérica.

3. Encuentre el nivel medio de azúcar en la sangre de 12 personas diabéticas a quienes se le aplicó la prueba de hemoglobina A_{1c} teniendo en cuenta que los exámenes mostraron los siguientes resultados:
6.5 6.4 5.0 7.9 5.0 6.0 8.0 6.5 6.1 9.2 6.5 6.2
Solución
x_1= 6.5, x_2= 6.4, x_3= 5.0, …, x_{12}= 6.22. Entonces, el nivel medio de azúcar en la sangre de esas 12 personas es:

$$\bar{x} = \frac{6.5 + 6.4 + 5.0 + ... + 6.2}{12} = \frac{79.3}{12} = 6.6$$

Una *representación física* de la media podría ser el imaginarnos una recta numérica sobre un punto de apoyo o un punto de balance o de equilibrio para las siguientes unidades de peso en libras: 1, 2, 3, 4 y 20.

Como se ilustra a continuación:

En este ejemplo, la media es poco representativa pues cuatro de los cinco datos están por debajo de ella.

Cuando hay números en los extremos o muy pequeños o muy grandes, la media es sensible a ellos y cambia de tal forma que no puede ser realmente representativa de la mayoría de las mediciones. En tal caso es más conveniente usar una medida de tendencia central más adecuada, *la mediana*, que es el valor debajo del cual se encuentra la mitad de los datos y arriba del cual se halla la otra mitad, y que definimos a continuación.

La mediana (Me)

Es una medida de tendencia central que se denota por Me. Para localizar la mediana de un conjunto de datos hay que hacer lo siguiente:
1. Ordenar los datos de menor a mayor.
2. Determinar si el número total de datos es impar o par.
3. Si el número de datos *n*, es impar, la mediana es el número que se encuentra en el centro de los datos ordenados; si el número de datos *n* es par, la mediana es el promedio de los dos números centrales.

Ejemplos:
1. Determine la mediana para los siguientes datos: 2, 4, 7, 8, 9.
Solución:
Primeramente, ordenamos los datos: 2, 4, 7, 8, 9.
La mediana es Me = 7, pues este dato es el valor central.

La posición de la mediana $d\,(Me)$ se determina por la fórmula:

$d\,(Me) = \frac{n+1}{2}$, donde n es el número de datos.

Para los datos del ejemplo , la posición de la mediana, es $d\left(Me\right)=\frac{n+1}{2}=\frac{5+1}{2}=\frac{6}{2}=3$.

El 3 indica la posición del valor de la mediana.

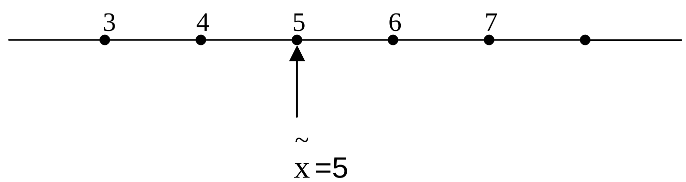
2. Calcule la mediana para los datos 6, 8, 7, 5, 3, 7.
Solución:
Ordenar los datos: 3, 5, 6, 7, 7, 8. Como el número de datos es par (n=6), se tiene que la posición de la mediana está dada por:

$$d\left(Me\right)=\frac{n+1}{2}=\frac{6+1}{2}=\frac{7}{2}=3.5$$

Esto significa que el valor de la mediana está entre el tercer y cuarto valor, por tanto

$$Me=\frac{6+7}{2}=\frac{13}{2}=6.5$$

Al igual que la media, la mediana se puede representar gráficamente, así los datos 3, 4, 5, 6 y 7 se ordenan de menor a mayor.



El 5 está en la tercera posición, o sea la central de los cinco números. Por lo tanto la mediana es 5.
No tiene sentido hallar la mediana, por ejemplo, del color del cabello de un grupo, pues no hay posibilidad de ordenar los colores de menor a mayor.

Ejercicios:

a. sean los datos 3,4,4,5,6,7,8,8,9,9,10. (n =11). Calcule la Me y la media.

b. Sean los datos que representan el número de transacciones en moneda extranjera 12 días hábiles (2 semanas) en un banco dado 9,7,9,7,10,8,12,14,13,16,14,15. (n =12). Calcule la Me y la media.

