



Data Science

Capítulo Estudiantil AAAIMX
Héctor Manuel Ruiz Juárez

INICIO



Primer caso de análisis

FACTURACIÓN POR ENCIMA DE \$ 30,000.00 POR ZONA



ID	Facturación	Zona
1	\$15000	SUR
.	.	.
.	.	.
50	\$18000	NORTE

Primeros pasos

1

IDENTIFICAR EL
TIPO DE VARIABLES

2

DETERMINAR EL
PORCENTAJE DE
CLIENTES POR ZONA

% de clientes por zona =

#facturas por zona/Total de facturas= %



Norte
 $31/50=0.62(62\%)$



Norte

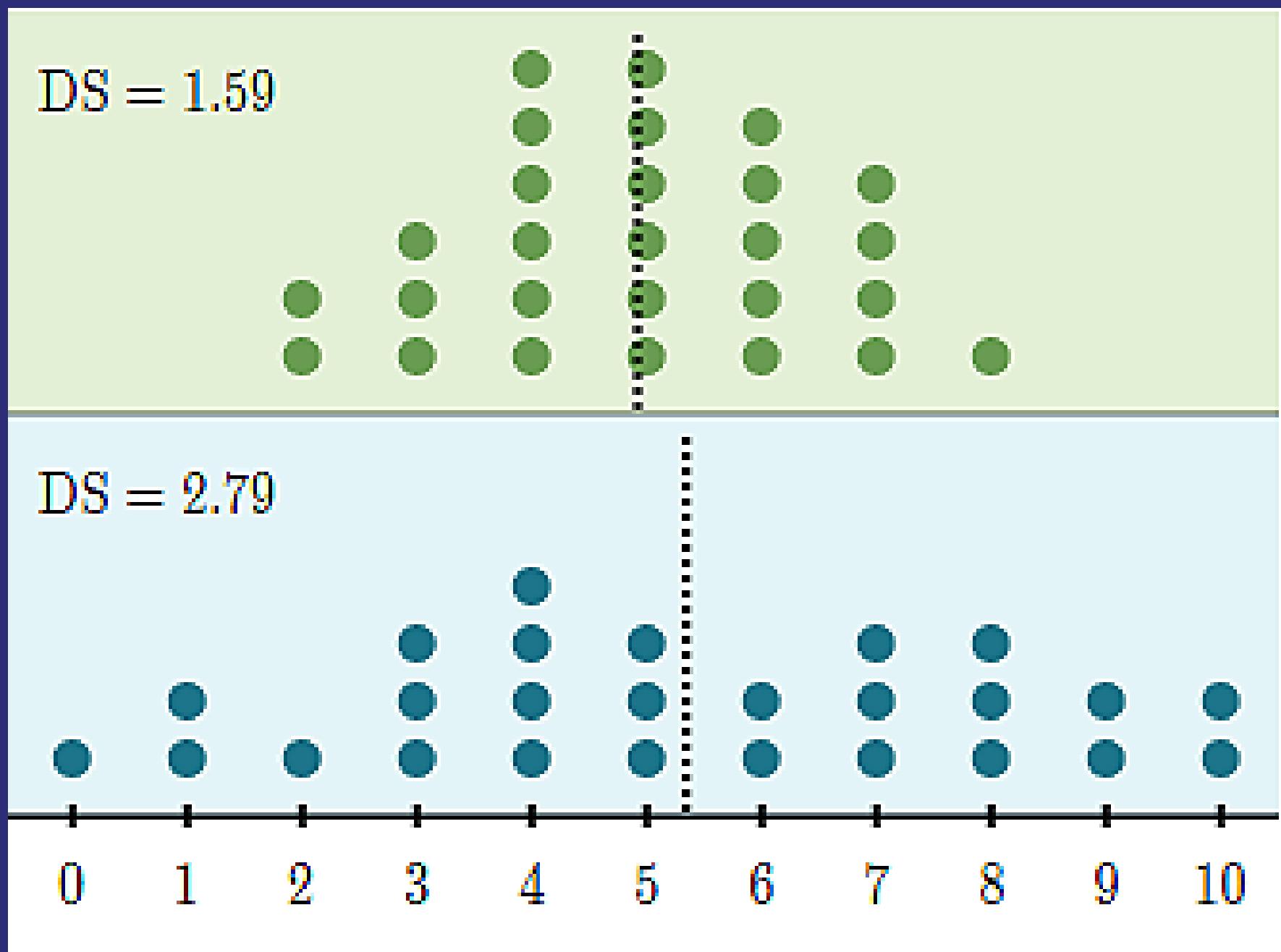
$31/50=0.62(62\%)$



Sur

$19/50=0.38(38\%)$

Media y Desviación Estándar



Norte
Promedio = \$28,468.4

Desviación estándar
\$10,174.24



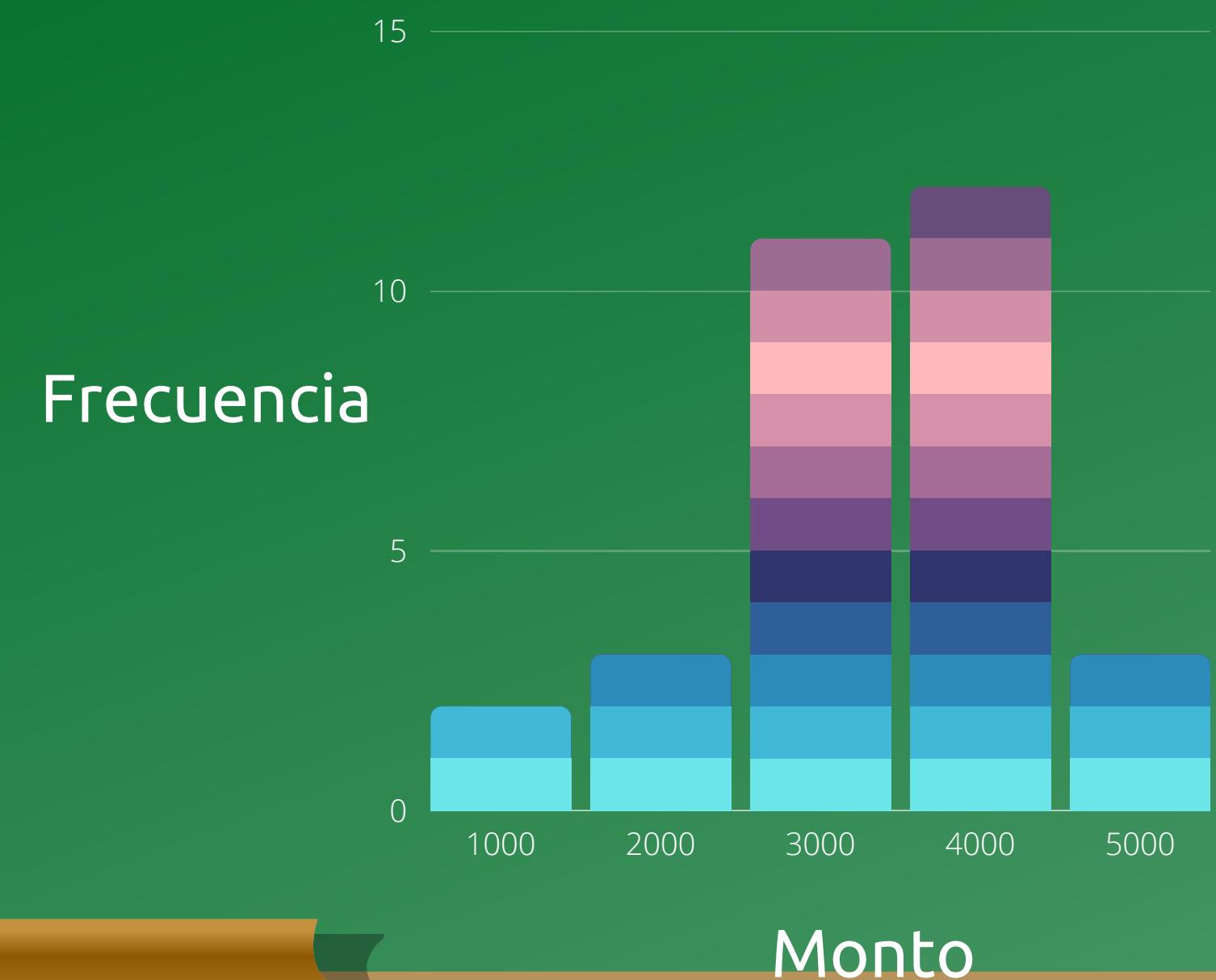
Norte
Promedio = \$28,468.4
Desviación estándar
\$10,174.24



