



# UDS



## Mi Universidad

Se realizó una encuesta a un centro multideportivo en el cual se tomó una muestra de 50 personas y se formuló las siguientes preguntas.

- Su edad
- Que deporte practica
- Cuantos años ha practicado deporte

Gracias a estas preguntas se puede hacer el uso de varias herramientas estadísticas para representar cada apartado de cada respuesta como por ejemplo: Tablas de frecuencias, caja de bigotes, diagramas de venn, probabilidades, entre otros; cada uno tiene una función diferente en el cual se explicará en esta actividad, tratando de ser lo más simple posible para comprender desde los aspectos más básicos, hasta como se realizó.

## VALORES PARA LA TABLA DE FRECUENCIAS EN INTERVALOS

Para hacer este tipo de tabla de frecuencia, se requiere saber el rango, la amplitud y el intervalo del número total de los datos a través de algún tipo de encuesta hecha a una muestra de población del tema de interés.

Rango (R)  
Fórmula

$$R = X_{\max} - X_{\min}$$

Intervalos (k)  
Fórmula

$$K = 1 + 3.322 \times \log(n)$$

Amplitud (A)  
Fórmula

$$A = R/k$$

1. **\*Rango\***: Es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos. Se utiliza para determinar la amplitud total de los datos.
2. **\*Intervalo\*** (o clase): Es un subconjunto de valores dentro del rango que se utiliza para agrupar los datos. Los intervalos pueden ser de igual o diferente amplitud.
3. **\*Amplitud\*** (o ancho de clase): Es la diferencia entre el límite superior y el límite inferior de un intervalo. La amplitud puede ser fija o variable.

