



Código  
**CH-FyA-0504**

## Guía 88: Organizando datos agrupados

Guía

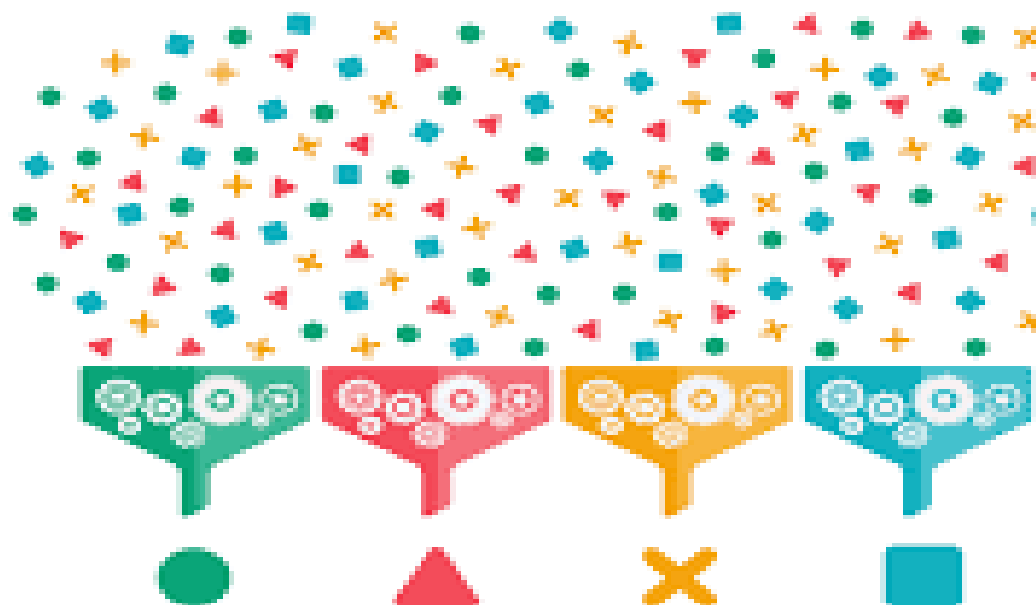
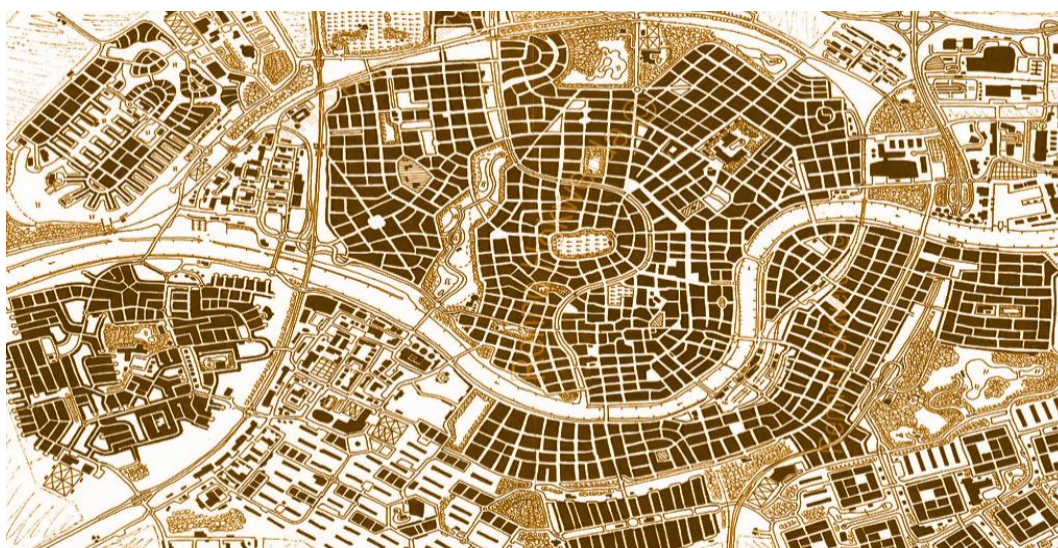
88

Meta 30

GRADO 9

GUÍA DEL ESTUDIANTE

# ORGANIZANDO DATOS AGRUPADOS





# Fe y Alegría

Movimiento de Educación Popular Integral y Promoción Social

## Guías de Aprendizaje de Cualificar Matemáticas Fe y Alegría Colombia

### Fe y Alegría Colombia

**Víctor Murillo**

**Director Nacional**

### Desarrollo de contenidos pedagógicos y educativos

**Jaime Benjumea - Marcela Vega**

### Autores de la guía 88

Silena Patricia Delbarre Lascarro, IED José Raimundo Sojo

Jeyson Rosales Mendoza, IED José Raimundo Sojo

### Coordinación pedagógica

Francy Paola González Castelblanco

Andrés Forero Cuervo

GRUPO LEMA [www.grupolema.org](http://www.grupolema.org)

### Revisores

Jaime Benjumea

Diego Fonseca Enciso, Colegio Las Mercedes

Francy Paola González Castelblanco

# ORGANIZANDO DATOS AGRUPADOS

## GRADO 9 - META 30 - PENSAMIENTO ALEATORIO

### Guía 88 (Duración 13 h)

#### ACTIVIDAD 1

- Comparación de dos o más poblaciones con respecto a una variable estadística.
- Comparación de medidas de tendencia central en dos o más conjuntos de datos.

#### ACTIVIDAD 2

- Aplicación: conjeturas sobre una población a partir de una muestra aleatoria.
- Variables bidimensionales

### Guía 89 (Duración 13 h)

#### ACTIVIDAD 1

- Medidas de dispersión: simetrías, cuartiles, diagramas de caja y bigotes.

#### ACTIVIDAD 2

- Introducción a hojas de cálculo y funciones básicas (promedio, moda, máximo, etc).

### Guía 90 (Duración 13 h)

#### ACTIVIDAD 1

- Permutaciones con y sin reemplazo, función factorial para conteo.
- Combinaciones y coeficiente binomial.

#### ACTIVIDAD 2

- Tablas, gráficas y diagramas estadísticos para comparar probabilidades observando simetrías y otros patrones.

**META DE APRENDIZAJE N. 30:** Explico semejanzas y diferencias entre dos poblaciones según atributos de mi interés, como la expectativa de edad en especies de animales, el nivel laboral según el género o el sueldo según la profesión, y lo uso para tomar decisiones que afectan a mi comunidad de forma positiva. Para ello, comparo medidas de tendencia (promedio, mediana, moda), y analizó la localización y dispersión de datos (simetría, cuartiles, diagrama de caja y bigotes) con hojas de cálculo; aprendo fórmulas de conteo (permutar con o sin reemplazo, factorial; combinar, coeficiente binomial) que aplico a mi vida diaria; comparó probabilidades usando diagramas de árboles, tablas de frecuencia y simetrías. Así, aprendo a comparar información de varias fuentes.