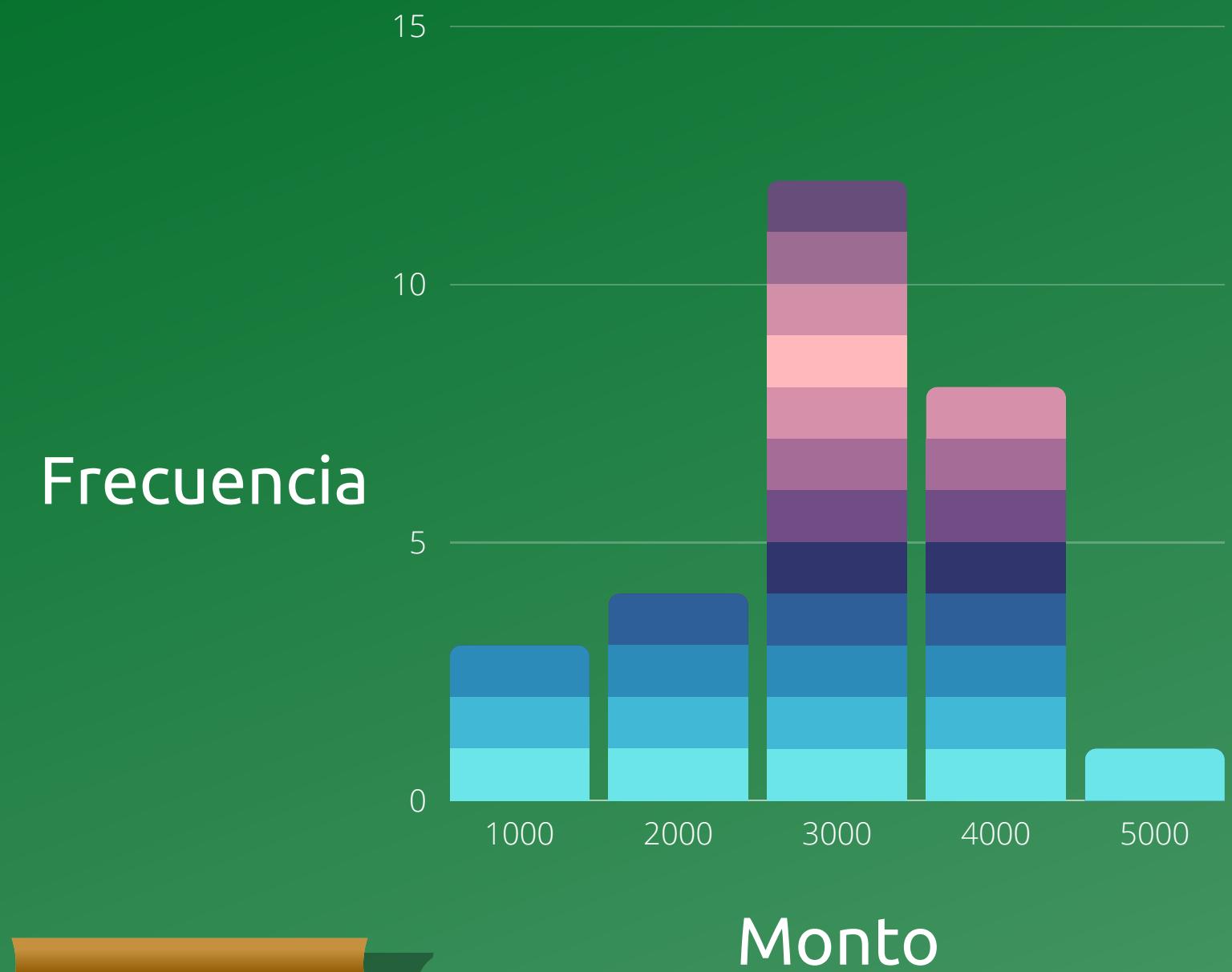
Sur
Promedio = \$24,450.8

Desviación estándar
\$10,267.66

Histograma de facturación Zona Norte

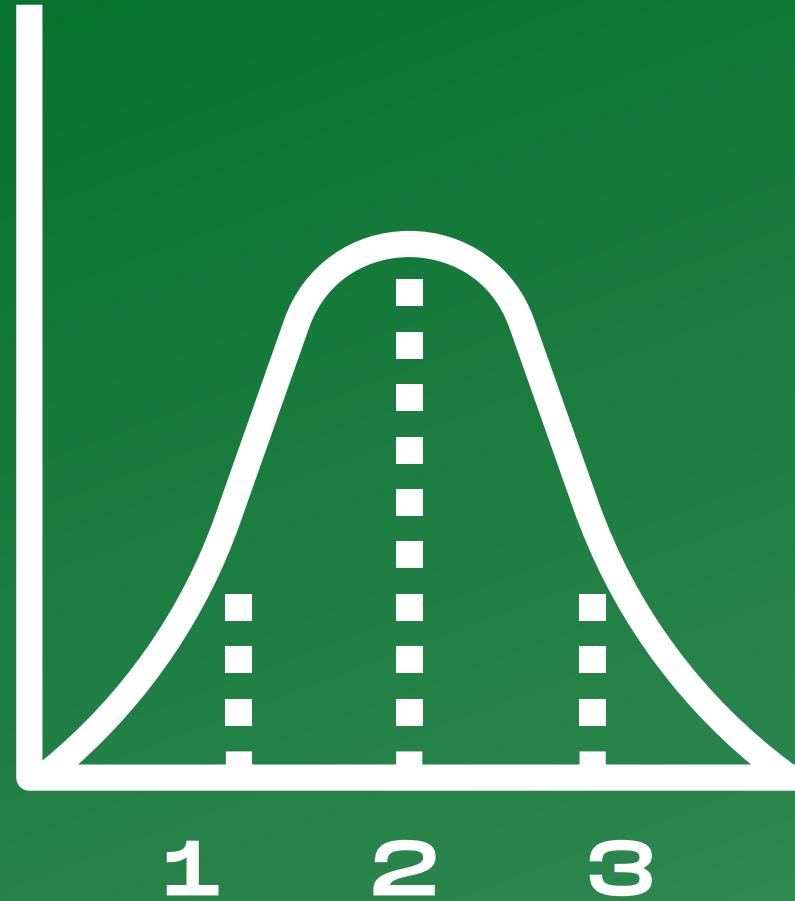


Histograma de facturación Zona Sur

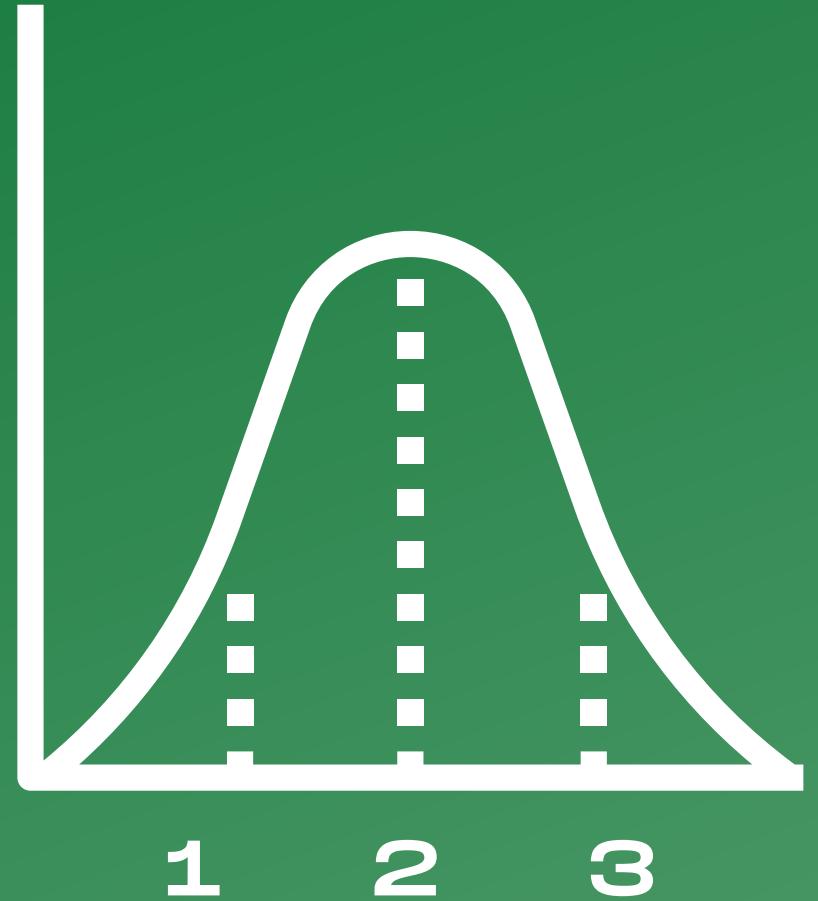


Probabilidad

Zona Norte



Zona Sur



Probabilidad

$P(\text{Ventas_Norte} > \$30,000)$

0.4417

$P(\text{Ventas_Sur} > \$30,000)$

0.2944

¿Qué se necesita
para hacer ciencia
de datos?



Estadística

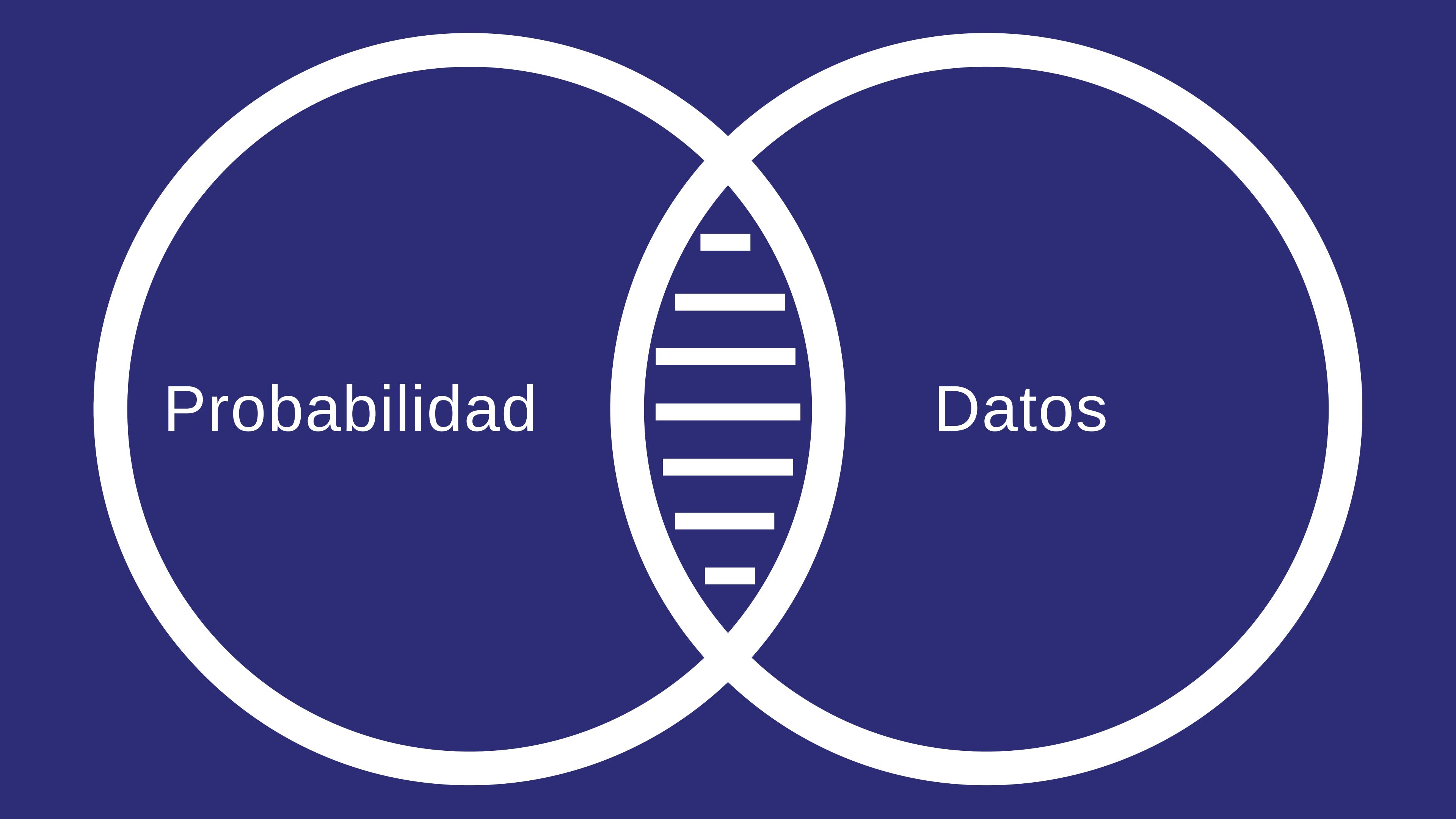
Recolectar
Organizar
Analizar
Interpretar
Presentar



Probabilidad

Medida numérica de la capacidad de ocurrencia de un evento cuyo resultado no se puede conocer con certeza





A Venn diagram consisting of two overlapping circles. The left circle is labeled "Probabilidad" and the right circle is labeled "Datos". The intersection area contains a white icon of a DNA double helix. The entire diagram is set against a dark blue background.