VALORES PARA LA TABLA DE FRECUENCIAS EN INTERVALOS  
APLICADO A LA ENCUESTA CON 50 DATOS DE EDAD DE LOS DEPORTISTAS

Rango (R)  
Fórmula

$$R = X_{\max} - X_{\min}$$

Sustitución

$$R = 40 - 17 = 23$$

Resultado

$$R = 23$$

Intervalos (k)  
Fórmula

$$K = 1 + 3.322 \times \log(n)$$

Sustitución

$$K = 1 + 3.322 \times \log(50)$$

Resultado

$$K = 6$$

Amplitud (A)  
Fórmula

$$A = R/k$$

Sustitución

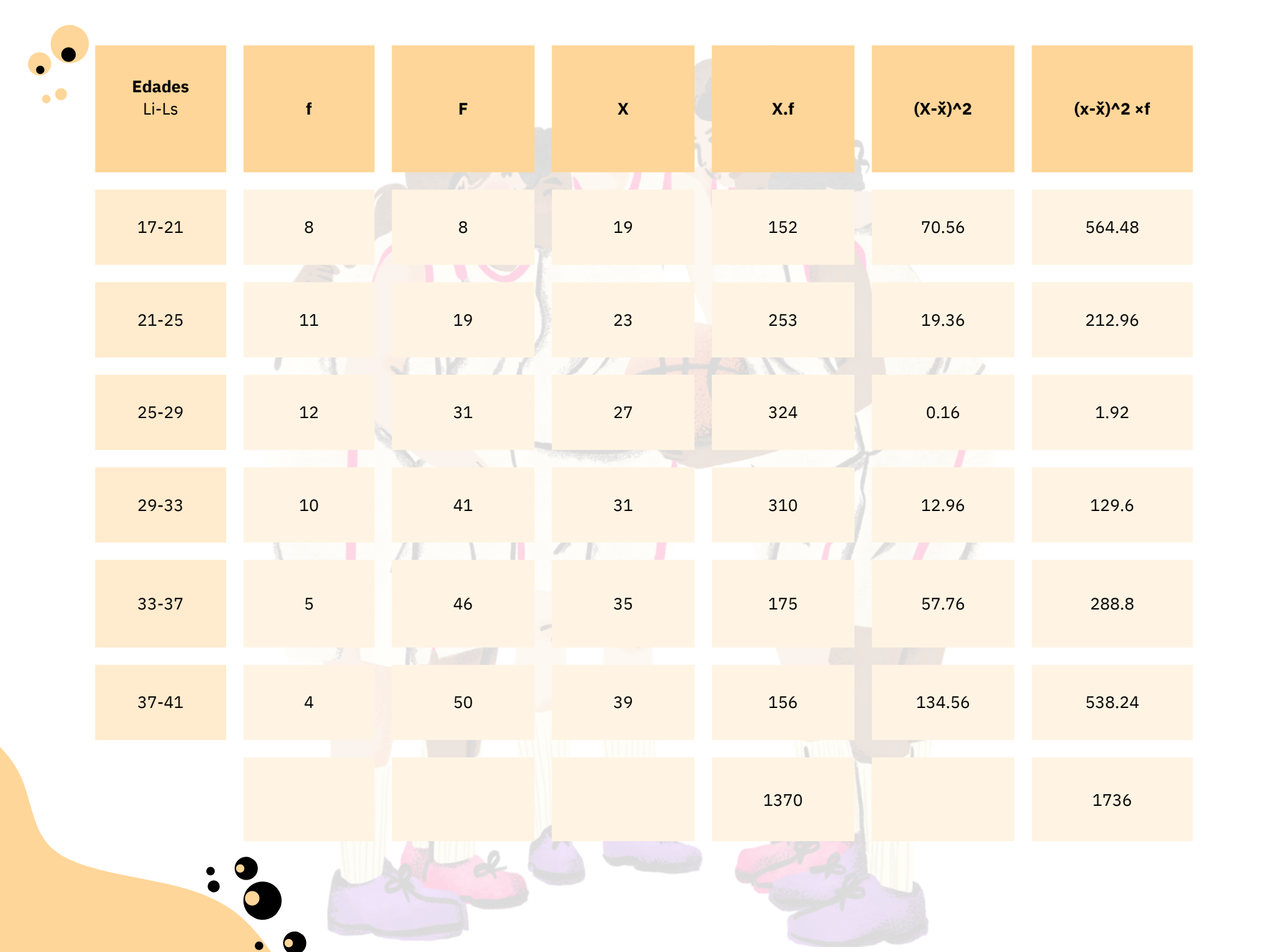
$$A = 23/6$$

Resultado

$$A = 4$$



CON ESTOS DATOS SE SACÓ LA  
TABLA DE FRECUENCIAS EN  
INTERVALOS



Edades Li-Ls	f	F	X	X.f	$(X-\bar{x})^2$	$(x-\bar{x})^2 \times f$
17-21	8	8	19	152	70.56	564.48
21-25	11	19	23	253	19.36	212.96
25-29	12	31	27	324	0.16	1.92
29-33	10	41	31	310	12.96	129.6
33-37	5	46	35	175	57.76	288.8
37-41	4	50	39	156	134.56	538.24
				1370		1736

# MEDIDAS DE TENDENCIA CENTRAL

LAS MEDIDAS DE TENDENCIA CENTRAL SON MEDIDAS ESTADÍSTICAS QUE PRETENDEN RESUMIR EN UN SOLO VALOR A UN CONJUNTO DE VALORES. REPRESENTAN UN CENTRO EN TORNO AL CUAL SE ENCUENTRA UBICADO EL CONJUNTO DE LOS DATOS. LAS MEDIDAS DE TENDENCIA CENTRAL MÁS UTILIZADAS SON: MEDIA, MEDIANA Y MODA.

## Moda Fórmula

$$Li + \frac{fi - fi-1}{(fi - fi-1) + (fi - fi+1)} \times A$$



ES EL VALOR QUE APARECE  
CON MAYOR FRECUENCIA EN  
UN CONJUNTO DE DATOS.