### PREGUNTAS ESENCIALES, GUÍA 88:

- ¿Qué relaciones y diferencias existen entre las variables cualitativas y cuantitativas estadísticas?
- ¿Qué significa que un conjunto de datos esté agrupado o no agrupado en un estudio estadístico?
- ¿De qué forma puedo comparar las medidas de tendencia central en dos o más conjuntos de datos?
- ¿Para realizar un estudio estadístico que técnica de recolección debo tener en cuenta?
- ¿Qué ventajas me ofrece organizar dos datos diferentes de una misma población en una tabla de doble entrada o en un diagrama de dispersión?

**EVIDENCIAS DE APRENDIZAJE, GUÍA 88**

- Realizo comparaciones entre dos o más poblaciones dada una variable estadística.
- Diferencio las medidas de tendencia central y las aplico a para datos agrupados.
- Comparo las medidas de tendencia central en datos agrupados para dos o más conjuntos de datos.
- Realizo estudios estadísticos con variables bidimensionales utilizando cualquier método de recolección de datos.
- Organizo y represento variables bidimensionales en tablas de doble entrada y diagramas de dispersión.
- Formulo conclusiones e hipótesis para una población a partir de información dada de una muestra.



## ACTIVIDAD 1: COMPARANDO MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS

Aprendamos a calcular las medidas de tendencia central para datos agrupados conociendo la población de estudio y teniendo en cuenta el tipo de variable estadística.

### A) Recordemos: población, muestra y variable estadística

#### RECUERDA QUE...

**LA POBLACIÓN:** Es un conjunto de datos que tienen una característica en común. Ejemplo: Los estudiantes de noveno grado de fe y alegría.

**LA MUESTRA:** es un subconjunto de esa población y que debe ser seleccionada de manera aleatoria. Ejemplo: Las niñas de noveno grado de fe y alegría.

**VARIABLE ESTADÍSTICA:** Es una cualidad o característica que posee un individuo. Ejemplo: color de ojos, edad, estatura, color favorito, etc. Las variables pueden ser cuantitativas y cualitativas.

- **Cuantitativas:** Son aquellas características o cualidades que se pueden contar o expresar a través de números. Esta se divide en dos, cuantitativa discreta y cuantitativa continua.
  - **Cuantitativa discreta:** aquella variable que utiliza valores enteros y no finitos. Por ejemplo, la cantidad de familiares que tiene una persona, tal como 2, 3, 4 o más.

- **Cuantitativa continua:** aquella variable que utiliza valores finitos y objetivos, y suele caracterizarse por utilizar valores decimales. Por ejemplo, el peso de una persona, tal como 64.3 kg, 72.3 kg, etc.

- **Cualitativas:** son aquellas características o cualidades que no pueden ser calculadas con números, sino que son clasificadas con palabras.

Se dividen en ordinales y nominales.

- **Cualitativa nominal:** aquellas variables que no siguen ningún orden en específico. Por ejemplo, los colores, tales como el negro, naranja o amarillo.
- **Cualitativa ordinal:** aquellas que siguen un orden o jerarquía. Por ejemplo, el nivel socioeconómico alto, medio o bajo.

**Práctica**

i) Para cada una de las siguientes situaciones identifica la población, muestra, variable estadística y tipo de variable:

- En un colegio de Fe y Alegría que cuenta con 1250 estudiantes, se realiza una encuesta a las mujeres para conocer su materia favorita. Completa lo siguiente:

Población: \_\_\_\_\_

Muestra: \_\_\_\_\_

Variable estadística: \_\_\_\_\_

Tipo de variable: \_\_\_\_\_

- Se les preguntó a los estudiantes de noveno grado de fe y alegría acerca de con cuántas personas viven en su casa.

Población: \_\_\_\_\_

Muestra: \_\_\_\_\_

Variable estadística: \_\_\_\_\_

Tipo de variable: \_\_\_\_\_

ii) Analiza la siguiente tabla y marca con una X en el tipo de variable que corresponde, ten en cuenta las siguientes convenciones:

**CN:** CUALITATIVA NOMINAL

**CO:** CUALITATIVA ORDINAL

**CD:** CUANTITATIVA DISCRETA

**CC:** CUANTITATIVA CONTINUA

	CN	CO	CD	CC
Color de ojos				
Número de goles realizados por un equipo.				
Desempeño académico				
Número de hijos				
Color favorito				
Estatura				

**Verifica las respuestas de la sección A con tu profesor.**

## B) Conceptos: Medidas de tendencia central para datos agrupados

### Exploración: Comparemos tendencias en dos barrios.

Realizas una encuesta a las mujeres de dos barrios de tu ciudad para conocer la edad en la que tuvieron su primer hijo(a). Quieres recoger esta información para analizar la edad promedio en la que las mujeres de tu barrio quedan en embarazo y compararla con otro barrio vecino.

En la tabla #1 te encuentras con los datos organizados según las edades y el número de mujeres que quedaron embarazadas en el barrio 1 y en la tabla #2 los datos según edades de las mujeres embarazadas en el barrio 2.

Tabla #1													
EDADES	14	15	16	17	18	19	20	21	22	23	24	25	26
# MUJERES	7	5	9	14	25	10	5	8	12	6	3	4	2

Tabla # 2											
EDADES	16	18	20	21	23	24	25	26	27	28	31
# MUJERES	8	21	15	13	16	10	9	5	7	4	2

Con la información de la tabla #1 te piden que calcules el promedio de edad en la que las mujeres del barrio 1 salen embarazadas por primera vez y que organices la información en una tabla con 7 intervalos o clases.

Para esto lo primero que debes hacer es calcular el **rango (R)**, restando el dato mayor menos el dato menor; es decir la edad máxima que en este caso es el 26 menos la edad mínima que en este caso es el 14.

$$R = \text{edad máxima} - \text{edad mínima} = 26 - 14 = 12$$





Con esto ya tienes el rango ( $R = 12$ ) y el intervalo ya sabemos que es  $[14, 26]$ .

Ahora vamos a calcular la amplitud ( $A$ ) dividiendo el rango entre el número de intervalos.

$$A = \frac{12}{7} = 1,7 \text{ cuando el resultado es decimal se aproxima.}$$

$$A = 2.$$

Con estos datos ya puedes organizar la información en la tabla como se muestra a continuación:

Edades	Marca de clase ( $x_i$ )	Frecuencia absoluta ( $f_i$ )
[14-16)	15	12
[16-18)	17	23
[18-20)	19	35
[20-22)	21	13
[22-24)	23	18
[24-26)	25	7
[26-28)	27	2
Total		110

Para calcular la marca de clase ( $x_i$ ), debemos sumar el límite inferior con el límite superior del intervalo y el resultado dividirlo entre la amplitud. Puedes guiarte de la siguiente fórmula:

$$X = \frac{\text{límite superior} + \text{límite inferior}}{A}, \text{ para el primer intervalo de la tabla [14-16) quedaría de la siguiente manera: } X = \frac{16+14}{2} = \frac{30}{2} = 15. \text{ De esta misma forma lo hacemos para el resto de los intervalos.}$$

Para la frecuencia absoluta ( $f_i$ ) debo sumar o contar cuantos datos entran en cada intervalo. Para el caso del primer intervalo [14-16) sumariamos las 7 mujeres de 14 años y las 5 de 15 años, quiere decir que  $7+5=12$ . La cantidad de mujeres que tuvieron su primer hijo a los 16 años no se sumaría en este intervalo sino en el siguiente.