Probabilidad

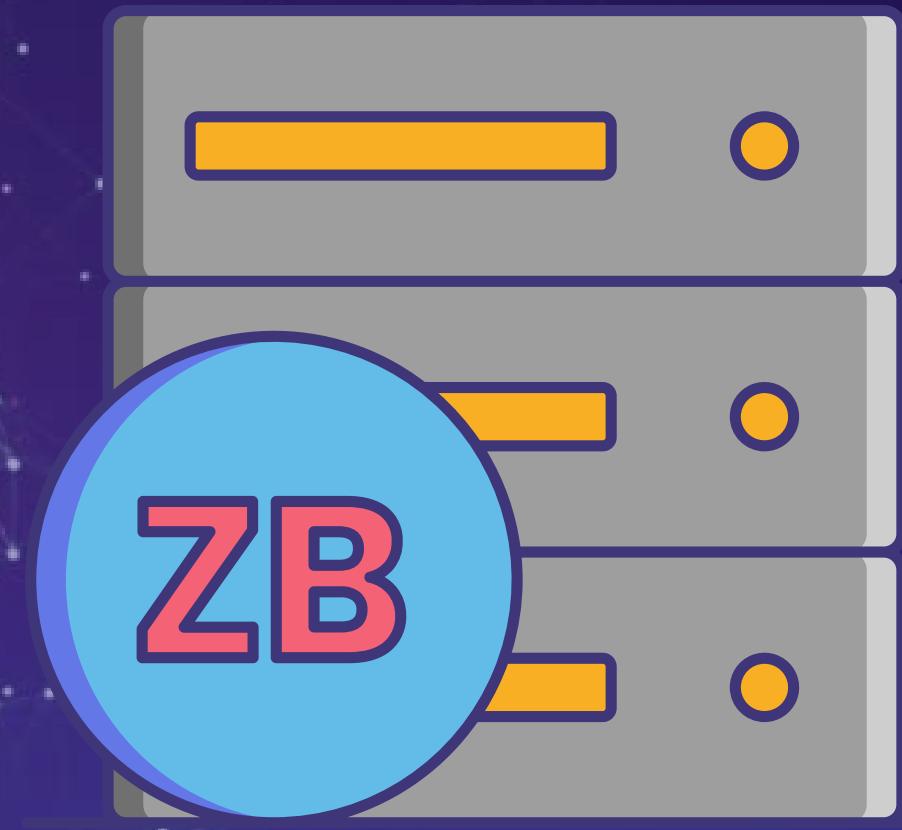
Datos

Datos

Representación
Simbólica de un
atributo o
medición



2035



Datos

Cualitativo



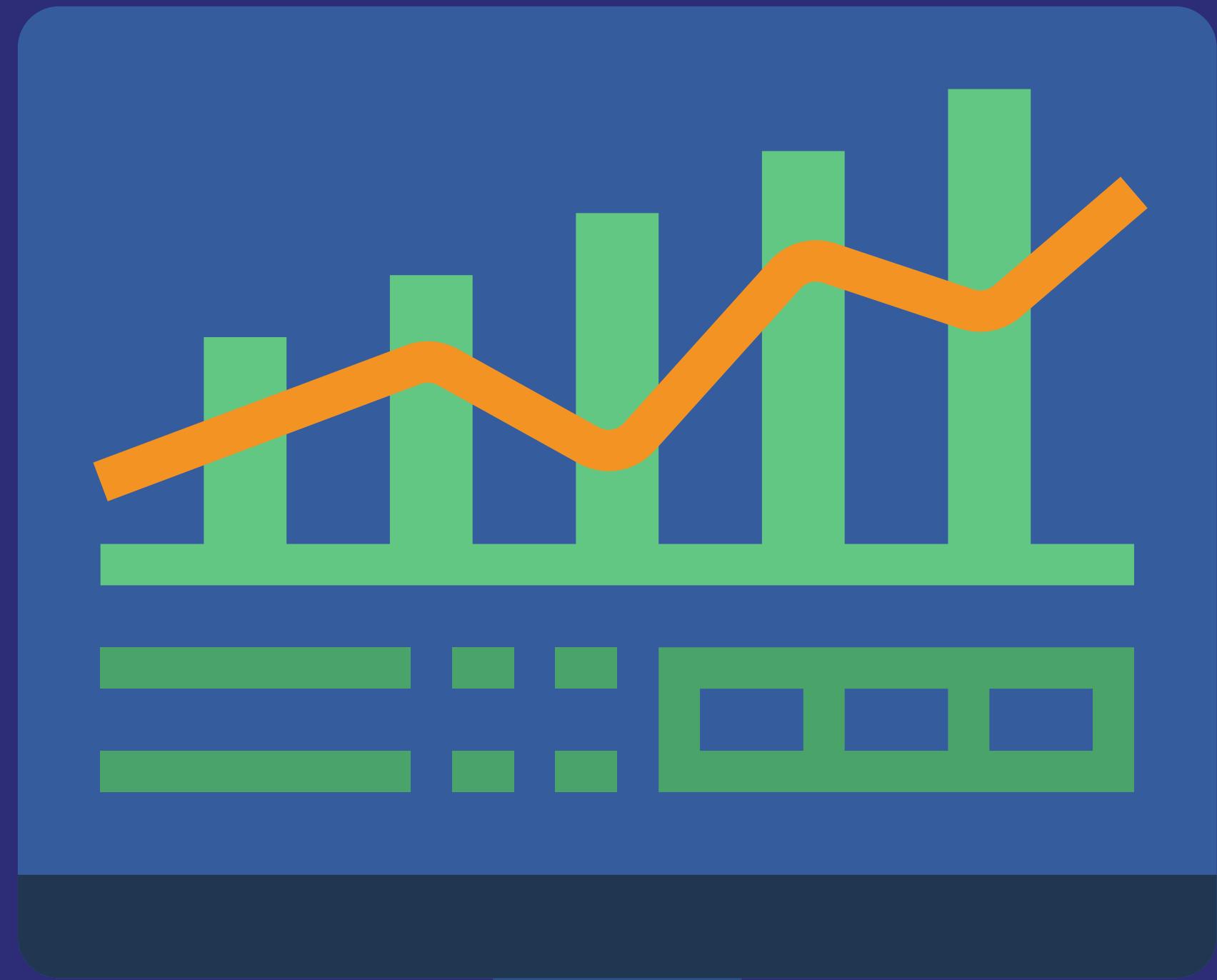
Nominal
Ordinal

Cuantitativo



Continuo
Discreta

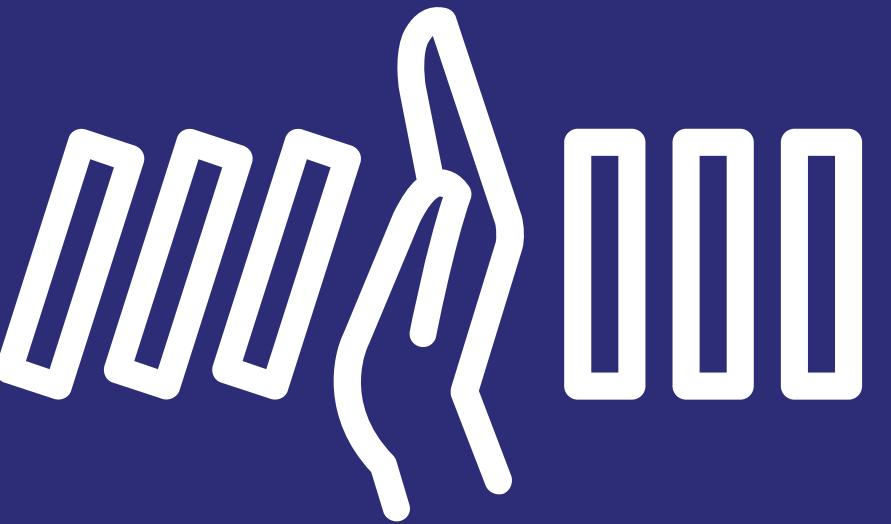
Estadística descriptiva



MODA



MEDIANA



MEDIA



VARIANZA

=VAR.P(B1:H1)

DESVIACIÓN ESTÁNDAR

=DESVEST.P(B1:H1)