## Promedio Fórmula

$$\bar{X} = \frac{\sum X \times f}{n}$$



ES LA SUMA DE TODOS LOS  
VALORES DIVIDIDA POR EL  
NÚMERO DE VALORES.

## Mediana Fórmula

$$Li + \frac{n/2 - F_{i-1}}{fi} \times A$$



ES EL VALOR CENTRAL DE UN  
CONJUNTO DE DATOS ORDENADOS,  
QUE SEPARA LA MITAD INFERIOR DE  
LA MITAD SUPERIOR.

# MEDIDAS DE TENDENCIA CENTRAL APLICADAS A LA TABLA DE FRECUENCIA DE UNA ENCUESTA CON 50 DATOS DE EDADES DE LOS DEPORTISTAS

Moda

Fórmula

$$Li + \frac{fi - fi-1}{(fi - fi-1) + (fi - fi+1)} \times A$$

Sustitución

$$25 + \frac{12 - 11}{25 + 1/3} \times 4 = 26.33$$

Resultado

$$Mo = 26.33$$

Promedio

Fórmula

$$\bar{X} = \frac{\sum X \times f}{n}$$

Sustitución

$$\bar{X} = 1,370 / 50 = 27.4$$

Resultado

$$\bar{X} = 27.4$$

Mediana

Fórmula

$$Li + \frac{n/2 - F_{l-1}}{fi} \times A$$

Sustitución

$$25 + \frac{25 - 19}{12} \times 4 = 27$$

Resultado

$$Me = 27$$

# VARIANZA Y DESVIACION ESTANDAR APLICADA A LA TABLA DE FRECUENCIA POR INTERVALOS CON DATOS DE UNA MUESTRA DE LA POBLACIÓN