**Responde:**

- a) Completa la siguiente tabla con la información de la tabla #2. Aquí utilizaremos 6 intervalos, el rango es igual a 15 y la amplitud es igual a 2,5 entonces lo aproximamos a 3.

Edades	Marca de clase ( $x_i$ )	Frecuencia absoluta ( $f_i$ )
[16-19)		
[19-22)		
[22-25)		
[25-28)		
[28-31)		
[31-34)		
Total		

Después de esto debes proceder a completar la tabla calculando la frecuencia acumulada ( $F_i$ ), relativa ( $n_i$ ) y porcentual (%).

Para calcular la frecuencia acumulada en la primera fila se coloca el mismo valor de la frecuencia absoluta de la primera marca de clase [14-16). Luego debes sumar el primer valor de la frecuencia absoluta que es 12 con el segundo que sería 23.

quedaría  $12+23=35$  y así sucesivamente hasta llegar al último intervalo cuyo resultado debe ser igual al total.

Para calcular la frecuencia relativa, se divide la frecuencia absoluta entre el total.

$$\frac{12}{110} = 0,11 \quad \frac{23}{110} = 0,21 \quad \frac{35}{110} = 0,32 \quad \frac{13}{110} = 0,12$$

Por último, para la frecuencia porcentual multiplicamos la frecuencia relativa por 100.

$$0,11 \times 100\% = 11\% \quad 0,21 \times 100\% = 21\%$$

$$0,32 \times 100\% = 32\%$$

Esta tabla representa la información del barrio 1:

Edades	Marca de clase ( $x_i$ )	Frecuencia absoluta ( $f_i$ )	Frecuencia acumulada ( $F_i$ )	Frecuencia relativa ( $n_i$ )	Frecuencia porcentual %
[14-16)	15	12	12	0,11	11%
[16-18)	17	23	35	0,21	21%
[18-20)	19	35	70	0,32	32%
[20-22)	21	13	83	0,12	12%
[22-24)	23	18	101	0,16	16%

[24-26)	25	7	108	0,06	6%
[26-28)	27	2	110	0,02	2%
Total		<b>110</b>		<b>1</b>	<b>100%</b>

Ahora completa la siguiente tabla con la información del barrio 2:

Edades	Marca de clase ( $x_i$ )	Frecuencia absoluta ( $f_i$ )	Frecuencia acumulada ( $F_i$ )	Frecuencia relativa ( $n_i$ )	Frecuencia porcentual %
[16-19)					
[19-22)					
[22-25)					
[25-28)					
[28-31)					
[31-34)					
Total					

Ahora vamos a calcular el promedio de la edad de mujeres embarazadas en el barrio 1. Para eso vamos a multiplicar la marca de clase con la frecuencia absoluta de cada intervalo así:  $(15 \times 12)$ ,  $(17 \times 23)$ ,  $(19 \times 35)$ ,  $(21 \times 13)$ ,  $(23 \times 18)$ ,  $(25 \times 7)$  y  $(27 \times 2)$ . Luego esos resultados los sumamos así:  $180 + 391 + 665 + 273 + 414 + 175 + 54 = 2152$ . Y por último ese resultado de la suma lo dividimos entre el total de los datos que es 110.

$\bar{x} = \frac{2152}{110} = 19,56$ . Podemos concluir que el promedio de edad en que las mujeres tienen a su primer hijo en el barrio 1 es de cerca de 19 años.



Responde:

- Calcula el promedio para el barrio 2.
- Cual es la diferencia de edades en que salen embarazadas las mujeres del barrio 1 con respecto a las del barrio 2.
- Menciona 3 conclusiones para la situación anterior.

#### Mini-explicación: Medidas de tendencia central para datos agrupados

Las medidas de tendencia central nos ayudan a

**Mediana ( $Me$ ) y clase mediana:** Es el valor de

comparar dos o más poblaciones dada una variable estadística en común. En el ejemplo de a continuación veremos los datos para el barrio 1, en especial el promedio de las edades de las mujeres y luego sacaremos conclusiones de los datos que obtuviste al calcular el promedio del barrio 2, y si esto influye o no en la problemática tratada de los embarazos a temprana edad.

### **Media aritmética para datos agrupados( $\bar{X}$ ):**

para calcularla se determina el cociente de la suma de los productos de la marca de clase ( $x_i$ ) con su correspondiente frecuencia ( $f_i$ ), dividido entre el total de los datos, N.

Tomando como ejemplo el ejercicio anterior sobre los embarazos, voy a calcular la media para el barrio 1.

$$\bar{X} = \frac{(15 \times 12) + (17 \times 23) + (19 \times 35) + (21 \times 13) + (23 \times 18)}{110}$$

Esto quiere decir que el promedio de edad en que las mujeres del barrio 1 tiene a su primer hijo es de 19 años.

**Moda( $M_o$ ) y clase modal:** La moda es el dato con mayor frecuencia absoluta. Cuando los datos están agrupados en clases, se toma como valor aproximado, la marca de la clase modal.

En el ejemplo anterior la clase o el intervalo que tiene mayor frecuencia absoluta es [18-20) con 35 mujeres. Lo que quiere decir que en el barrio 1 la mayoría de mujeres tienen su primer hijo entre los 18 y los 20 años.

la variable, de tal forma que el número de valores menores que él es igual al número de valores mayores que él. Esta depende del orden de los datos y no de su valor.

Para calcular la mediana tomando como referencia el ejemplo anterior, tenemos en cuenta la columna de la frecuencia acumulada ( $F_i$ ) y se calcula la mitad de los datos así:  $\frac{110}{2} = 55$

En la tabla observamos cual es el intervalo cuya frecuencia acumulada es mayor que 55.

Edades	Marca de clase ( $x_i$ )	Frecuencia absoluta ( $f_i$ )	Frecuencia acumulada ( $F_i$ )
[18-20)	19	35	70

una vez identificada la clase mediana, calculamos su valor aproximado de la siguiente forma:

$$Me = \frac{18 + 20}{2} = \frac{38}{2} = 19$$

Por lo tanto la mediana o el valor central es 19 años.

## Paso 1: Redes sociales



Hemos preguntado a un grupo de 40 estudiantes de noveno grado de fe y alegría el tiempo que pasan al día navegando en las redes

sociales, las respuestas fueron dadas en horas como se muestra en la tabla 1. Se les realizó la misma encuesta a 40 estudiantes de décimo grado y las respuestas se muestran en la tabla 2.