La moda

Es la medida de tendencia central que más fácilmente se determina. Para datos libres es el valor que aparece con mayor frecuencia o sea el que más se reporta. Se denota por Mo.

Ejemplos:

1. En el conjunto de datos 3, 3, 5, 6, 8,5,3,8 la moda es 3.
2. Algunos conjuntos de datos pueden tener dos modas (bimodal). La moda en el siguiente conjunto de datos 1, 4, 4, 4, 7, 3, 3, 4, 2, 7, 8, 3, 3, son 3 y 4.

¿un conjunto de datos puede no tener moda?.

3. El responsable de una biblioteca de cierta Escuela ordenó un estudio del tiempo que un alumno tiene que esperar (en minutos) para que le sea entregado el libro solicitado para consulta. Los datos fueron tomados durante un día normal a una muestra de 20 alumnos: 6, 9, 5, 5, 7, 3, 5, 2, 4, 9, 8, 2, 3, 7, 7, 8, 3, 3, 4, 3

a) hallar: media, moda, mediana

b) ¿Cuánto tiempo máximo, se debe suponer que el 75% de los alumnos debe esperar para obtener su libro de consulta?

Solución:

a) $\overline{X} = \dfrac{6+2*8+3*5+3*7+5*3+2*4+2*2+2*9}{20} = 5.15$

Como hay 20 datos y ordenando los datos tenemos que la mediana es Me=5, (es el promedio de los valores que están en la posición 10 y 11, el cual resulto ser 5 en los dos casos)

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

La moda Mo es 3 es el dato que más se repite, con una frecuencia de 5.

Ejercicios:

1. Se analizan los notas de 20 alumnos en el curso de matemática recogiéndose los siguientes datos: 38, 44, 62, 52, 66, 42, 60, 72, 96, 80, 42, 66, 60, 36, 70, 54, 60, 78,78, 84

a. Cuantos estudiantes aprobaron el curso, en porcentaje?

b. Cuál es la nota promedio de los alumnos?

c. Localicé la Mediana.

d. Cuál es la moda

2. El Profesor Pérez tiene 6 hijos, de los cuales 3 son trillizos y 2 mellizos. Si al calcular la media la mediana y moda de estas edades resultaron 10, 11 y 12 respectivamente. Halle la diferencia entre la máxima y mínima edad.

Guía de trabajo

1. Encuentre el peso promedio de l@s alumn@s de una aula de mi escuela.

2. En el conjunto de datos 2, 7, 15, 25, 35, 25, 12, 25, 10, 25; encuentre: la media, la moda y la mediana.
3. Recabar los meses de nacimiento de cada uno de ustedes. Encuentre la moda.
4. Averigua los años de servicios de l@s profesores de tu aula. ¿Cuál es la media? ¿Cuál es la moda? ¿Cuál es la mediana?

Consideraremos ahora que los datos están agrupados es decir están representados en una TDF.

MEDIA ARITMETICA

La media aritmética, representa el valor promedio de los datos, se denota por:

$\bar{\mathbf{x}}$: media aritmética para una muestra

μ : media aritmética para una población.

Estaremos utilizando la notación para una muestra, aunque los cálculos son validos para las medidas o parámetros poblacionales.

Para datos agrupados se define como:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i m_i}{n}$$

Dónde:

f: es la frecuencia absoluta,
m: es la marca de clase y
n: el número de observaciones.
k: número de clases

LA MEDIANA

Es un valor central que tiene la característica de dividir en dos partes iguales las observaciones. Un 50% de las observaciones son menores o iguales a la mediana y el otro 50% mayor o igual, se denota por M_e.

Para datos agrupados se define:

$$M_e = Li + \left[\frac{\frac{n}{2} - F}{f} \right] . C$$

Dónde:

Li: Límite inferior real del intervalo que contiene a la M_e.
F: Frecuencia acumulada anterior al intervalo que contiene a la M_e.
f: Frecuencia absoluta del intervalo que contiene a la M_e.
C: Amplitud de clase.

Para obtener la Mediana, identificamos la clase que contiene a la mediana, la cual será donde la primera frecuencia acumulada es mayor o igual a: n/2.