## Varianza Fórmula (Muestra)

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

## Desviación E Fórmula

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

## Sustitución

$$S^2 = \frac{\sum 1736}{50-1}$$

$$S^2 = \frac{1736}{49}$$

$$S^2 = 35.42^2$$

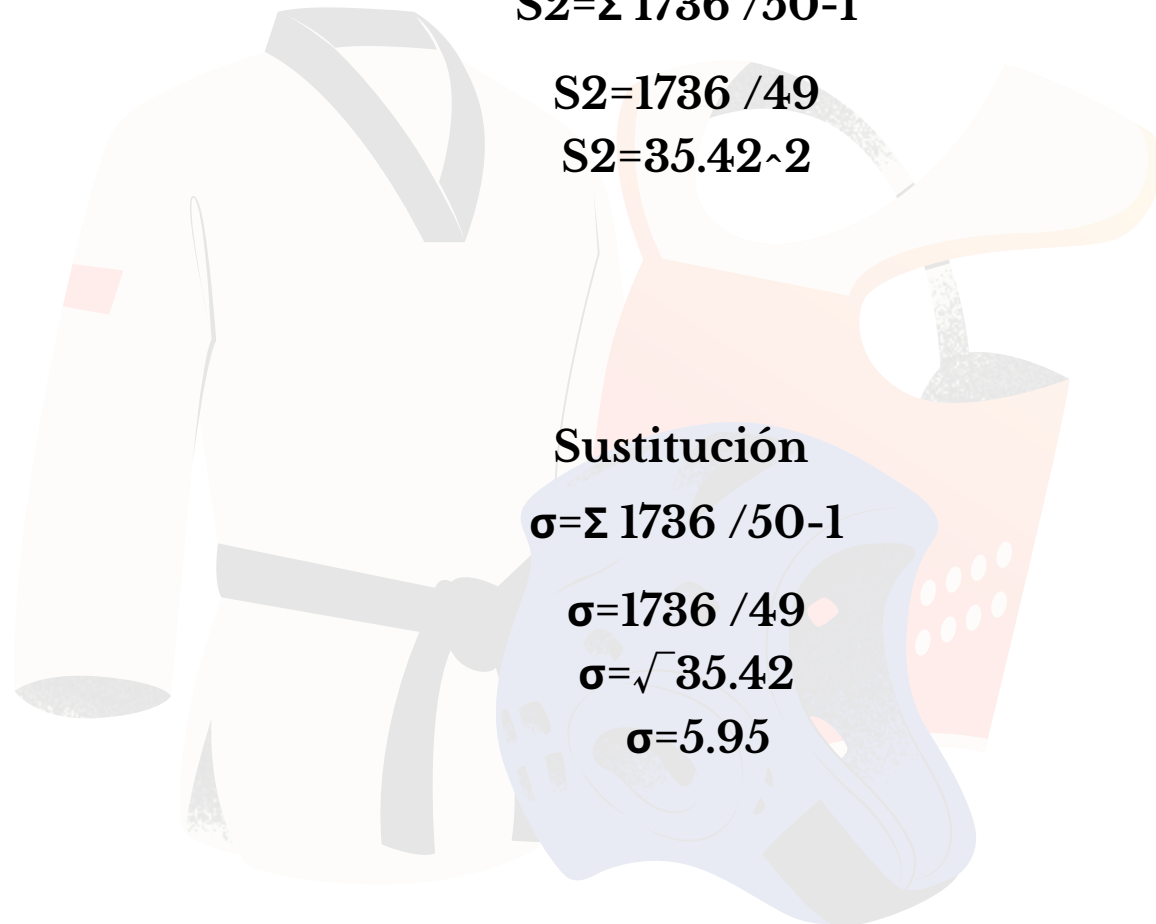
## Sustitución

$$\sigma = \frac{\sum 1736}{50-1}$$

$$\sigma = \frac{1736}{49}$$

$$\sigma = \sqrt{35.42}$$

$$\sigma = 5.95$$



# VARIANZA Y DESVIACION ESTANDAR

La varianza y la desviación estándar son dos medidas estadísticas que sirven para medir la dispersión de conjuntos de datos. La varianza mide la cantidad de variabilidad en un conjunto de datos y se utiliza para comparar dispersión, calcular covarianza y modelar procesos aleatorios. La desviación estándar, por otro lado, es una medida más intuitiva que se expresa en la misma unidad que los datos y se utiliza para establecer límites de confianza, comparar datos e identificar valores atípicos.

Varianza  
Fórmula  
(Población)

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Varianza  
Fórmula  
(Muestra)

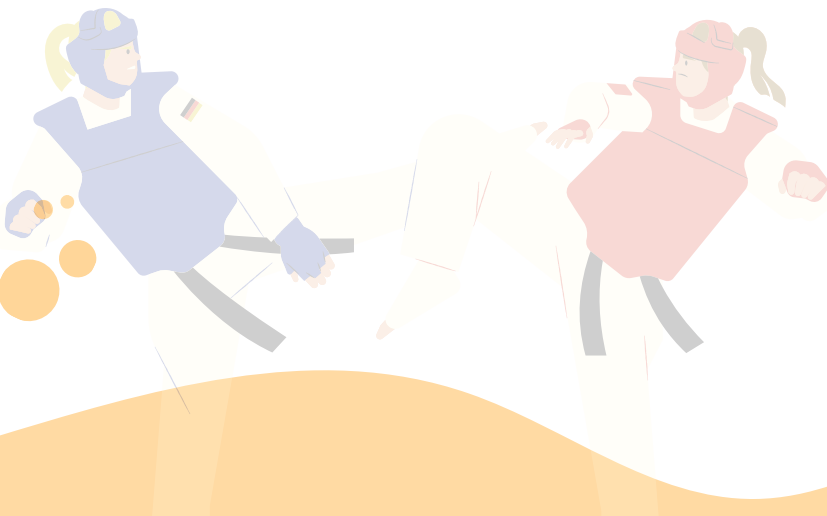
$$S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Desviación E  
Fórmula

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

O

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





# DIAGRAMA CAJA DE BIGOTES

Un diagrama de caja y bigotes es un gráfico que muestra la distribución de datos en cuartiles, la media y los valores atípicos. También se le conoce como box plot en inglés.