**Tabla 1.**

3	2	6	0	1	2	1	5
4	8	3	3	4	2	5	1
5	2	4	3	3	2	3	1
5	6	7	1	2	2	3	1
2	3	4	8	5	3	5	2

**Tabla 2.**

4	7	2	7	9	5	3	5
4	6	5	7	5	4	3	3
7	5	4	2	5	3	4	2
4	4	3	1	6	3	6	1
2	3	5	0	8	4	3	3

Ahora vamos a organizar la información de la tabla 1 correspondiente al noveno grado en una tabla de

frecuencia y agrupamos los datos en 5 intervalos. Y hacemos lo mismo con la información del grado décimo.

**Tabla de frecuencias (noveno grado)**

HORAS	Marca de clase $X_i$	Frecuencia absoluta $f_i$	Frecuencia acumulada $F_i$
[0-2)	1	7	7

[2-4)	3	18	25
[4-6)	5	10	35
[6-8)	7	3	38
[8-10)	9	2	40
TOTAL		40	

En la tabla anterior se puede observar que el intervalo con mayor frecuencia absoluta es el [2-4) el cual corresponde a la clase modal con 18 estudiantes. Eso quiere decir que la mayoría de los estudiantes de noveno grado dedican entre 2 y 3 horas a navegar en las redes sociales.

Para calcular el promedio o la media multiplicamos la marca de clase con respectiva frecuencia y la suma de todas lo dividimos entre el total de los datos que en este caso es 40 estudiantes.

$$\bar{x} = \frac{(1 \times 7) + (3 \times 18) + (5 \times 10) + (7 \times 3) + (9 \times 2)}{40}$$

$$\bar{x} = \frac{7 + 54 + 50 + 21 + 18}{40} = \frac{150}{40} = 3,75$$

Aproximando el valor a un número entero tenemos que el promedio de tiempo que navegan los estudiantes de noveno grado en las redes sociales al día es de 4 horas.

**Tabla de frecuencias (décimo grado)**

HORAS	Marca de clase $X_i$	Frecuencia absoluta $f_i$	Frecuencia acumulada $F_i$
-------	----------------------	---------------------------	----------------------------

[0-2)	1	3	3
[2-4)	3	13	16
[4-6)	5	15	31
[6-8)	7	7	38
[8-10)	9	2	40
TOTAL		40	

En la tabla anterior se puede observar que el intervalo con mayor frecuencia absoluta es el [4-6) el cual corresponde a la clase modal con 15 estudiantes. Eso quiere decir que la mayoría de los estudiantes de décimo grado dedican entre 4 y 5 horas a navegar en las redes sociales.

Para calcular el promedio o la media multiplicamos la marca de clase con su respectiva frecuencia y la suma de todas lo dividimos entre el total de los datos que en este caso es 40 estudiantes.

$$\bar{x} = \frac{(1 \times 3) + (3 \times 13) + (5 \times 15) + (7 \times 7) + (9 \times 2)}{40}$$

$$\bar{x} = \frac{3 + 39 + 75 + 49 + 18}{40} = \frac{184}{40} = 4,6$$

Aproximando el valor a un número entero tenemos que el promedio de tiempo que navegan los estudiantes de décimo grado en las redes sociales al día es de 5 horas.

Concluimos que la diferencia del tiempo que dedican los estudiantes de noveno grado a navegar en las redes sociales en el día con respecto a décimo grado es de una hora.

## Paso 2: Completa este ejemplo: Consumo de metros cúbicos de agua

Se les pidió a 25 estudiantes de Fe y Alegría averiguar cuántos metros cúbicos de agua habían consumido en su casa en el mes de enero y febrero y estas fueron las respuestas.

ENERO					FEBRERO				
14	15	18	21	16	10	8	12	9	15
13	13	15	14	14	12	15	14	13	9
15	18	16	17	15	10	10	9	12	14
12	15	18	17	16	15	12	14	13	10
18	15	14	16	16	9	8	12	14	12

Con base a la información anterior nos piden que calculemos el promedio de metros cúbicos de agua que se consumieron en los meses de enero y febrero en las casas de los 25 estudiantes. Nos piden que hagamos



una tabla de frecuencias con 5 intervalos o clases para el mes de enero y 4 intervalos para el mes de febrero.

Las tablas que se presentan a continuación se encuentran incompletas, ayúdanos con la información que hace falta en los espacios que están sombreados.

CONSUMO ENERO					
MINUTOS	MARCA DE CLASE	F. ABSOLUTA	F. ACUMULADA	F. RELATIVA	F. PORCENTUAL
[12-14)	13			0,12	
[14-16)		10			40%
[16-18)	17		20	0,28	
[18-20)		4			
[20-22)	21				4%
TOTAL					

CONSUMO FEBRERO					
MINUTOS	MARCA DE CLASE	F. ABSOLUTA	F. ACUMULADA	F. RELATIVA	F. PORCENTUAL
[8-10)	9			0,24	
[10-12)		4			16%
[12-14)	13		18	0,32	
[14-16)		7			
Total					100%



**Responde:**

- ¿Cuál es el promedio de consumo de agua en los meses de enero y febrero?
- ¿Cuál es la diferencia de dichos promedios?
- ¿Qué puedes concluir de esta información?

---

**Paso 3: 1-2-4: Tu turno (individual, en parejas y en grupos de 4)**



De manera individual piensa en un conjunto de datos numéricos en donde sea posible construir una tabla de frecuencias completa y calcular las medidas de tendencia central.



Comparte tu problema con un compañero. Tu pareja deberá realizar la tabla de frecuencia y calcular las medidas de tendencia central de tu problema y tu la tabla de su problema.



Ahora reúnanse con otra pareja y compartan sus tablas elaboradas. Dialoguen sobre lo que hicieron y el proceso que llevaron dándose retroalimentación entre ustedes.

Finalmente busquen a su profesor o profesora para dialogar y compartir lo que hicieron y aclarar los conceptos.

## C) Resuelve y practica

- 1) Una tienda en línea registra el tiempo que tarda la empresa de correos en hacer llegar su mercadería a los clientes. los tiempos en días registrados son los siguientes.

2 7 10 16 19

22 6 25 5 20

13 32 13 29 18

20 13 6 12 35

- Con los datos obtenidos, calcula el rango, la amplitud y elabora una tabla de frecuencias con 5 intervalos.
- Calcula las medidas de tendencia central.

- 2) La tienda decide contratar una nueva empresa de correos para entregar su mercancía y toma el tiempo en días que demora en entregarlos y los registra en la siguiente tabla.

12	8	7	10	13
16	15	12	9	14
8	10	13	8	12
15	9	16	15	11

- Con los datos obtenidos, calcula el rango, la amplitud y elabora una tabla de frecuencias con 5 intervalos.
- Calcula las medidas de tendencia central.
- Realiza una comparación de los datos obtenidos en los puntos 1 y 2 y concluye

cuál empresa de correos le es más conveniente para la tienda.