LA MODA:

Para datos agrupados, la moda está en la clase de mayor frecuencia, dicha clase es la clase modal. La moda si existe puede ser no única y se denota por M_o, se define como:

$$M_o = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] . C$$

Dónde:

Li: Límite interior real del intervalo Modal.
Δ_1: Mayor frecuencia menos la inmediata anterior
Δ_2: Mayor frecuencia menos la inmediata posterior
C: Amplitud de Clases

MEDIDAS DE VARIABILIDAD:

Son medidas que expresan variabilidad o dispersión alrededor de un valor promedio. Una vez localizado el centro de distribución de un conjunto de datos nos vemos en la necesidad de medir el grado en que se dispersan o diseminan los datos alrededor del valor promedio.

Una característica de casi todos los datos es que los valores no son todos iguales de ser así el grado de dispersión alrededor del promedio sería cero. Para medir el grado de dispersión lo hacemos a través de la varianza y desviación estándar.

Varianza

Expresa variabilidad al cuadrado de los datos alrededor del valor promedio se representa por S^2 (para una muestra) y como σ^2 (para una población) y se obtiene:

Datos agrupados

$$S^2 = \frac{\sum m_i^2 f - n\bar{x}^2}{(n-1)}$$

mi: marca de clase
f: frecuencia
n: total de datos

Datos no agrupados

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

x: datos
$\bar{x}$: media aritmética
n: total de datos

La varianza es un valor no negativo.

Desviación Estándar

Expresa variabilidad lineal de los datos alrededor del valor promedio. Es la raíz cuadrada de la varianza.
Se representa por: S: para una muestra y como: σ: para una población.

Se obtiene: tanto para datos agrupados como para no agrupados.

Coeficiente de Variación

$$C_V = \frac{s}{\bar{x}}$$

Se utiliza para comparar la variabilidad de dos o más distribuciones, será más representativa (menos variabilidad) la que tenga el C_V más pequeño.

Ejemplos:

1. El número de accidentes en 60 operaciones de máquinas que fueron analizados en 3 meses consecutivos, se dan en la tabla siguiente:

# de accid.	frecuenc.
0	27
1	12
2	6
3	6
4	3
5	3
6	2
7	1

Utilizando expresiones correspondientes a estos datos, hallar: Moda, Mediana, Media y la desviación estándar. (Interprete)
Solución:
Este ejercicio conviene resolverlo para datos no agrupados, para la moda y mediana:

Mo = 0 (lo más frecuente es que ocurra cero accidentes en las 60 operaciones de máquinas)
Me = 1, (en el 50% de las operaciones de máquinas, el # de accidentes es menor que 1)

Para calcular la media y desviación estándar se pueden usar las fórmulas para datos agrupados, (considerando el # de accidentes como las marcas de las clases)

$$\overline{X} = \frac{(0)(27)+(1)(12)+(2)(6)+(3)(6)+(4)(3)+(5)(3)+(6)(2)+(7)(1)}{60} = 1.467$$

El análisis lo puede hacer a conveniencia es decir puedo aproximarlo a 1 o a 2, en este caso respetaremos las normas de redondeo y aproximaremos a 1.
Para la desviación estándar obtenemos el siguiente valor con el uso de la formula

$$S^2 = \frac{(0)(27)+(1)(12)+(4)(6)+(9)(6)+(16)(3)+(25)(3)+(36)(2)+(49)(1)-60(1.467)^2}{59}$$

$$= \frac{334-60(2.1521)}{59} = \frac{204.8747}{59} = 3.472$$

De aquí que $S = \sqrt{3.472} = 1.863$

Comparado con la media este valor es muy alto, lo que indica que los datos están bastantes variables.

4. Los siguientes son las ganancias de cierre en miles de córdobas de dos almacenes en cinco viernes consecutivos:

Almacén A: 18, 17 5/8, 18 ¼, 17 3/8, 18 1/8
Almacén B: 20 ¼, 20, 19 7/8, 20 ½, 20 3/8

¿Cuál de estos almacenes tiene ganancias de cierre más estable?