Características :

Se utiliza para comparar distribuciones de datos numéricos

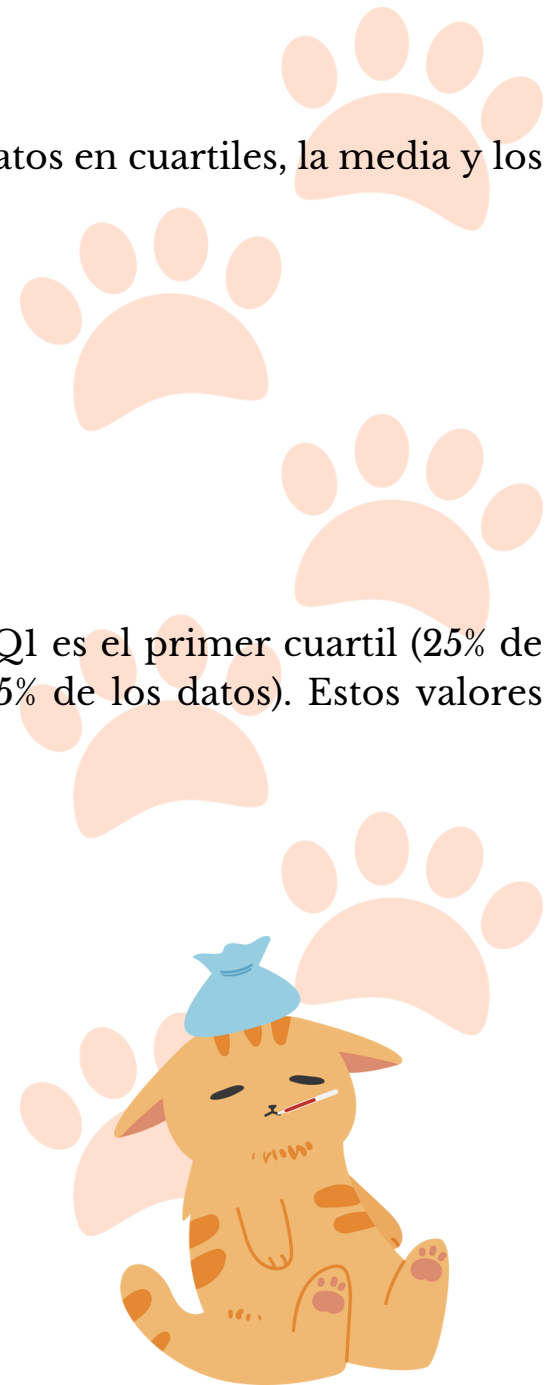
Se basa en las medianas

Muestra la morfología y simetría de los datos

Permite identificar valores atípicos

Muestra el 50% de los valores centrales datos datos

En una caja de bigotes,  $Q_1$ ,  $Q_2$  y  $Q_3$  representan los cuartiles de los datos.  $Q_1$  es el primer cuartil (25% de los datos),  $Q_2$  es la mediana (50% de los datos) y  $Q_3$  es el tercer cuartil (75% de los datos). Estos valores ayudan a visualizar la distribución de los datos y a identificar patrones.