- 2) Una empresa que realiza baterías recargables, desea saber el rendimiento de su producto y para ello separaron una muestra de 100 baterías las cuales fueron probadas en juguetes para saber cuantas horas funcionaba antes de necesitar ser recargada nuevamente. En la siguiente tabla se muestran los datos ordenados de manera agrupada.



Tiempo de duración (horas)	Frecuencia absoluta (fi)	Frecuencia relativa (hi)	Marca de clase (Xi)	Frecuencia absoluta por marca de clase (Xi*fi)	Frecuencia absoluta acumulada (Fi)
11-15	9	0,09	13	117	9
16-20	20	0,2	18	360	29
21-25	13	0,13	23	299	42
26-30	10	0,1	28	280	52
31-35	13	0,13	33	429	65
36-40	18	0,18	38	684	83
41-45	9	0,09	43	387	92
46-50	8	0,08	48	384	100
<b>TOTAL</b>	<b>100</b>	<b>1</b>		<b>2940</b>	

Con base a la información anterior.

- Calcula el promedio de horas que dura la batería antes de volver a cargarla?
  - ¿Qué porcentaje de baterías duran entre 31 y 35 horas?
  - ¿Qué porcentaje de baterías duran entre 11 y 30 horas?
- 3) El profesor de educación física llevó al salón de clases una báscula, para saber el peso en kilogramos de 20 estudiantes de su clase. A

continuación se presentan los resultados obtenidos.

[58-62)			
Total			

56	61	53	62	56	46	49	52	61	62
58	59	58	57	51	49	52	53	61	59

Complete la siguiente tabla de frecuencias con los intervalos propuestos. Recuerde que la marca de clase es el punto medio de cada intervalo y en la última columna debe multiplicar la marca de clase con la frecuencia absoluta.

Peso (Kg)	Marca de clase ( $x_i$ )	Frecuencia ( $f_i$ )	$x_i * f_i$
[46-50)			
[50-54)			
[54-58)			

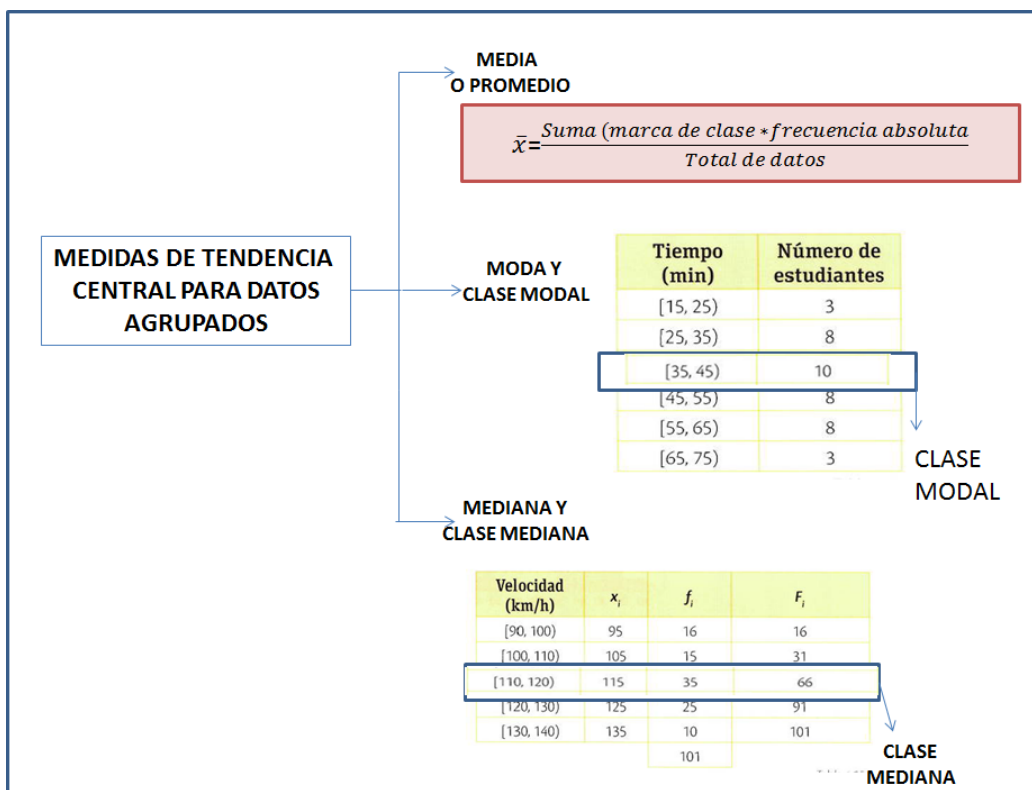
- ¿Entre que pesos se encuentra la mayoría de los estudiantes?
- ¿Qué porcentaje de estudiantes está entre 54 y 58 kilogramos?
- Calcula el promedio del peso de los estudiantes.

### PROBLEMA DE KHAN ACADEMY

TEMA: Introducción a la estadística: moda, media y mediana.

<https://es.khanacademy.org/math/ap-statistics/summarizing-quantitative-data-ap/measuring-center-quantitative/v/statistics-intro-mean-median-and-mode>

## D) Resumen



### Ten en cuenta que:

- Estamos trabajando únicamente variables cuantitativas, quiere decir datos numéricos.
- Es importante definir las poblaciones con las que se quiere trabajar para obtener mejores resultados y ver las comparaciones.
- Las tablas de frecuencias me permitan organizar de manera rápida y fácil la información para luego sacar conclusiones.
- Para calcular la moda y la mediana en datos agrupados es fundamental identificar primero la clase a la que pertenecen.

## E) Valoración

### i) Califica tu comprensión por tema en tu cuaderno

Evidencias	●○○ Todavía no entiendo los conceptos	●●○ Voy bien pero quiero más práctica	●●● Comprendí muy bien el tema
1. Realizo comparaciones entre dos o más poblaciones dada una variable estadística.			
2. Diferencio las medidas de tendencia central y las aplico en un estudio estadístico para datos agrupados.			
3. Comparo las medidas de tendencia central en datos agrupados para dos o más conjuntos de datos.			

### ii) Preguntas de comprensión

1) Para realizar un estudio estadístico, ¿Es necesario trabajar con toda la población?

V( ) F( )

2) ¿Las variables cuantitativas son las que arrojan únicamente valores numéricos?

V( ) F( )

3) El promedio o media aritmética se puede calcular para variables cualitativas.

V( ) F( )

4) Menciona 3 variables cuantitativas.

---



---



---

5) Menciona 3 variables cualitativas.

---



---



---

### iii) Resuelvo un problema

Andrés y Juan son estudiantes de noveno grado y han estado dialogando acerca de sus notas. Andrés dice que tiene mejores notas que Juan y viceversa. Ayudemos a los chicos a calcular el promedio de sus notas para ver a quién le va mejor.