Solución:
Primero hay que calcular la media de cada almacén y luego la desviación estándar.

$$\overline{X_A} = \frac{18+17\frac{5}{8}+18\frac{1}{4}+17\frac{3}{8}+18\frac{1}{8}}{5} = 17\frac{7}{8}$$

$$\overline{X_B} = \frac{20\frac{1}{4}+20+19\frac{7}{8}+20\frac{1}{2}+20\frac{3}{8}}{5} = 20\frac{1}{5}$$

$$S_A{}^2 = \frac{18^2+(17\frac{5}{8})^2+(18\frac{1}{4})^2+(17\frac{3}{8})^2+(18\frac{1}{8})^2-5\left(17\frac{7}{8}\right)^2}{4}$$

$$= \frac{1598.1094-1597.5781}{4} = 0.1328$$

Entonces de aquí $S_A = \sqrt{0.1381} = 0.3644$

$$S_B^2 = \frac{\left(20\frac{1}{4}\right)^2+\left(20\right)^2+\left(19\frac{7}{8}\right)^2+\left(20\frac{1}{2}\right)^2+\left(20\frac{3}{8}\right)^2-5\left(20\frac{1}{5}\right)^2}{4}$$

$$= \frac{2040.4687-2040.2000}{4} = \frac{0.2687}{4} = 0.0672$$

Luego $S_B = \sqrt{0.0672} = 0.2592$

De lo anterior calculamos

$$CV_A = \frac{S_A}{\overline{X}_A} = \frac{0.3644}{17\frac{7}{8}} = 0.02038,\quad CV_B = \frac{S_B}{\overline{X}_B} = \frac{0.2592}{20\frac{1}{5}} = 0.01283$$

Por lo tanto el almacén B tiene precios de cierre más estables.

5. Calcular la media, la mediana, la moda y la desviación estándar de la siguiente distribución de frecuencias

CALIFICACIONES	f_i	mi	f_ix_i
22 - 32	1	27	27
33 - 43	2	38	76
44 - 54	5	49	245
55 - 65	2	60	120
66 - 76	9	71	639
77 - 87	9	82	738
88 - 98	10	93	930
99 - 109	5	104	520
110 - 120	3	115	345
121 - 131	4	126	504
	$\Sigma f_i = 50$		$\Sigma f_ix_i = 4144$

Solución

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \cdot m_i}{\sum_{i=1}^{n} f_i} = \frac{4144}{50} = 82.88 \approx 82.9$$

$$Me = 77 + \left(\frac{25-19}{9}\right)11 = 77 + \frac{6}{9} * 11 = 84.333$$

$$Mo = 88 + \left(\frac{1}{1+5}\right)11 = 88 + \frac{1}{6} * 11 = 89.83$$

$$S = \sqrt{\frac{1*27^2 + 2*38^2 + 5*49^2 + 2*60^2 + 9*71^2 + 9*82^2 + 10*93^2 + 5*104^2 + 3*115^2 + 4*126^2 - 50*82.88^2}{49}}$$

$$= \sqrt{\frac{372456 - 343454.72}{49}} = \sqrt{\frac{29001.28}{49}} = \sqrt{591.863} = 24.33$$

6. Un laboratorio de medicamentos utiliza camiones ligeros, sus controles reveló las siguientes millas recorridas por galón de combustible consumido. Calcular la media y la mediana.

NUMERO DE MILLAS RECORRIDAS	NUMERO DE CAMIONES
10 - 13	2
13 - 16	5
16 - 19	10
19 - 22	8
22 - 25	3
25 - 28	2
TOTALES	30

Solución

La media representa el consumo de todos los camiones, es decir 18.6 mph representa las millas que recorren todos los camiones.

$$\overline{X} = \frac{2*11.5 + 5*14.5 + 10*17.5 + 8*20.5 + 3*23.5 + 2*26.5}{30} = \frac{558}{30} = 18.6$$

La clase donde está la mediana es la tercera o sea la 16 – 18, dado que ahí cae el valor $\frac{N}{2}$ = 15. Ahora observemos el valor que toman:

Li = 16, $\frac{N}{2}$ = 15, F_a = 7, f = 10 y C = 3

Por tanto, $M_e = 16 + \left(\frac{15 - 7}{10}\right) \cdot 3 = 16 + (0.8) \cdot 3 = 16 + 2.4 = 18.4$

Esto significa que el 50% de los camiones ligeros, recorren menos de 18.4 millas por galón de combustible consumido y el otro 50% recorre más de 18.4 millas por galón de combustible empleado. Ese valor, 18.4 mph, está en el centro de todos los datos.