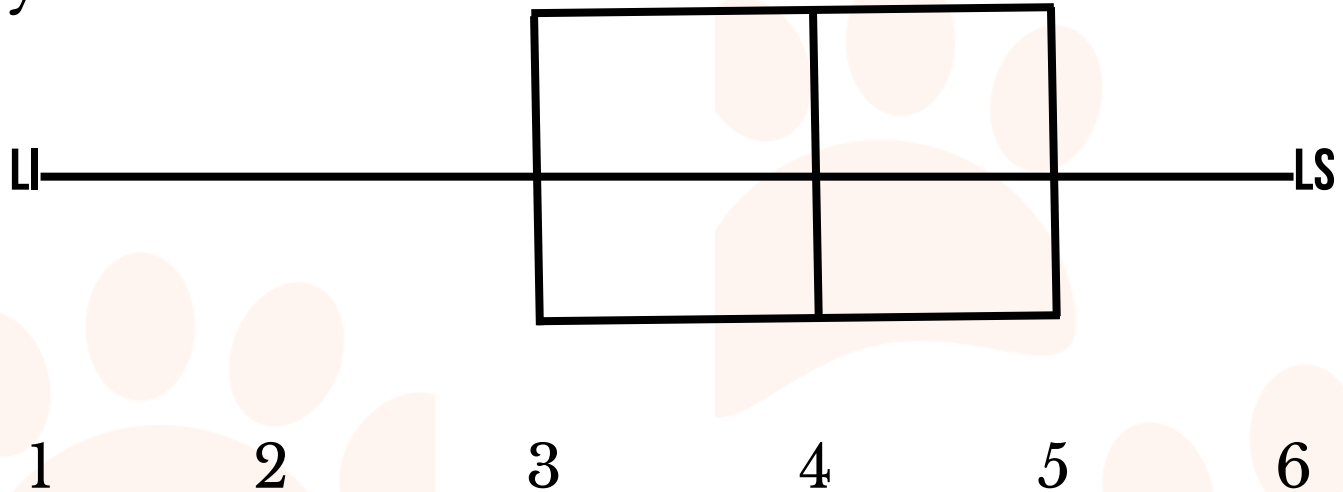
# DIAGRAMA CAJA DE BIGOTES APLICADO A EL TIEMPO EN AÑOS PRACTICANDO UN DEPORTE

DATOS:

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

$Q1=3$     $Q2=4$     $Q3=5$

EJ  
 $Q2 = N/2$   
 $Q2 = 50/2$   
 $Q2 = \text{cae en } 4y4$   
 $Q2 = 4+4/2$   
 $Q2 = 4$



# PROBABILIDAD

A background illustration featuring a stylized sun with rays in the upper right and a tennis net in the lower right, set against a light orange and white background.

## PROBABILIDAD

Probabilidad de la unión de dos sucesos cualesquiera: La probabilidad de la unión de dos sucesos cualesquiera es igual a la probabilidad del primero, más la probabilidad del segundo, menos la probabilidad de la intersección

## Fórmulas

$$A \cup B = P(A) + P(B)$$

Sirve para calcular la probabilidad sin que el otro suceso se repita

$$A \cup B = P(A) + P(B) - P(A \cap B)$$

Sirve para calcular si las dos probabilidades se cumplen

# DIAGRAMAS DE VENN

Un diagrama de Venn es un gráfico que utiliza círculos superpuestos para mostrar las relaciones entre grupos de datos. Son útiles para comparar datos, medir probabilidades y organizar visualmente la información.