NOTAS DE JUAN					NOTAS DE ANDRÉS				
9	8	9	6	10	7	10	8	9	10

a) Calcula el promedio de las notas de ambos estudiantes.

b) Menciona 2 conclusiones para lo anterior.



## ACTIVIDAD 2: APLIQUEMOS LO APRENDIDO!

Aprendamos a hacer conjeturas y plantear hipótesis para una población de gran tamaño partiendo de una muestra tomada de ella.

### A) Recordemos: Técnicas de recolección de datos

#### RECUERDA QUE...

Las principales técnicas de recolección de datos son:

**Encuesta:** Con esta técnica de recolección de datos da lugar a establecer contacto con las unidades de observación por medio de los cuestionarios previamente establecidos.

**Entrevista:** La entrevista es una situación de interrelación o diálogo entre personas, el entrevistador y el entrevistado.

**Análisis documental:** se recolectan datos de fuentes secundarias. Libros, boletines, revistas, folletos, y periódicos se utilizan como fuentes para recolectar datos sobre las variables de interés.

#### Práctica

i) Escoge una de las técnicas anteriores y recoge 40 datos de tu interés. Puedes escribirlos en el recuadro siguiente.


ii) Identifica la población, muestra, variable de estudio y el tipo de variable a la que pertenece.  
iii) Responde las siguientes preguntas y Justifica:

- ¿Si se escogiera un nuevo dato, como seria con relación a los 40 de la tabla?
- ¿Si se realizara el ejercicio nuevamente y se escogieran 40 datos nuevos, estos serían totalmente diferentes, o habrían similitudes?
- ¿Se puede dar conclusiones generales para toda la población, solo con la información de la tabla o es necesario recoger los datos de toda la población?

iv) Plantea 3 conclusiones para la población, tomando como referencia los resultados obtenidos.

**Verifica las respuestas de la sección A con tu profesor.**

## B) Conceptos: Variables bidimensionales

### Exploración: Prueba Saber vs Estrato socio-económico



Realizas una investigación para conocer como influye el estrato socioeconómico en los resultados de las pruebas Saber Noveno. Le preguntas a 50 estudiantes de grado 10° por los resultados obtenidos en el año anterior en las pruebas Saber Noveno, los cuales están en un rango de  $[1, 100]$  y les preguntas por su estrato Socio-económico.

Los resultados los organizas en pares ordenados (puntajes, Estrato) esto con el fin de mantener las dos respuestas de un estudiante juntas. por ejemplo, Andrés es un estudiantes de estrato 3, que obtuvo un puntaje de 45, por

lo tanto sus respuestas estarán representadas como (45,3).

Los resultados que obtuviste son los siguientes:

(62,4) (42,2) (12,1) (73,4) (31,2) (67,5) (95,6) (55,3) (74,4) (68,3) (93,7) (39,1) (25,2) (23,2)  
(56,3) (31,2) (42,1) (83,5) (94,6) (73,5) (84,4) (73,4) (67,1) (85,3) (43,2) (89,4) (83,4) (91,6)  
(45,3) (85,5) (76,5) (67,3) (49,2) (95,6) (85,6) (57,3) (94,7) (95,7) (83,7) (28,1) (29,2) (53,2)  
(75,3) (83,5) (54,4) (62,4) (98,7) (35,1) (42,3) (51,4)

Para organizar la información y poder dar conclusiones, creas una tabla que te relaciona los puntaje y el estrato, como ya tienes conocimiento de datos agrupados decides colocar los estratos en intervalos de rango  $[1,10]$ .

	1	2	3	4	5	6	7
[1,10]							
[11,20]	I						
[21,30]	I	III					
[31,40]	II	II					
[41,50]	I	III	II				

[51,60]		I	III	II			
[61,70]	I		II	II	I		
[71,80]			I	III	II		
[81,90]			I	III	III	I	I
[91,100]						IIII	IIII

De la tabla anterior se puede agregar tanto la suma de cada estrato como de cada intervalo de puntaje. Con lo que se tiene la tabla de doble entrada

	1	2	3	4	5	6	7	Total
[1,10]	0	0	0	0	0	0	0	0
[11,20]	1	0	0	0	0	0	0	1
[21,30]	1	3	0	0	0	0	0	4
[31,40]	2	2	0	0	0	0	0	4
[41,50]	1	3	2	0	0	0	0	6
[51,60]	0	1	3	2	0	0	0	6
[61,70]	1	0	2	2	1	0	0	6
[71,80]	0	0	1	3	2	0	0	6
[81,90]	0	0	1	3	3	1	1	7
[91,100]	0	0	0	0	0	4	4	8
Total	6	9	9	10	6	5	5	50



### Responde

- ¿Cuántos estudiantes de estrato 6 obtuvieron un puntaje entre 91 y 100?
- ¿Cuántos estudiantes de estrato 3 obtuvieron un puntaje de 70 o más?
- ¿Los puntajes más altos los obtuvieron los estudiantes de que estrato?
- ¿Dónde se ubica el puntaje de la mayoría de los estudiantes?
- ¿Cómo influye el estrato socioeconómico en los resultados de las pruebas Saber Noveno?
- Considera los pares ordenados obtenidos al inicio como puntos y ubícalos en el plano cartesiano.

### Mini-Explicación: VARIABLES ESTADÍSTICAS BIDIMENSIONALES

Se llama variable estadística bidimensional al estudio de dos variables estadísticas unidimensionales en una misma población. Así, tendremos que recoger de cada individuo dos datos. Estos datos correspondientes a un mismo individuo no se pueden separar, pues vamos a estudiar la relación que existe entre ellos. Los podremos recoger formando pares de elementos  $(x,y)$ . Así,  $x$  es unidimensional, y también  $y$  y  $(x,y)$  es el dato bidimensional.

**TABLA DE DOBLE ENTRADA:** Está formada por tantas filas y columnas como valores tengamos de cada una de las variables, más una fila y una columna más para indicar los totales. Está indicada para casos con bastantes datos, en los que para cada valor de una variable, existen varios valores de la otra.

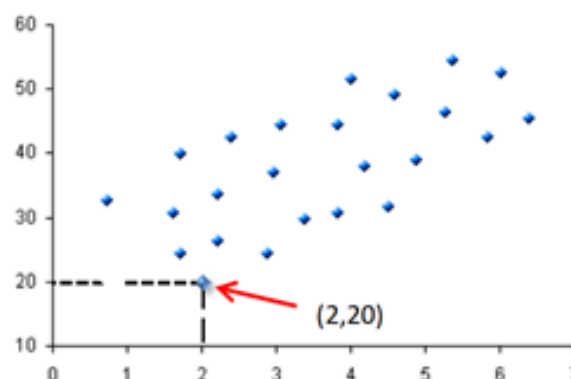
En la siguiente tabla, se organizan los datos de una encuesta. Se puede observar que 7 personas escogieron lápiz y cuaderno. Para los totales se debe sumar las celdas de cada columna y cada fila, al final en la esquina inferior izquierda, va el total de encuestados (20). Al sumar las celdas rojas debe dar 20 al igual que sumar las celdas amarillas.

