

INSTITUTO TÉCNICO INDUSTRIAL FRANCISCO JOSE DE CALDAS

FUNDAMENTOS DE ESTADÍSTICA

La Estadística método científico utilizado en la recolección, organización, análisis, interpretación numérica de la información y su respectiva representación, ya sea en tablas o gráficas.

Estadística deductiva o descriptiva: Presenta una serie de datos sin llegar a conclusiones.

Estadística Inductiva o de inferencia: **Llega** a conclusiones de un grupo mayor, basados en la información de un grupo menor. Permite predecir actuaciones.

Estadística analítico-descriptiva: Conjunto sistemático de procedimientos para observar y describir numéricamente el fenómeno y describir las leyes que regulan la aparición, transformación y desaparición del mismo.

ELEMENTOS BÁSICOS EN ESTADÍSTICA

1. **Población:** Colección de individuos, objetos o medidas que tienen una característica común.
2. **Muestra:** Subconjunto de la población
3. **Distribución de frecuencia: Cuadro de datos organizados**
4. **Variable:** Es una característica de los elementos de la muestra. El valor de la variable es la medida de la característica. Las variables pueden ser de dos clases: Discretas y Continuas.
 - **Variables discretas:** Las variables que al tabularse toman solamente valores de tipo entero.
 - **Variables continuas:** Las variables que al tabularse pueden tomar valores decimales.
5. **Frecuencia Absoluta:** Número de datos que hay en cada intervalo. Se denota con la letra f.
6. **Frecuencia acumulada:** Se obtiene sumando cada frecuencia absoluta con las anteriores. La última debe ser igual a la totalidad de los casos.
7. **Frecuencia relativa:** Se obtiene dividiendo cada frecuencia absoluta por el número total de datos
8. **Frecuencia relativa acumulada:** Se obtiene sumando cada una de las frecuencias relativas con todas las anteriores.
9. **Intervalo:** Número de categorías en que se divide una variable. También se llama **clase**.
10. **Amplitud:** Diferencia entre dos límites inferiores o dos límites superiores de dos intervalos consecutivos.
11. **Marca de Clase:** Punto medio de cada intervalo y es el valor representativo de cada uno de ellos.

EJEMPLO: Se quiere averiguar el salario promedio de los vendedores ambulantes de la capital. Para saberlo, entrevistamos a 100 personas dedicadas a esta actividad. Determinar los elementos básicos de la Estadística:

Solución.

Población: Los salarios de todos los vendedores ambulantes.

Muestra: Los 100 salarios de las personas encuestadas.

Distribución de frecuencias:

SALARIOS (miles de pesos)	No. DE VENDEDORES
200-299	9
300-399	10
400-499	31
500-599	22
600-699	16
700-799	12
TOTAL	100

Variables: Salarios, vendedores ambulantes. Los salarios es una variable cuantitativa continua porque puede tomar cualquier cantidad de dinero entre \$200.000 y \$799.000.

Intervalo: Ejemplo. 22 de la columna número de vendedores significa que 22 vendedores ambulantes están en la categoría o intervalo de los que ganan entre \$500.000 y \$599.000 mensuales

Amplitud: Ejemplo: entre 400 y 3000 ó entre 499 y 399 el resultado es 100. Cada intervalo tiene una amplitud o tamaño igual a 100.

Marca de clase: Para hallarla tomamos cualquier intervalo y hallamos su valor promedio; para hallarlo tomamos el límite inferior y el límite superior del intervalo y se divide entre 2. Por ejemplo: $\frac{300+399}{2} = 349.5$ Este valor representa la marca de clase de ese intervalo y es el valor más representativo de cada intervalo.