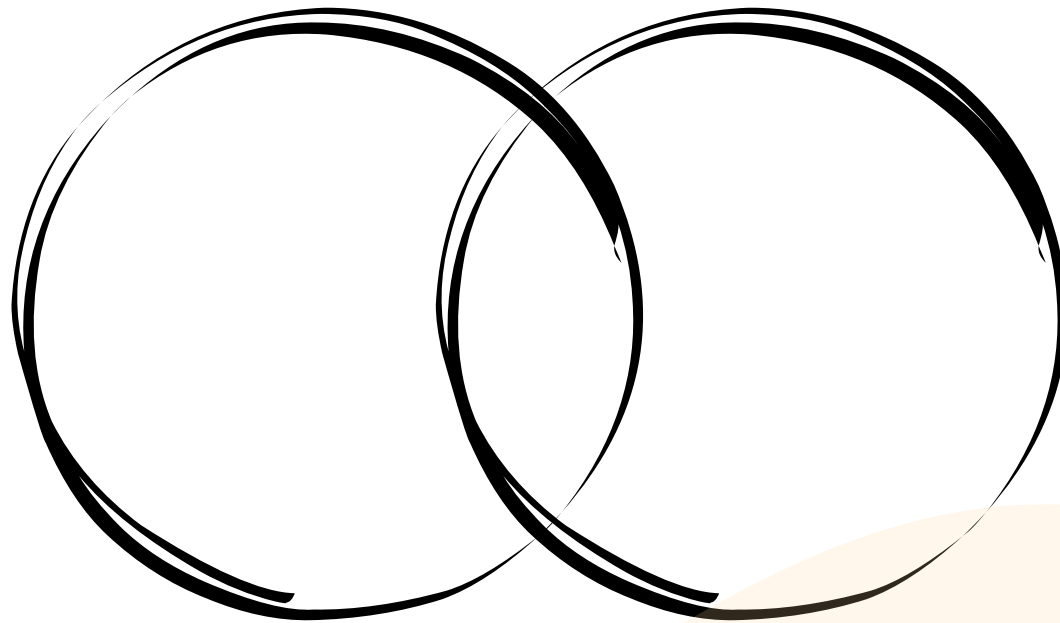
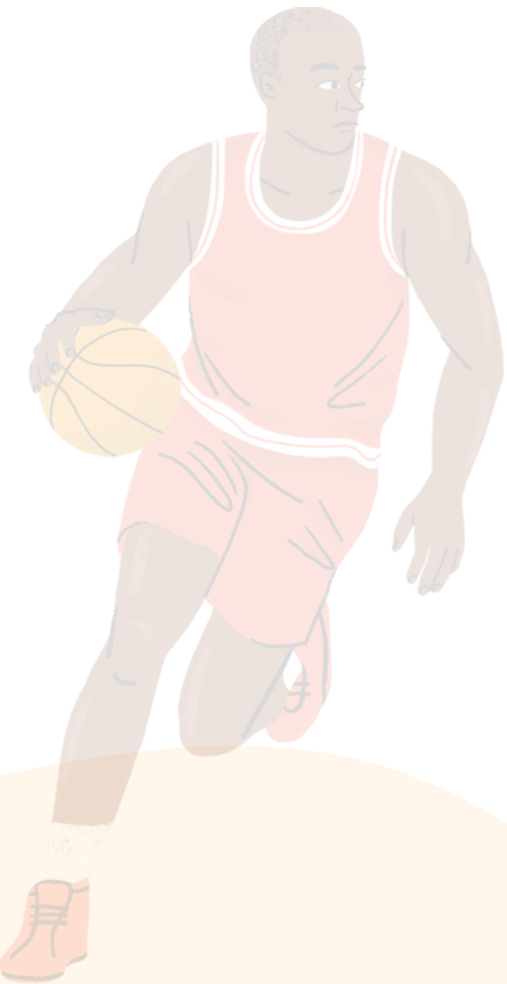
Características

Cada círculo representa un grupo de datos o concepto.

Las secciones superpuestas representan las cualidades que comparten los grupos de datos.

El área de superposición se llama "intersección".

El área que rodea los círculos engloba los datos que no entran en ninguna categoría.



Para aplicar el diagrama de venn y probabilidad se utilizo el conjunto de datos proveniente de las encuestas, siendo estas los deportes que se practican y cuantos deportes practica cada persona.

**Aqui entra la probabilidad**

- A) Que probabilidad hay que la persona haga 2 deportes
- B) Que probabilidad hay que haga deporte de contacto

**Datos:**

10 personas practican 2 deportes  
10 personas practican Muay thai  
50 personas encuestadoas

$$A \cup B = P(10/50) + P(10/50) - P(1/50)$$

$$A \cup B = (10/50) + (10/50) - (1/50)$$

$$A \cup B = 10 + 10 - 1 = 19/50$$

$$A \cup B = 0.38 \times 100 = 38\%$$