	Cuaderno	Block	Total
Lápiz	7	6	$7+6= 13$
Lapicero	5	2	$5+2= 7$
Total	$7+5= 12$	$6+2= 8$	20

### DIAGRAMA DE DISPERSIÓN

Consiste en dos ejes perpendiculares, en cada uno de ellos colocaremos los valores de cada una de las variables. Para representar el dato correspondiente al par  $(x,y)$ , colocaremos un punto en las mismas coordenadas.

Para hacer un diagrama de dispersión, se dibuja una L, y se ubican los datos en eje X y en el eje Y, para enumerar se debe tener en cuenta los números de cada dato. Para graficar el dato por ejemplo el  $(2,20)$ , se busca el punto donde se encuentren el 2 del eje horizontal, con el 20 del eje vertical. Tal como se señala en la imagen.





### Paso 1: Ejemplo: Estudiar minutos extras

Hemos tomado 30 estudiantes de Fe y Alegría, después de la entrega de informes y se les registraron:  
 $x$  = Minutos extras que toma para estudiar, y  
 $y$  = # de ambientes en los que no alcanzó la meta deseada.

Partiendo de la información obtenida se organizaron los resultados en la siguiente tabla de doble entrada.

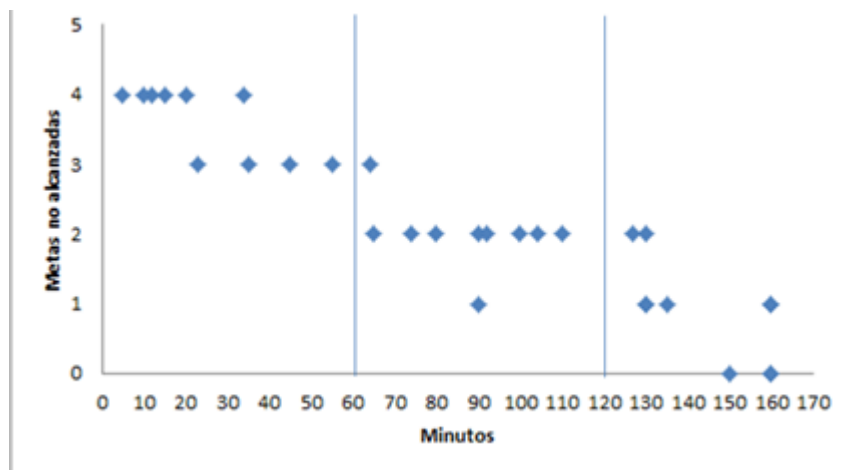
La columna amarilla corresponde a las metas no alcanzadas por los estudiantes y la fila azul los minutos extras utilizados para estudiar.

	[0,59]	[60,119]	[120,179]	TOTAL
0	0	0	2	2
1	0	1	3	4
2	0	8	2	10
3	4	1	0	5
4	6	0	0	6
TOTAL	10	10	7	27

Si queremos saber por ejemplo cuántos minutos extras utilizaron los estudiantes que no alcanzaron 3 metas, podemos observar que de 5 estudiantes, 4 estudian máximo 1 hora extra y un estudiante máximo 2 horas.

Con la tabla de doble entrada podemos tener una idea de la relación de estos datos, pero el

diagrama de dispersión deja en evidencia el comportamiento.



Podemos concluir del diagrama que los estudiantes que implementan menos tiempo para estudiar tienen menos probabilidad de alcanzar las metas de los ambientes o que a medida que los estudiantes implementaron más tiempo de estudio se redujeron las metas no alcanzadas.

### Paso 2: Completa este ejemplo: La marca favorita



Se realiza una encuesta a 100 niños sobre su zapato de preferencia.


Para esto les preguntó color y marca, se obtuvieron los siguientes resultados:


	Nike	Adidas	Puma	Total
Negro	22	?	8	42
Rojos	5	10	6	?
Blancos	?	1	?	14
Azul	10	4	9	?
Total	45	?	28	?


Se sacaron las siguientes conclusiones

1. La marca favorita de los niños encuestados es \_\_\_\_\_ con 45 votos.
2. Los zapatos \_\_\_\_\_ de color \_\_\_\_\_ son los menos escogido por los niños, con solo 1 voto.
3. Los zapatos Nike Azul tuvieron la misma cantidad de voto que los zapatos \_\_\_\_\_, ambos con \_\_\_\_\_ votos.

**Paso 3: 1-2-4 tu turno (individual, en parejas y en grupo de 4)**

 Crea un problema con un conjunto de 40 datos (Variables bidimensionales), escríbelo en tu cuaderno y recuerda escribir cada dato de la forma (x,y).

 Comparte tu problema con un compañero. Tu pareja deberá realizar la tabla de doble entrada de tu problema y tu la tabla de su problema.

 Júntense con otra pareja y compartan sus tablas. Cada pareja debe realizar el diagrama de dispersión de 1 de las tablas de la otra pareja.  
Finalmente, busquen a su profesor para dialogar y compartir sus creaciones, aclarando los conceptos.

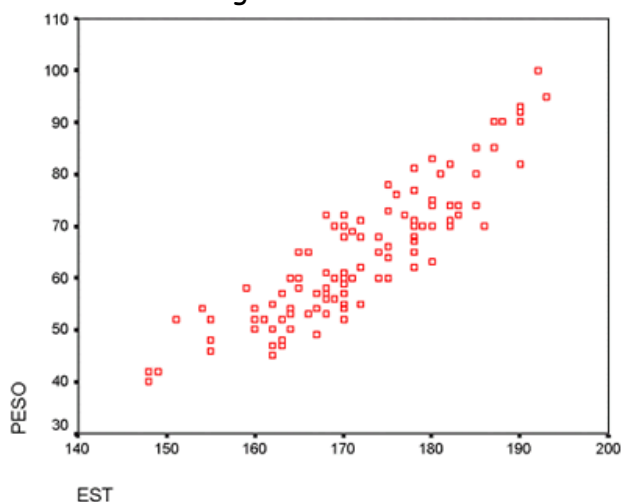


## C) Resuelve y practica

- 1) Gracias a diversos estudios, se ha establecido la siguiente tabla que permite conocer si un hombre se encuentra en sobrepeso o con falta de peso, según su estatura.