SALARIOS (MILES DE PESOS)	N. DE VENDEDORES FRECUENCIA	MARCA DE CLASE	AMPLITUD	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
200-299	9	250	100	9	9/100= 9%	9%
300-399	10	350	100	19	10/100=10%	19%
400-499	31	450	100	50	31/100=31%	50%
500-599	22	550	100	72	22/100=22%	70%
600-699	16	650	100	88	16/100=16%	88%
700-799	12	750	100	100	12/100=12%	100%
TOTAL	100				100/100=1=100%	

EJERCICIOS:

Las notas obtenidas en 801 conformado por 30 estudiantes fueron:

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

Teniendo en cuenta la información anterior Determinar: Población, Muestra, tipo de variable, Tamaño de la muestra, tabla de frecuencia, frecuencia absoluta, acumulada y relativa, elabore una distribución de frecuencias por intervalo de amplitud 3, encuentre la marca de clase de cada intervalo

MEDIDAS DE TENDENCIA CENTRAL

La Media Aritmética: es un valor numérico definido como el coeficiente entre la suma de los valores de las variables y el número total de datos Se denota por \bar{x} que se lee equis barra.

La Mediana: La mediana de una serie ordenada de datos es aquel valor que se encuentra equidistante de los extremos de ella. Se denota Me.

La Moda: Es el dato o valor que más se repite en una serie.

EJEMPLO: Se pregunta a 40 familias por el número de hijos de cada una de ellas y se obtienen los siguientes datos

Variable (No. De hijos)	FRECUENCIA	VARIABLE POR FRECUENCIA	FRECUENCIA ACUMULADA
0	4	0	4
1	12	12	16
2	7	14	23
3	10	30	33
4	7	28	40
5	$\sum f = 40$	$\sum (v.f) = 84$	

$$Media = \frac{Suma\ de\ (variable\ por\ frecuencia)}{Número\ de\ datos} = \frac{84}{40} = 2.1$$

En promedio cada familia tiene 2.1 hijos, que se puede aproximar a 2 por ser una variable discreta.

Mediana = 23

Moda = 12

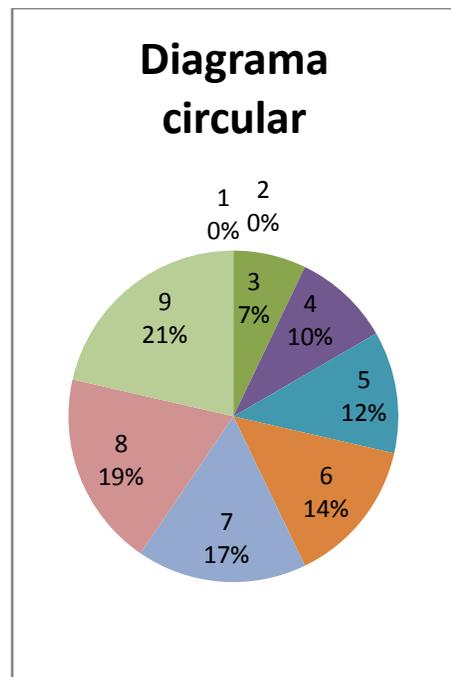
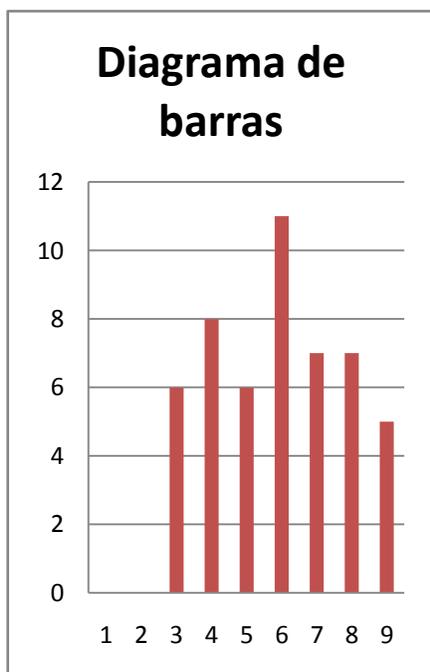
REPRESENTACIÓN GRAFICA DE DATOS

- **DIAGRAMA DE BARRAS**
- **DIAGRAMA CIRCULAR**
- **HISTOGRAMAS**
- **POLOÍGONOS DE FRECUENCIA**

DIAGRAMAS DE BARRAS. Las barras que se usan son rectángulos separados entre sí y contruidos en forma vertical, la altura del rectángulo está representada por la frecuencia absoluta de cada dato y va en el meje vertical del plano.