MEDIA Y DESVIACIÓN ESTÁNDAR

HISTOGRAMA



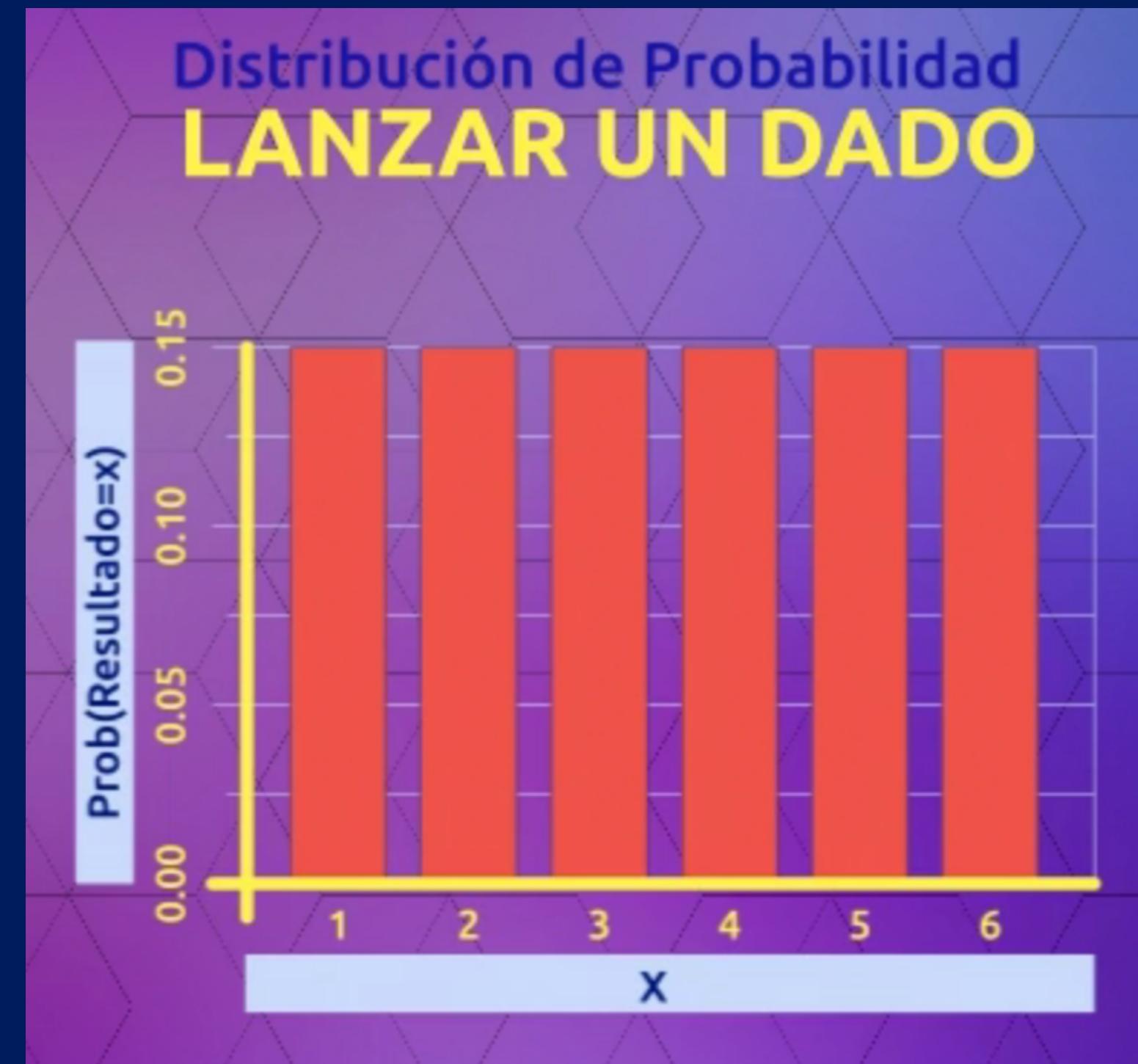


[https://quizizz.com/join?
gc=672552&source=liveDashboard](https://quizizz.com/join?gc=672552&source=liveDashboard)