7. Una máquina automática de una embolletadora de refrescos, llena vasos de refresco para ser distribuido en algunos centros turísticos de Managua y parece que está trabajando de manera errática. Una verificación de los pesos del contenido de un cierto número de vasos reveló lo siguiente:

PESOS (EN GRAMOS)	NUMERO DE VASOS
130 - 139	2
140 - 149	8
150 - 159	20
160 - 169	15
170 - 179	9
180 - 189	7
190 - 199	3
200 - 209	2
TOTALES	66

Solución
La clase modal es la tercera o sea la 150 - 159; con una f = 20 que es la mayor frecuencia absoluta. De esa manera:
Lri = 149.5 y los valores de $\Delta_1 = 20 - 8 = 12$; de $\Delta_2 = 20 - 15 = 5$ y el de C = 10, luego aplicamos la fórmula

$$M_o = Li + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) \cdot C = 149.5 + \left(\frac{12}{12+5}\right)(10) = 149.5 + \frac{120}{17} = 149.5 + 7.058 = 156.56$$

Calcule la media y la mediana de este ejercicio.

Ejercicios:
1. El Contador de un almacén desea estimar el balance promedio en dólares de las 1000 cuentas de
crédito que maneja el almacén. La distribución de frecuencias se presenta en la tabla y fue construida a partir de una muestra de 100 cuentas seleccionadas al azar en los archivos de crédito del almacén.

BALANCE GENERAL	NUMERO DE CUENTAS
0-20	10
20-40	15
40-60	40
60-80	22
80-100	13
	100

Calcular la media, la mediana, la moda y la desviación estándar de las cuentas de crédito que maneja el almacén.

2. Los siguientes datos representan el número de interrupciones en 15 días de trabajo debidas a fallas mecánicas en una planta procesadora de alimentos 3, 0, 5, 1, 3, 1, 3, 2, 2, 0, 2, 1, 2, 4, 3 . Calcule el número promedio de interrupciones y la desviación estándar del número de interrupciones.

3. Dos países A y B venden la misma materia prima en el mercado mundial a los siguientes precios por kilogramo en el trascurso de 6 meses:

Mes	Cotizaciones en dólares por País	
	A	B
1	4.9	2.9
2	5.0	3.8
3	2.6	3.0
4	4.5	3.5
5	2.3	3.7
6	4.1	5.0

Realice un análisis de precios de este producto para ambos países comparando los coeficientes de variación y diga a que país se le presentan condiciones de mercado más favorables.

Trabajo en grupo o individual:

1. Una empresa industrial agrupó sus fábricas de acuerdo con el valor de la producción anual de cada una, se obtuvo la siguiente distribución

Producción (millones de córdobas)	Número de Fábricas
40-45	7
45-50	10
50-55	11
55-60	9
60-65	8
65-70	7

a. Determine la producción anual promedio de las fábricas
b. Determine la desviación estándar de las producciones
c. Calcule el coeficiente de variación

2. La siguiente tabla muestra los gastos en energía eléctrica en un mes de 50 casas de un barrio de Managua

Gastos en KW/MES	Número de casas
75-100	4
100-125	8
125-150	15
150-175	13
175-200	7
200-225	3

Calcule el gasto promedio de las 50 casas

a. Calcule la mediana y la moda en el gasto
b. Calcule la desviación estándar

3. Considere el costo de la canasta básica de 10 ciudades de Nicaragua: 9500, 8500, 8500, 10000, 9000, 9500, 10500, 11000, 10500, 9500.
a. Calcule la media y la desviación estándar.
b. Encuentre la mediana y la Moda

4. La siguiente tabla muestra los consumos quincenales registrados (cienes de córdobas) en 30 familias de un barrio del departamento de Managua

Gasto quincenal (cienes de córdobas)	Números de familias
34-42	4
42-50	8
50-58	10
58-66	5
66-74	3

a. Calcule el consumo promedio de las 30 familias en una quincena
b. Calcule la desviación estándar del consumo promedio de las 30 familias
c. Encuentre la mediana y la moda en el consumo quincenal