DIAGRAMA CIRCULAR. Hace notar las diferencias en las proporciones o porcentajes. El proceso para realizar el diagrama circular (pastel) consiste en dividir los 360° de la circunferencia, proporcionalmente al valor de cada frecuencia absoluta.

Dato	f
0	0
0	0
3	6
4	8
5	6
6	11
7	7
8	7
9	5

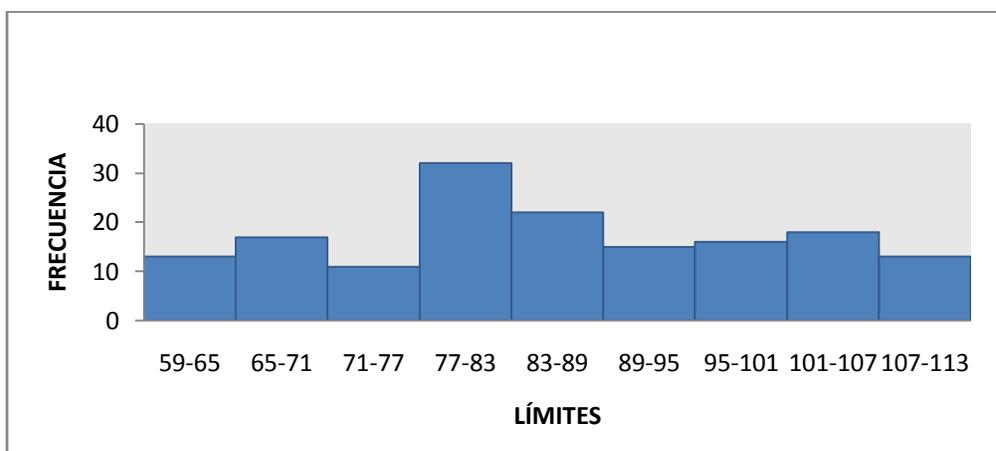


HISTOGRAMAS. Es una forma de presentar gráficamente las frecuencias de los distintos intervalos.

Se diferencia de los diagramas de barras en detalles como:

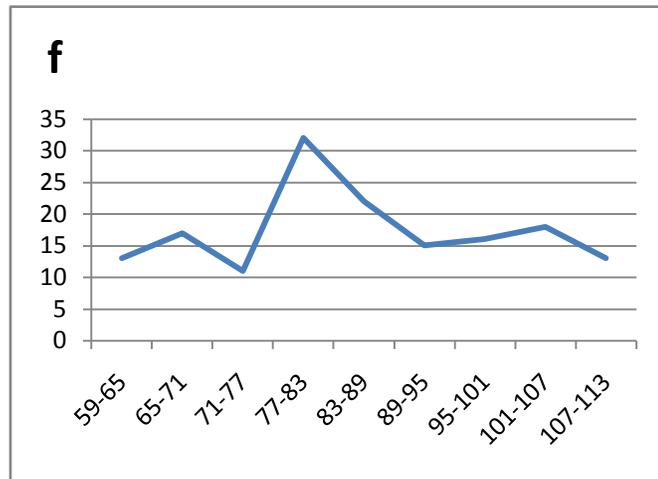
- La frecuencia no está determinada por la altura de los rectángulos sino por el área de ellos, se suele tomar como base de cada rectángulo una dimensión igual a uno.
- Los rectángulos se dibujan generalmente sin dejar espacios entre ellos.

Límites reales	f
59-65	13
65-71	17
71-77	11
77-83	32
83-89	22
89-95	15
95-101	16
101-107	18
107-113	13



Polígonos de frecuencia. En el histograma se marca en la parte superior el punto medio (correspondiente a la marca de clase), unimos todos los puntos por medio de segmentos de recta y obtenemos el polígono de frecuencias.