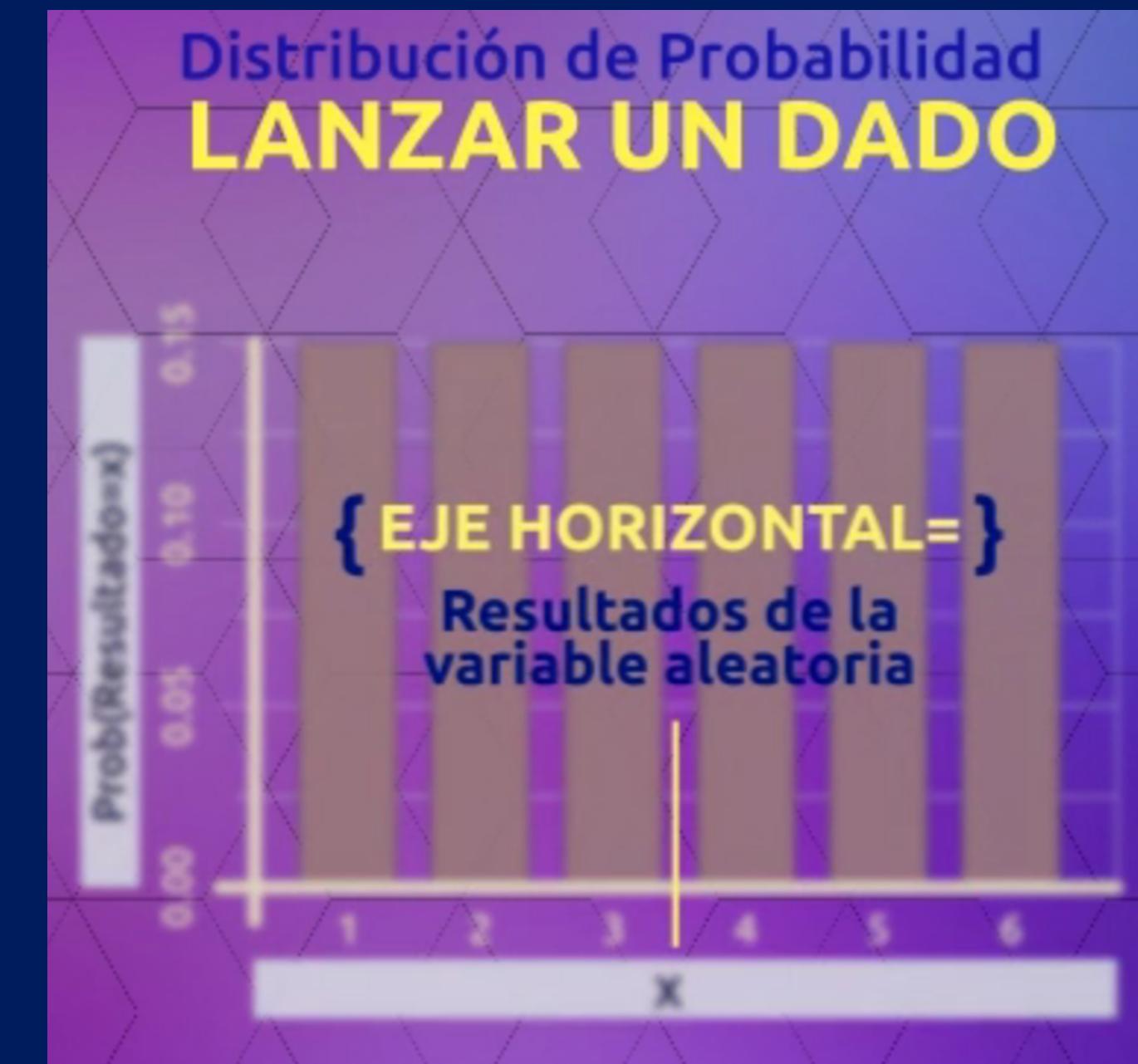
672552



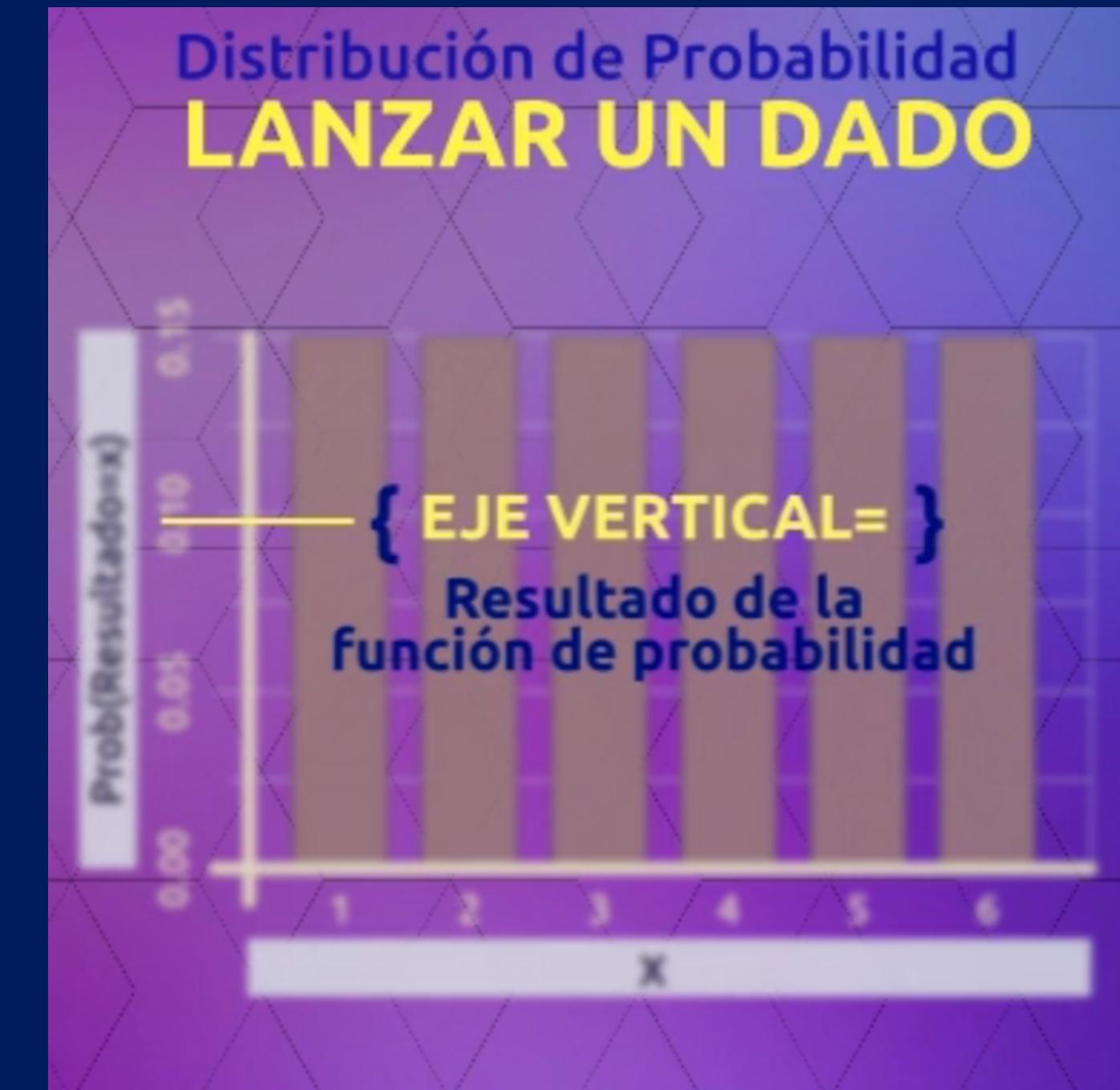
Variable Aleatoria



Variable Aleatoria



Variable Aleatoria



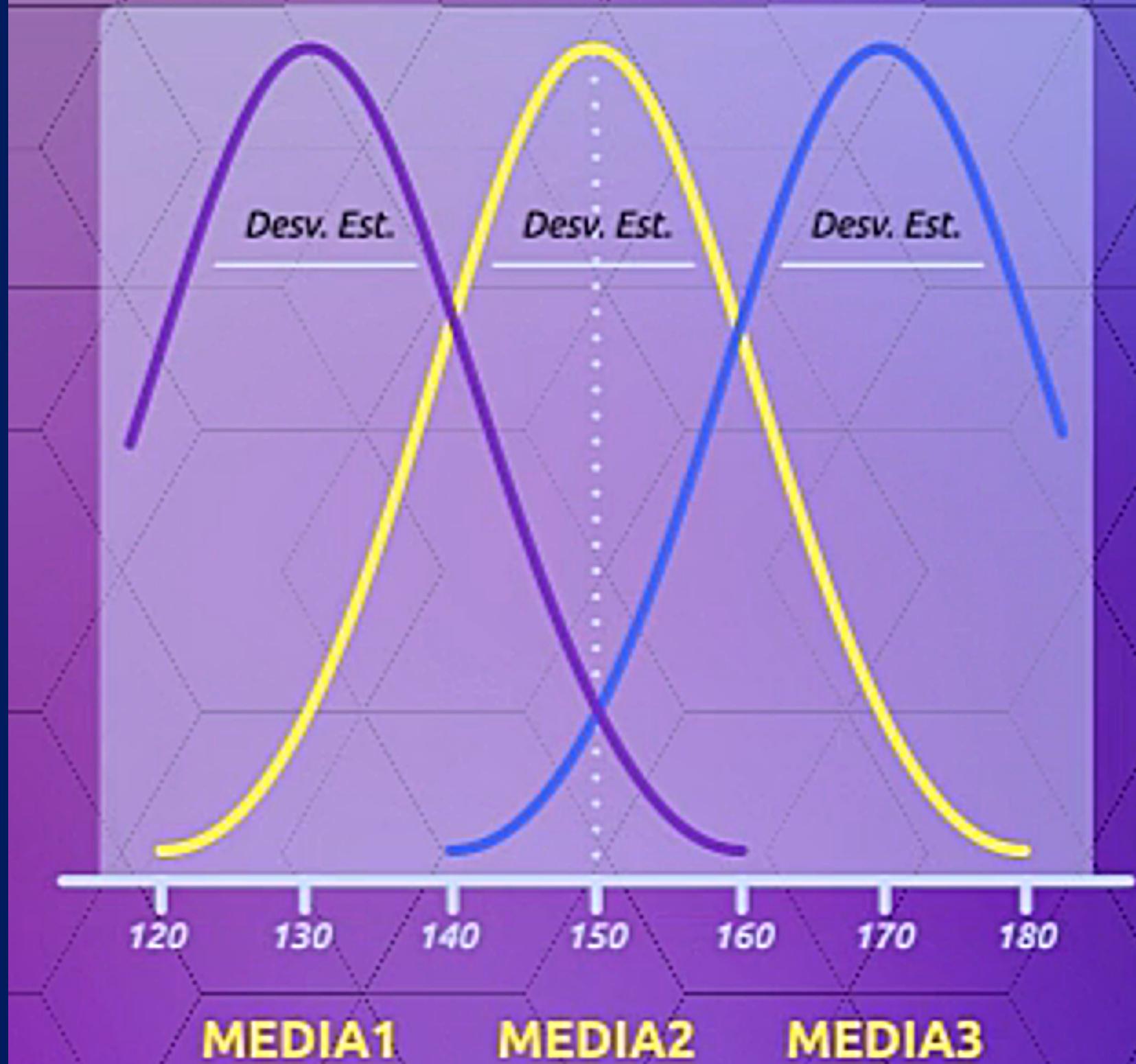
DISTRIBUCIÓN DE PROBABILIDAD UNIFORME

FUNCIONES PARA VARIABLES CONTINUAS

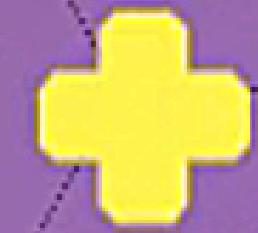
**FUNCIONES DE
DENSIDAD**

Misma desviación estándar
Diferente alineación

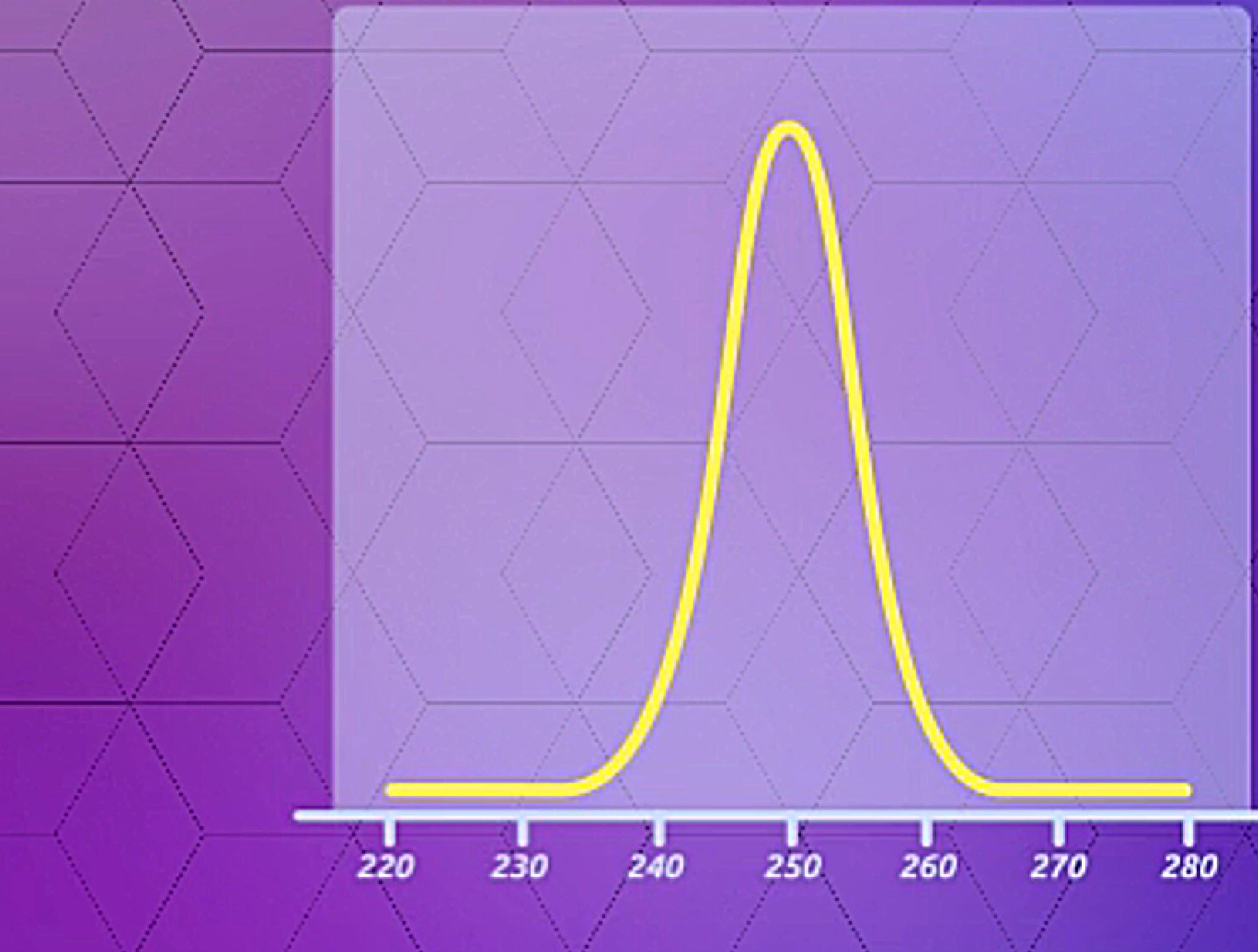
Distribución de normal **DIFERENTE MEDIA**



LEPTOCÚRTICA



Mayor
Concentración



PLATICÚRTICA

Menor
Concentración



220 230 240 250 260 270 280

MESOCÚRTICA



Evaluación II



[https://quizizz.com/join?
gc=789610&source=liveDashb
oard](https://quizizz.com/join?gc=789610&source=liveDashboard)





PROBABILIDAD

Medida numérica de la capacidad de ocurrencia de un evento.