Estatura	Peso mínimo	Peso ideal	Peso máximo
150 cm	43 kg	47 kg	54 kg
153 cm	45 kg	49 kg	56 kg
156 cm	47 kg	51 kg	58 kg
159 cm	49 kg	53 kg	60 kg
162 cm	51 kg	55 kg	62 kg
165 cm	53 kg	57 kg	64 kg
168 cm	55 kg	59 kg	66 kg
171 cm	57 kg	61 kg	68 kg
174 cm	59 kg	63 kg	70 kg
177 cm	61 kg	65 kg	72 kg
180 cm	63 kg	67 kg	74 kg
183 cm	65 kg	69 kg	76 kg
186 cm	67 kg	71 kg	78 kg
189 cm	69 kg	73 kg	80 kg

Se realizó un estudio donde se le preguntó a un grupo de hombres su estatura y peso, se Obtuvieron los siguientes resultados:



En base al diagrama anterior responde:

- ¿Entre los hombres que pesan menos de 160cm, alguno tiene Sobrepeso? ¿Por qué?
- ¿Cuánto pesa la persona más alta encuestada?
- ¿Los hombres que miden 190 cm o más, tienen sobrepeso? ¿por qué?
- ¿En qué intervalo de edades se encuentran la mayoría de los hombres encuestados?
- Ubica tu pareja ordenada (estatura, peso) en el diagrama de dispersión y observa si te encuentras en el peso ideal.
- fórmula 3 conclusiones de estatura vs peso.

2) Realiza una encuesta a 40 de tus compañeros del colegio, donde les preguntes la estatura (cm) y la edad (años), y en tu cuaderno realiza lo siguiente:

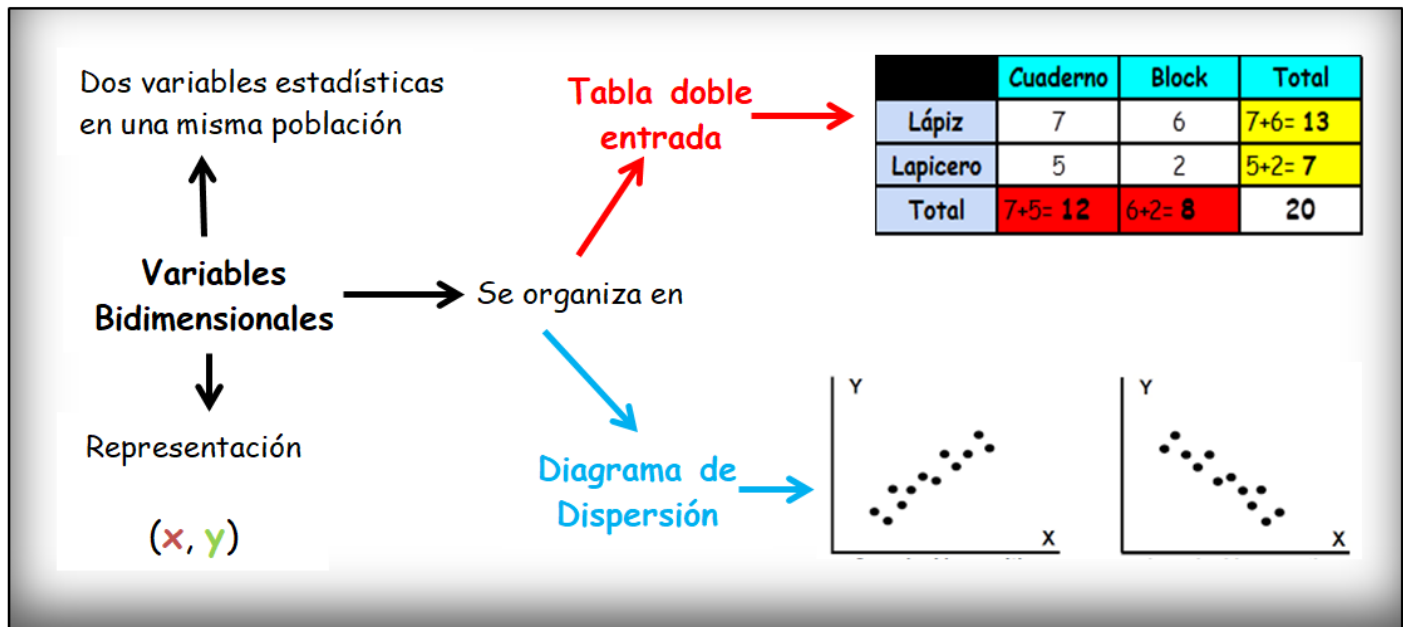
- Escribe los datos de la forma (edad, estatura).
- Tabla doble entrada
- Diagrama de dispersión.
- Fórmula 3 conclusiones.

### PROBLEMAS DE KHAN ACADEMY

Tema: Construir un diagrama de dispersión  
(Mira los videos y responde las preguntas)

<https://es.khanacademy.org/math/cc-eighth-grade-math/cc-8th-data/cc-8th-scatter-plots/v/constructing-scatter-plot>

## D) Resumen



Ten en cuenta que:

- Para un estudio estadístico de variables Bidimensionales se debe recoger de cada individuo dos datos.
- La tabla doble entrada se utiliza cuando se recogen muchos datos.
- Para graficar en un diagrama de dispersión se debe tener claro a que eje le corresponde cada variable.
- El diagrama de dispersión es importante para determinar qué relación hay entre las variables.

## E) Valoración

### i) Califica tu comprensión por tema en tu cuaderno

Evidencias	●○○ Todavía no entiendo los conceptos	●●○ Voy bien pero quiero más práctica	●●● Comprendí muy bien el tema
Realizo estudios estadísticos con variables Bidimensionales utilizando cualquier método de recolección de datos.			
Organizo y represento las variables bidimensionales en tablas de doble entrada y diagrama de dispersión.			
Formulo conclusiones e hipótesis para una población a partir de información dada de una muestra.			

### ii) Preguntas de comprensión

Tabla. Tipo de pelo de mujer.

	Liso	Rizo	Total
Largo	14	8	22
Corto	9	9	18
Total	23	17	40

1) El total de mujeres con pelo liso es  
[ ] 14 [ ] 23

2) ¿Cuántas mujeres se encuestaron?  
\_\_\_\_\_

3) ¿Cuántas mujeres tienen el pelo largo y con rizos?  
[ ] 17 [ ] 8

4) Las mujeres de cabello corto con rizos son la misma cantidad de mujeres con pelo corto liso.  
[ ] Verdadero  
[ ] Falso

5) ¿Qué conclusión es cierta?  
[ ] Es más probable encontrar una mujer de pelo liso largo.  
[ ] Solo existen 9 mujeres con el pelo corto liso.

### iii) Resuelvo un problema

Se realizó un estudio de las horas que duerme una persona según su edad, para esto se preguntó a 20 personas con distintas edades en el rango de [1,20]. Pero se olvidaron de colocar los siguientes datos.

(3,13) (9,10) (18, 8.5) (20, 7)

- ¿Podrías ubicar los datos en el siguiente diagrama de dispersión?
- Se concluyó que la edad no influye en nada con las horas de sueño. ¿Qué les dirías tú y por qué?