Límites reales	f
59-65	13
65-71	17
71-77	11
77-83	32
83-89	22
89-95	15
95-101	16
101-107	18
107-113	13



EJERCICIO

Los siguientes datos corresponden al número de goles anotados en 45 fechas del campeonato nacional de fútbol.

datos
número de goles

16, 20, 14, 18, 24, 32, 19, 23, 16, 18, 23, 17, 30, 24,
26, 18, 25, 31, 17, 26, 28, 21, 28, 28, 22, 20, 20, 24, 19, 29, 15, 23,
30, 27, 23, 25, 18, 23, 19, 17, 16, 24, 23, 28, 30, 18, 21

- **Elabore la tabla de frecuencia acumulada**
- **Elabore una distribución de frecuencias por intervalo de amplitud 3**
- **Encuentre la marca de clase de cada intervalo**
- **Represente la situación en un diagrama de barras, diagrama circular, histograma y diagrama poligonal**
- **En cuántas fechas se marcaron entre 28 y 30 goles?**
- **Qué intervalo corresponde al mayor número de fechas?**
- **Qué intervalo corresponde a 5 fechas?**
- **Halle la media, la mediana y la moda**

PERCENTILES

Percentil: Es el valor en que se divide porcentualmente el conjunto de datos. Sirven para indicar la posición de un valor dentro de un conjunto de datos ordenados.

EJEMPLO: Qué significa que el puntaje de 137 esté en rango percentil de 68?

Significa que el 68% del grupo obtuvo una calificación menor o igual a 137 o también que el 32% tiene una calificación igual o superior a este resultado.

CALCULO DE PERCENTIL:

- 1. Se construye un gráfico de distribución de frecuencia**
- 2. Se elabora el gráfico correspondiente a la distribución de frecuencias relativas acumuladas.**
- 3. Buscar el percentil que corresponde localizando este número en el eje de las abscisas (eje X) y se traza una perpendicular al eje Y, por este punto que corte la curva.**

EJEMPLO: Buscar el percentil que corresponde al puntaje de aptitud física de Fernando que fue de 125., según la información suministrada por el entrenador.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

PUNTAJE	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
0-19	5	5	$\frac{5}{90} = 0.055 = 5.5\%$	5.5%
20-39	3	8	$\frac{3}{90} = 0.033 = 3.3\%$	8.8%
40-59	12	20	$\frac{12}{90} = 0.133 = 13.3\%$	22.1%
60-79	10	30	$\frac{10}{90} = 0.111 = 11.1\%$	33.2%
80-99	14	44	$\frac{14}{90} = 0.155 = 15.5\%$	48.7%
100-119	23	67	$\frac{23}{90} = 0.255 = 25.5\%$	74.2%
120-139	11	78	$\frac{11}{90} = 0.122 = 12.2\%$	86.4%
140-159	12	90	$\frac{12}{90} = 0.133 = 13.3\%$	99.7% \cong 100%
TOTAL 90				

EJERCICIO: Los siguientes datos corresponden a una prueba de oratoria y otra de ortografía realizadas el día del idioma.

Puntaje de prueba de oratoria

PUNTAJE	FRECUENCIA (No. Estudiantes)
0-2.9	4
3-5.9	8
6-8.9	5
9-11.9	15
12-14.9	24
15-17.9	18
18-20.9	9

Puntaje de la prueba ortográfica

PUNTAJE	FRECUENCIA
0-4.9	7
5-9.9	13
10-14.9	21
15-19.9	14
20-24.9	8
25-30	11

De acuerdo con esta información:

- a. Cuántos estudiantes se presentaron a cada prueba?
- b. Copia en tu cuaderno y completa la distribución de frecuencia agregando la frecuencia acumulada, la frecuencia relativa y la frecuencia relativa acumulada.
- c. Con los datos obtenidos elabore un polígono de frecuencias relativas acumuladas.
- d. Si Laura obtiene 10 puntos en la prueba de oratoria y 14 puntos en la prueba de ortografía, en cuál de las dos pruebas le fue mejor?
- e. Si Miguel obtiene 18 puntos en la prueba de ortografía, en qué cuartil será ubicado?
- f. Con la gráfica que realizaste, qué puntaje corresponde aproximadamente al 60% de la prueba de oratoria?
- g. Qué puntajes están por debajo del 50% en la prueba de ortografía?
- h. Qué puntajes están por encima del primer cuartil en la prueba de oratoria?
- i. Averigua cómo se llaman las particiones de 10 en 10 que se hacen en una distribución de frecuencias relativas acumuladas.

MEDIDAS DE DISPERSIÓN

Las **medidas de dispersión** indican qué tan lejos o cerca se encuentran unos datos de otros en una distribución de frecuencias. La medida representativa más utilizada para analizar la dispersión de datos es la **media aritmética**.

EJEMPLO: Si Pedro obtiene una calificación de 75 puntos en un grupo cuya media fue 68 puntos. Cómo se ubica de acuerdo a la media?

Para comparar la calificación de Pedro, hallamos la distancia de su calificación a la media: $75-68=7$.

La calificación de Pedro se encuentra 7 puntos por encima de la media.

Analizaremos las siguientes medidas de dispersión: Rango, varianza, desviación media, desviación estándar.

1. **RANGO:** Medida de dispersión que indica la distancia entre el valor mayor y el valor menor en un grupo de datos.

EJEMPLO: Cuál es el rango en los puntajes de una prueba de levantamiento de pesas, si quien ocupó el primer puesto obtuvo 254 puntos y el último 137?

Solución: Para determinar el rango se halla la diferencia entre el puntaje más alto y el puntaje más bajo así:

$$\text{Rango} = 254 - 137 = 117 \text{ puntos}$$

EJERCICIO: Hallar el rango en cada caso:

- En una prueba de aptitud verbal la nota más alta fue 125 y la más baja 37.
- El mejor encestador del equipo de baloncesto anotó 48 puntos y el jugador que hizo menos cestas anotó 37.

2. **DESVIACIÓN MEDIA (D.M):** Es una medida de dispersión que indica la medida de las distancias de los datos a la media aritmética. Se obtiene sumando las distancias de los datos respecto a la media y luego dividiendo por el número de datos.

EJEMPLO: El equipo de baloncesto del ITI necesita seleccionar un nuevo jugador. Los candidatos para ingresar son: Andrés y Pablo, cuyas anotaciones en los 5 últimos partidos son las siguientes:

Jugadores	P ₁	P ₂	P ₃	P ₄	P ₅	Media \bar{x}	Rango
Andrés	3	5	2	4	6	4	4
Pablo	7	1	2	7	3	4	6

Qué hacer para escoger al candidato? S debe comparar cada puntaje con la media

$x - \bar{x}$ Andrés
3-4= -1
5-4= 1
2-4= -2
4-4= 0
6-4= 2
Total 0

$x - \bar{x}$ Pablo
7-4= 3
1-4= -3
2-4= -2
7-4= 3
3-4= -1
Total 0

$ x - \bar{x} $ Andrés
3-4= -1 =1
5-4= 1 =1
2-4= -2 =2
4-4= 0 =0
6-4= 2 =2
Total 6

$ x - \bar{x} $ Pablo
7-4= 3 =3
1-4= -3 =3
2-4= -2 =2
7-4= 3 =3
3-4= -1 =1
Total 12

Para hallar la desviación media se divide el total de cada tabla entre el número de datos:

- La desviación media para Andrés es de $\frac{6}{5} = 1.2$
- La desviación media para Pablo es de $\frac{12}{6} = 2.4$. Las anotaciones son más dispersas con respecto a la media.