CLÁSICA



FRECUENTISTA



TELCEL: 29,7%



73,4%

TELCEL Y
VISA:

20,3%

$$P(\text{Telcel o Visa}) =$$

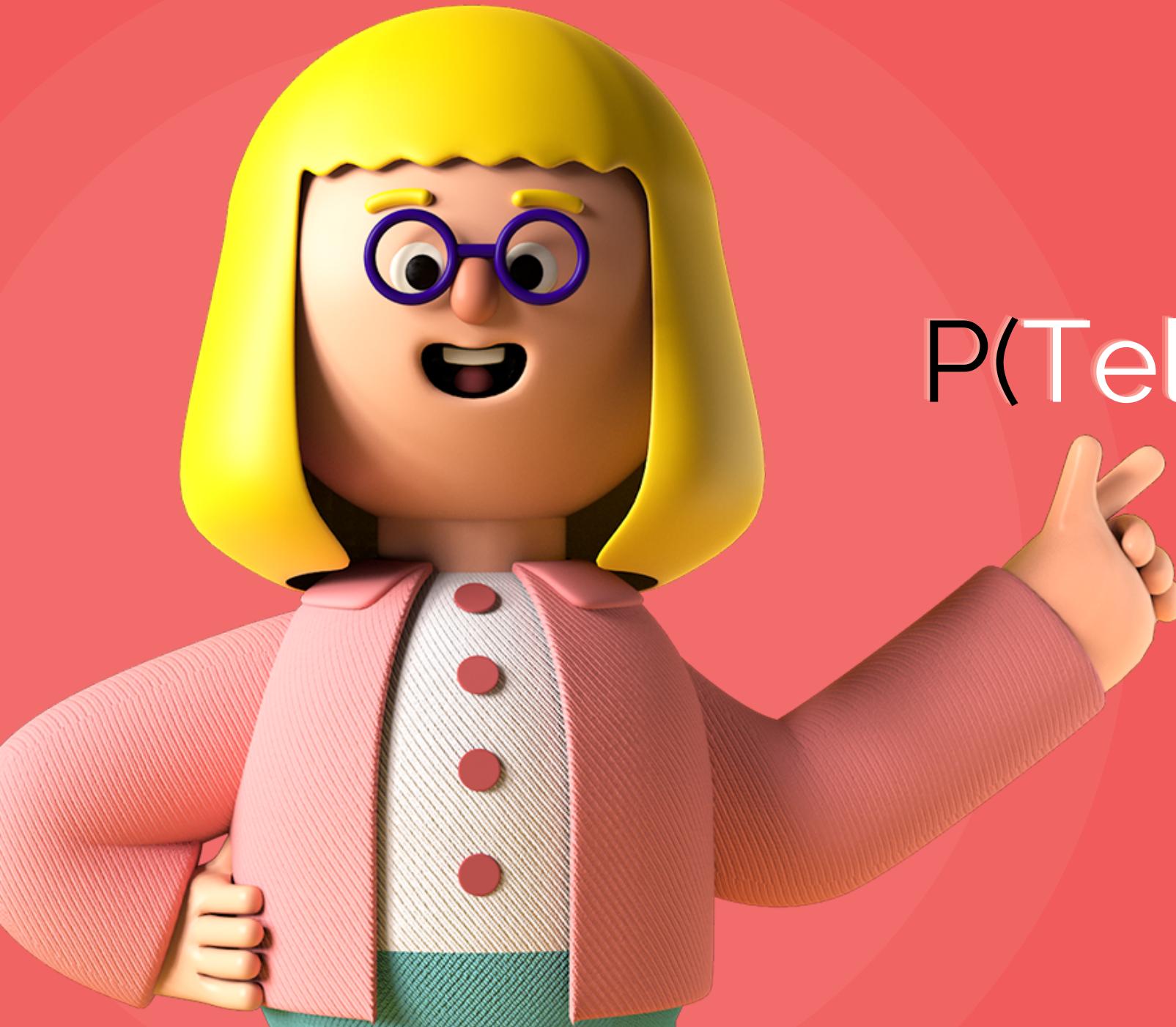
$$P(\text{Telcel}) + P(\text{Visa}) - P(\text{Telcel y Visa}) =$$

$$0.297 + 0.734 - 0.203 = 0.828$$

REGLA DE ADICIÓN

$P(\text{Telcel o Visa}) =$

$P(\text{Telcel}) + P(\text{Visa}) - P(\text{Telcel y Visa})$



PROBABILIDAD COMPLEMENTARIA DE UN EVENTO



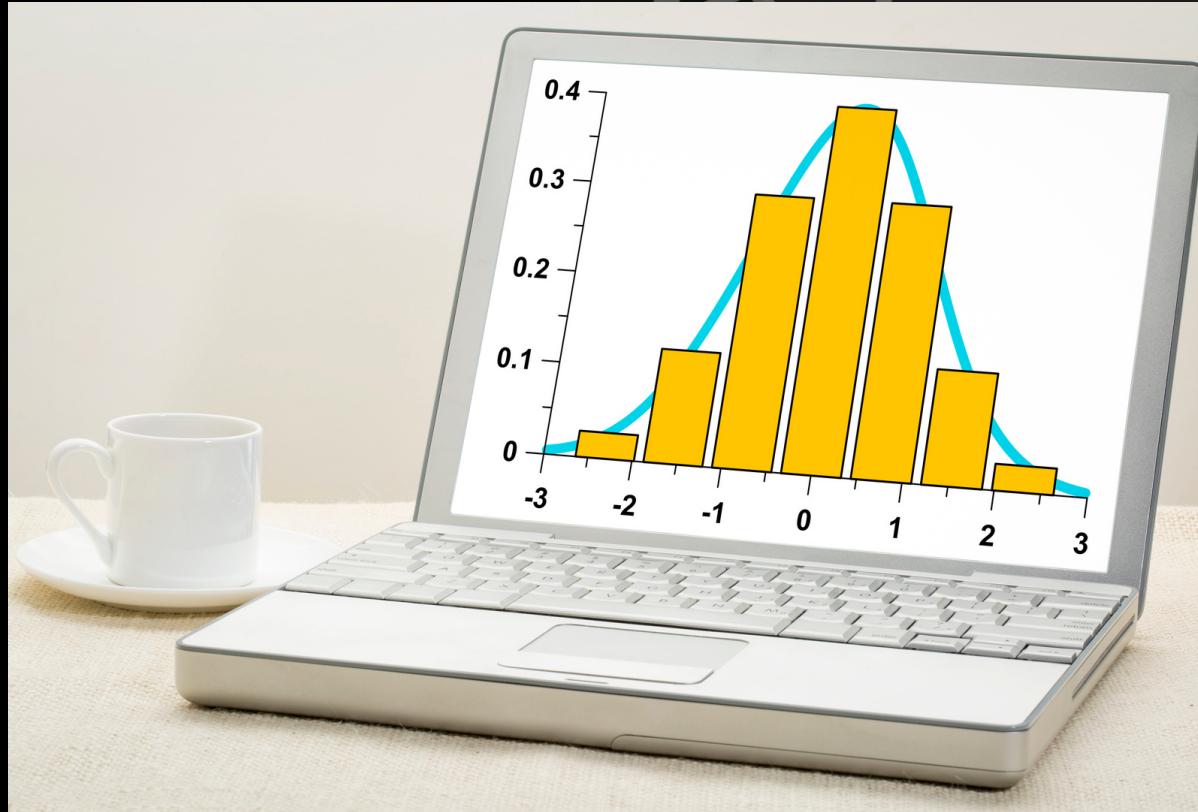
PROBABILIDAD COMPLEMENTARIA DE UN EVENTO



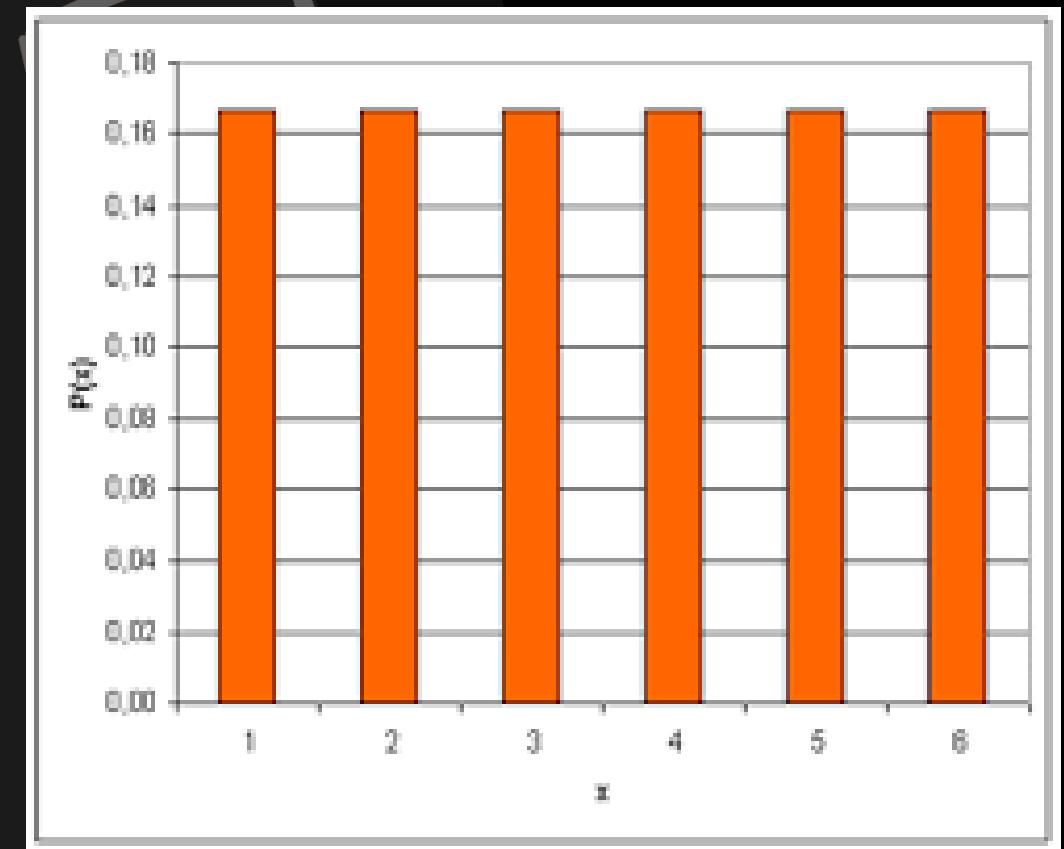
$$P(\text{No Visa}) = 1 - P(\text{Visa}) = 1 - 0.734 = 0.266$$



NORMAL



UNIFORME



1

100 %

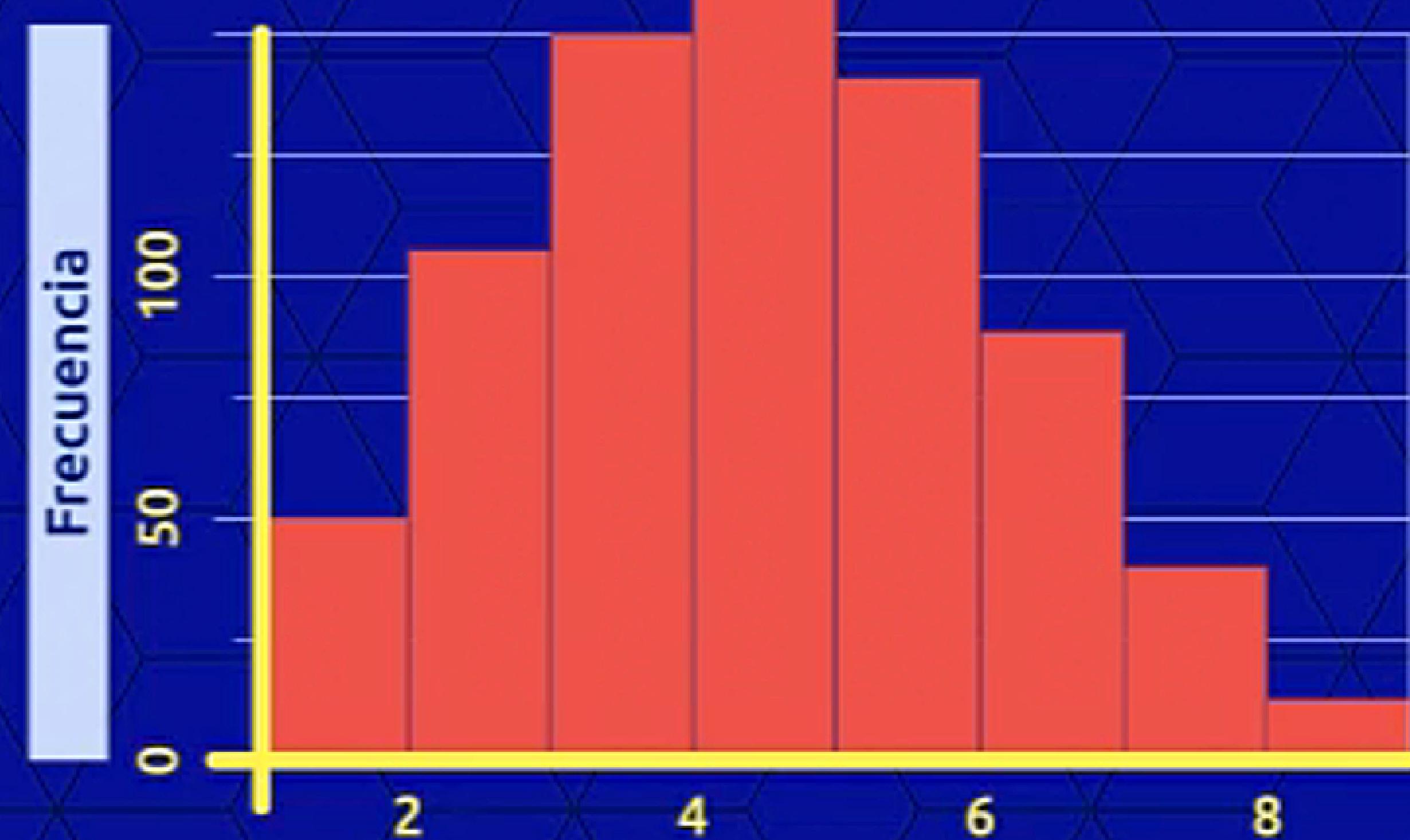
Ejemplo

Identificar el número de clientes diarios que recibimos en nuestra empresa para calcular algunas probabilidades.