EJERCICIO: Escriba el promedio de las desviaciones con respecto a la media de los siguientes datos: {23,19,46,35,21,50}

3. **VARIANZA:** La varianza es una medida de dispersión que se halla mediante la suma de los cuadrados de las desviaciones respecto a la media dividida entre el número de datos. Se simboliza S^2 .

EJEMPLO: Cuál es la varianza de los datos: 67, 43, 75, 39 y 62

Primero calculamos el promedio de los datos $\bar{X} = \frac{67+43+75+39+62}{5} = \frac{286}{5} = 57.2$

Después hacemos la diferencia de cada calificación con la media y elevamos esa diferencia al cuadrado.

$X - \bar{X}$	$(X - \bar{X})^2$
39-57.2= -18.2	331.24
43-57.2=-14.2	201.64
62-57.2=4.8	23.04
67-57.2=9.8	96.04
75-57.2=17.8	316.84
0	968.80

La varianza es el promedio de estos cuadros.

$$\text{Varianza} = \frac{968.8}{5} = 193.76 \text{ puntos}^2$$

4. **DESVIACIÓN ESTÁNDAR:** La desviación estándar es una medida de dispersión que indica qué tan alejado, en promedio, se encuentra un conjunto de datos en relación con su promedio.
La desviación estándar se obtiene extrayendo la raíz cuadrada de la varianza. Se simboliza con la letra S, y conserva la unidad de medida de los datos.

PROCEDIMIENTO PARA HALLAR LA DESVIACIÓN ESTÁNDAR

JUGADORES	NÚMEROS DE GOLES X
Alberto	8
Bernardo	7
Carlos	7
Danilo	6
Eduardo	5
Fabio	5
Germán	5
Héctor	3
Ignacio	1
Jaime	1

Figura 1

X	$X - \bar{X}$	$(X - \bar{X})^2$
8	3.2	10.24
7	2.2	4.84
7	2.2	4.84
6	1.2	1.44
5	0.2	0.04
5	0.2	0.04
5	0.2	0.04
3	-1.8	3.24
1	-3.8	14.44
1	-3.8	14.44
48		53.60

Figura 2

La figura 1 . muestra los 10 primeros jugadores de un campeonato de fútbol.

Analicemos que tan alejados se encuentran en promedio de la media del número de goles.

- Se calcula el promedio del total de goles $\bar{X} = \frac{48}{10} = 4.8$ goles En promedio cada jugador marcó 4.8 goles.
- Se hace la diferencia de cada dato con el promedio.
- Cada una de estas diferencias se eleva al cuadrado
- Se suman los resultados anteriores: Suma $(X - \bar{X})^2$ (ver figura 2)
- Se calcula la varianza dividiendo el resultado anterior por el número de datos: $\text{Varianza} = \frac{\text{Suma de } (x - \bar{x})^2}{\text{Número de datos}} = \frac{53.6}{10} = 5.36 \text{ goles}$
- Para conocer la desviación estándar se extrae la raíz cuadrada de la varianza:

$$\text{Desviación estándar} = \sqrt{\text{varianza}} = S = \sqrt{5.36 \text{ goles}^2} = 2.315 \text{ goles}$$

Los jugadores entonces se encuentran a 2.315 goles del promedio general de anotaciones.

EJERCICIO:

Concesionario	No. De vehículos
A	51
B	37
C	48
D	41
E	37
F	43
G	54
H	41
I	46
J	50

En una encuesta se establece el número de carros vendidos en el mes de abril por 10 concesionarios (ver tabla).

De acuerdo con esta información:

- Halla la media de carros vendidos durante el mes de abril.
- Cuál es el rango de esta distribución?
- Cuáles son los concesionarios que están por encima de la media?
- Qué concesionarios deben incrementar su publicidad para alcanzar su promedio mensual?
- Calcula la varianza de datos
- Cuál es el promedio de las desviaciones con respecto a la media?
- Se puede afirmar que el concesionario F está más cerca del promedio mensual que el concesionario E?