HISTOGRAMA

clientes diarios



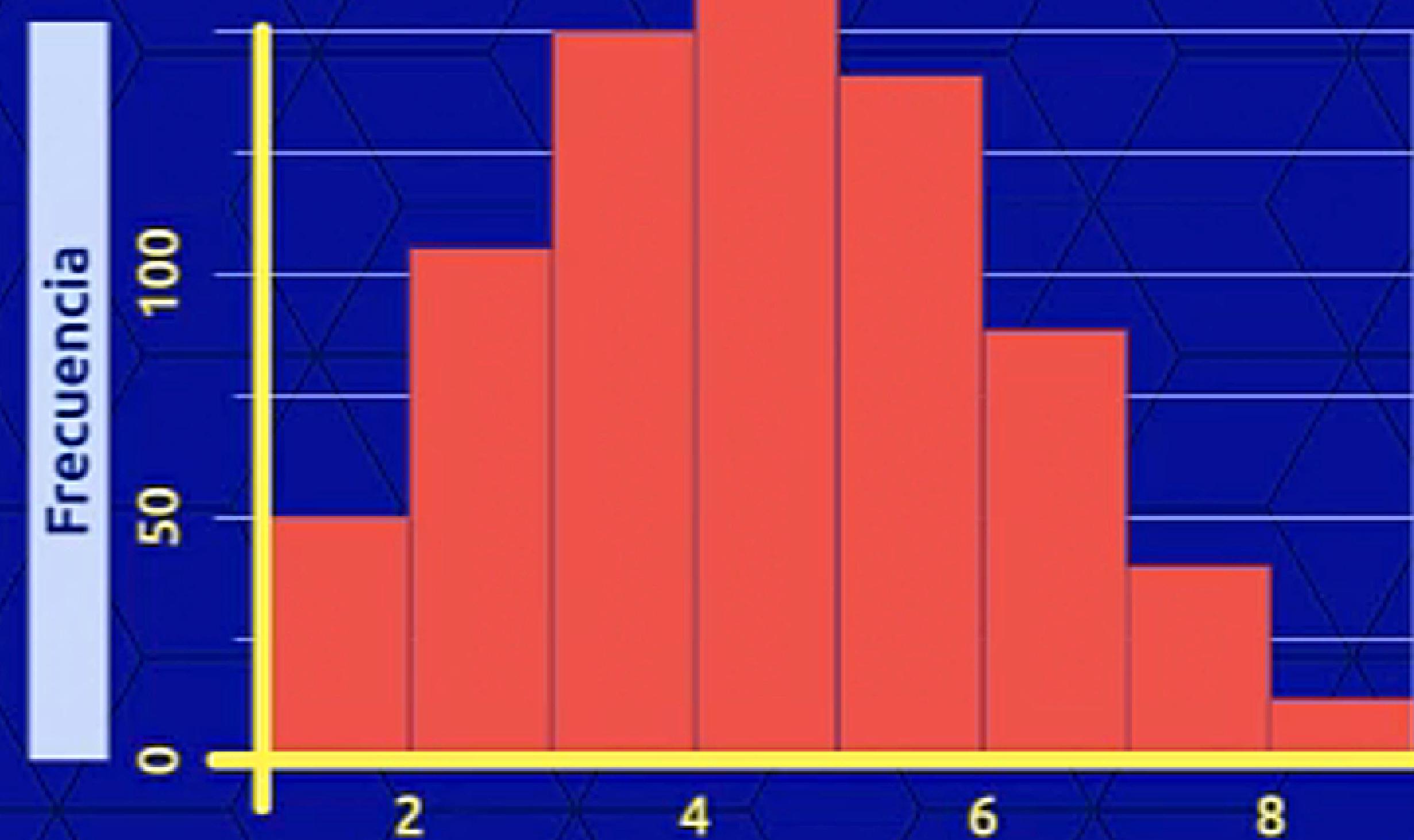
HISTOGRAMA

clientes diarios

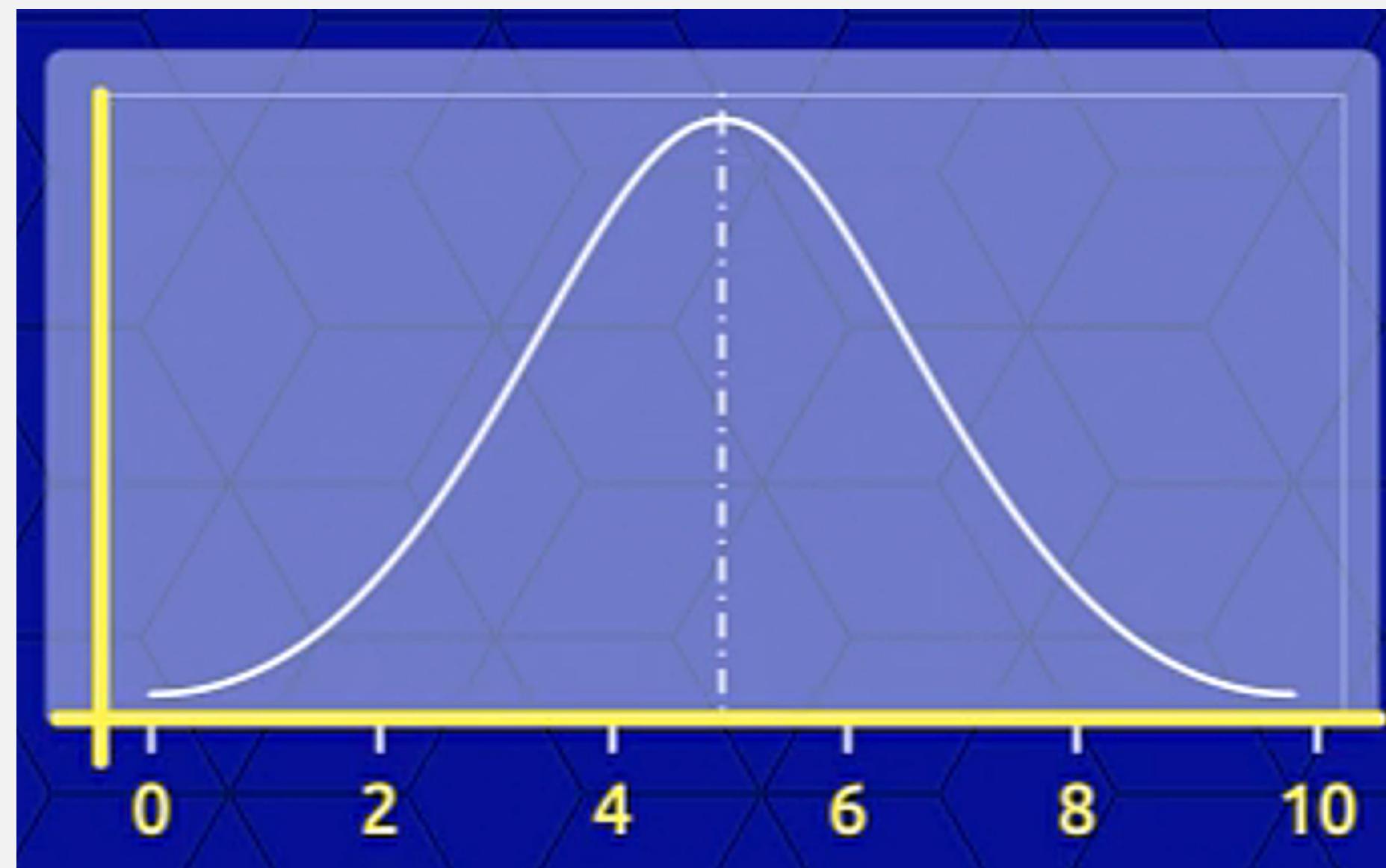
Distribución de
probabilidad
#Clientes

Promedio = 4.855

Desviación estándar= 1.653



Distribución normal

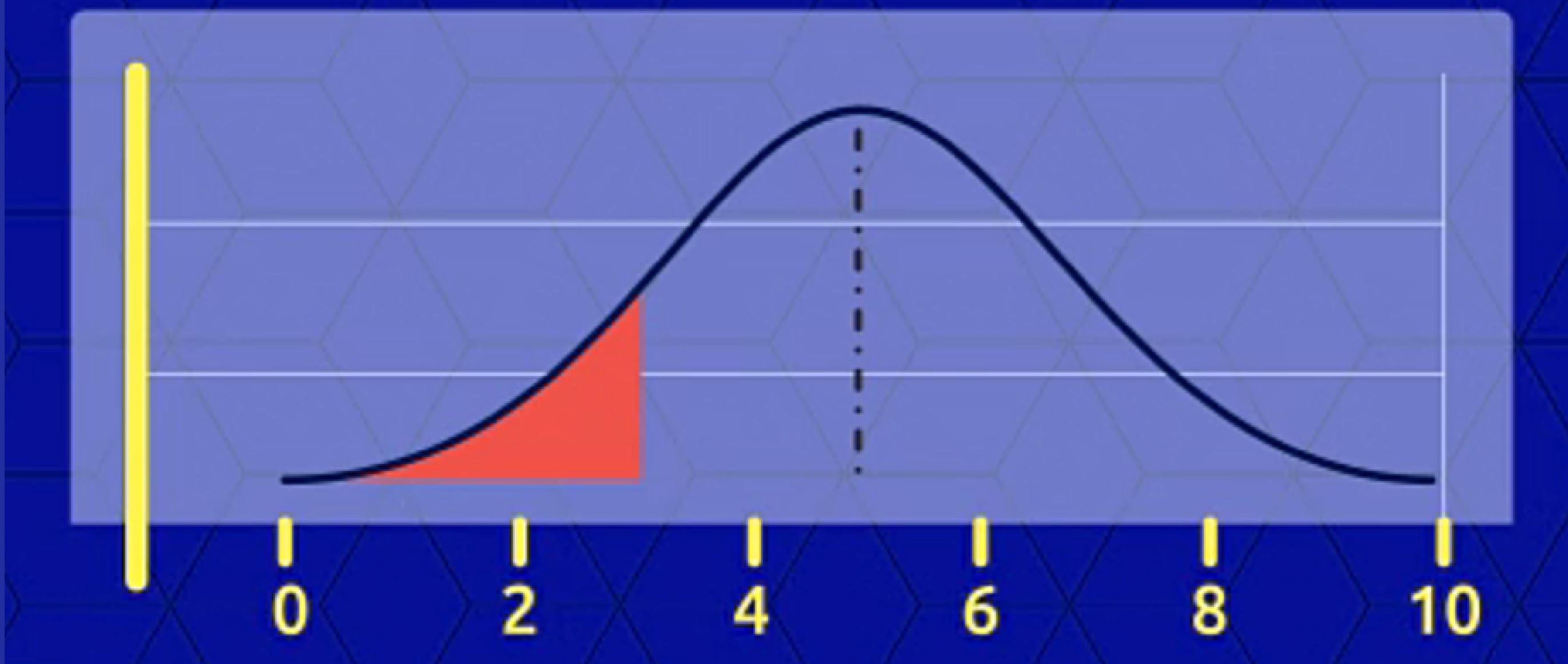


=DIST.NORM.NO

- **X**
- **MEDIA**
- **DESV_ESTÁNDAR**
- **acomulado**

Distribución de Probabilidad

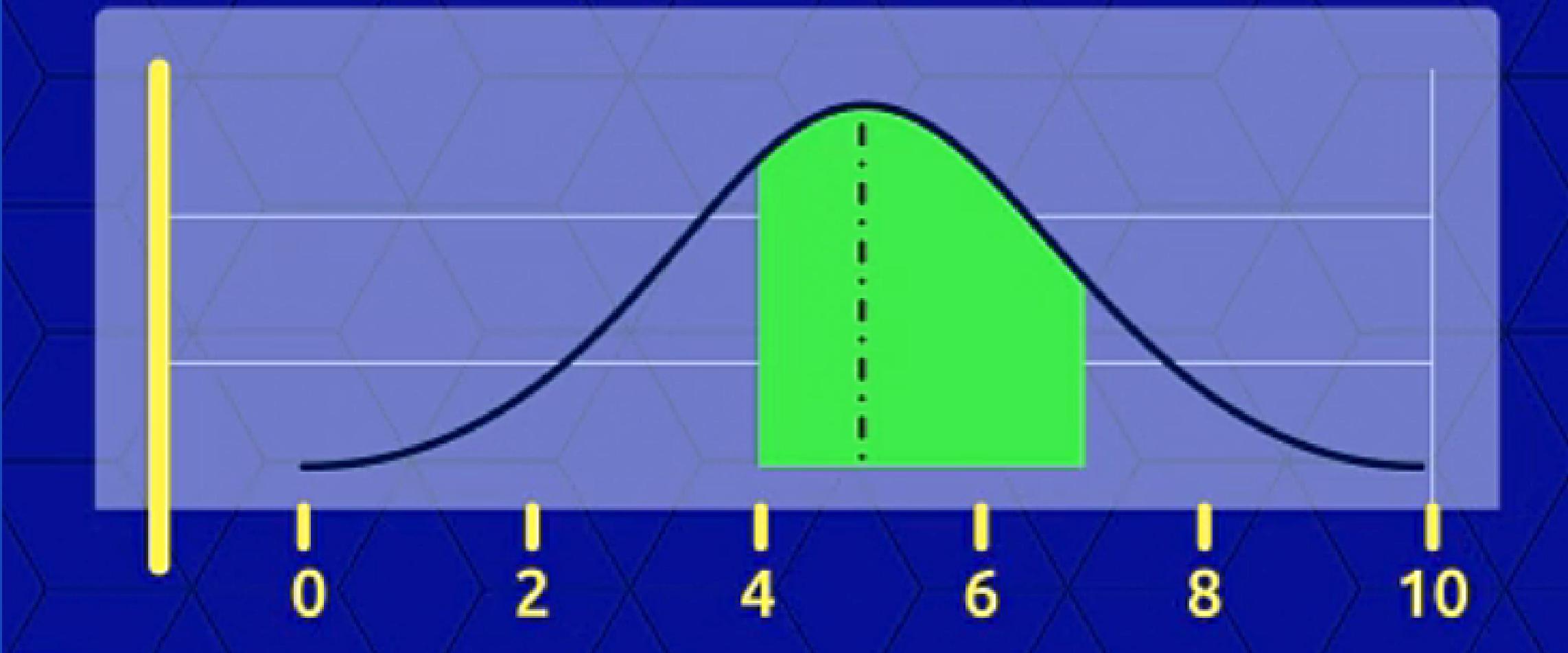
DE CLIENTES



Distribución de Probabilidad # DE CLIENTES



Distribución de Probabilidad **# DE CLIENTES**

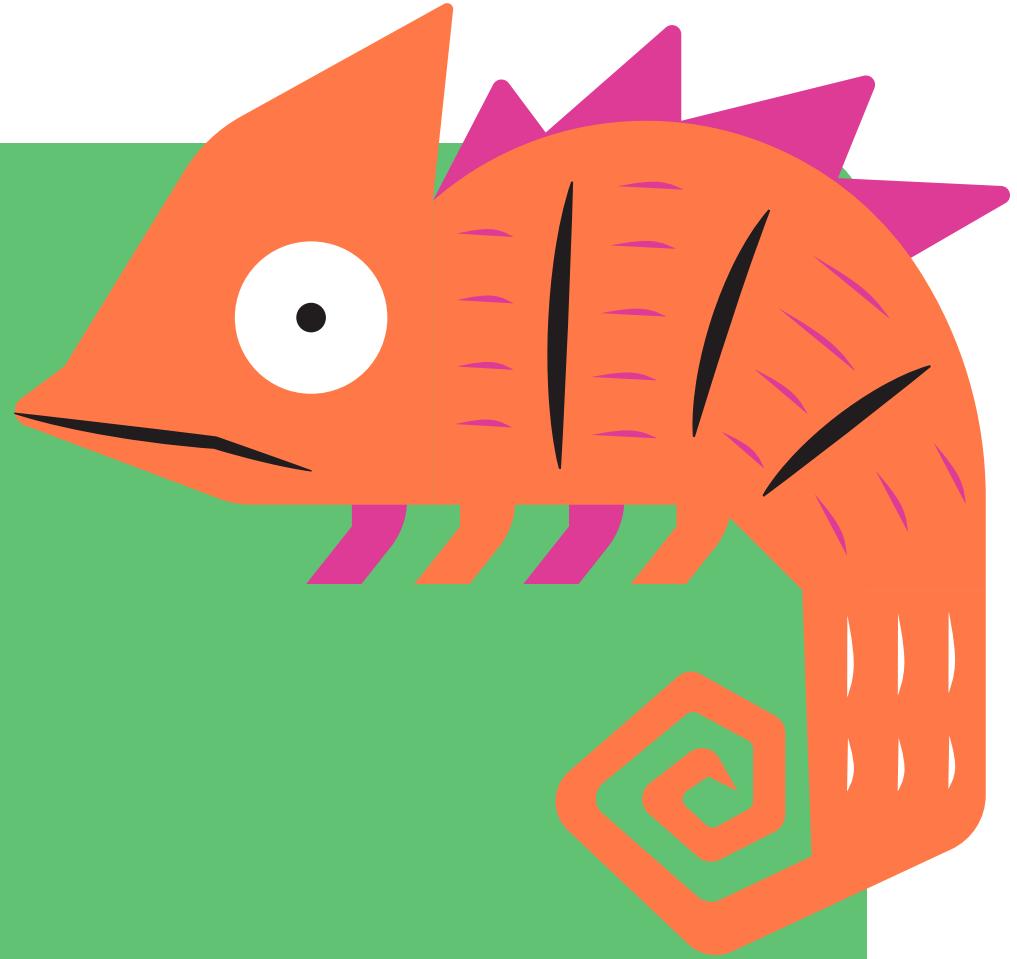




ÚLTIMA RONDA

Evaluación III

[https://quizizz.com/join?
gc=854951&source=liveDashboard](https://quizizz.com/join?gc=854951&source=liveDashboard)





FIN

academy