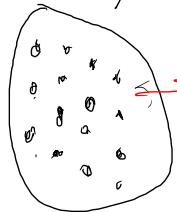


Fundamentos de Estadística.

$X = \{ \text{conjunto de datos} \}$.



→ proveer información

Statística descriptiva (Descripción de datos)

promedio X

→ Variables (cuantitativa → cantidad
cualitativa → cualidades)

Probabilidad
→ medida de la
posibilidad

P [o -]

Cuantitativa → cantidad
Cualitativa → cualidades.

Ejemplo: $X = \{1, 2, 3, 4, 5\}$

• $\bar{X} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$. \rightarrow Característica cuantitativa.

- Son números naturales mayores e iguales a 1 pero menores e iguales a 5.
- Son los cinco primos \exists^+ si en el caso. \rightarrow Cualitativa

↳ El objetivo de analizar un conjunto de datos es obtener conocimiento de los datos.

Herramientas para Procesar Datos

int, float, bool, str

- lógica computacional → Tipos de datos
Condicionales → If, Else
Ciclos → for, While
Funciones → def foo(): , math.sqrt(2)
Objetos → Clases, Paquetes.
Estructura de datos
 - listas, tuplas
 - Diccionarios
- Comparte Scripts
Git
- Persistencia de Información → xlxs
.txt
.CSV
.SQL → Bases de datos
modelado Base de datos
- Excel.

Estudios sobre una Población

↳ Encuestas sobre preferencias Políticas

• Población = {conjunto Universal de datos en contexto}

Ej: Elecciones de Alcaldé para Bogotá

Población = | Todos los habitantes inscritos en el censo electoral
↓
5.935.722

• muestra = | Un sub-conjunto representativo de la Población?
1% de 5.935.722 = 59.357,12

Regla de 3

$$\frac{1\%}{100} \times 5.935.722 = \frac{(1\%)(5.935.722)}{100\%} \rightarrow (0,01)$$

muestra → estadísticas (media)

Promedio.

→ Promedio.

Íd (59.357)	Id	Nombre	Edad	Genero	Estado	Localidad	Escolar
				H/mn	↓		P
					2		B
					3		Pre
					4		Post

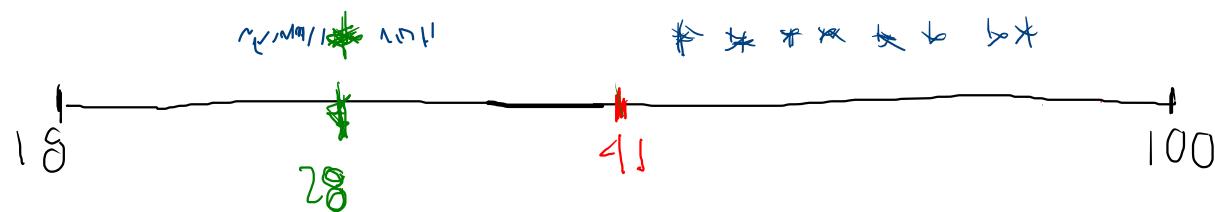
Promedio de Edad por Genero: Piso 18
Techo 100

3
4
5
6

Promedio de Edad por Genero por Localidad

Promedio de Edad por Genero por Escolaridad

Promedio de Edad / media



$$\text{Punto medio} \\ \frac{100 - 18}{2}$$

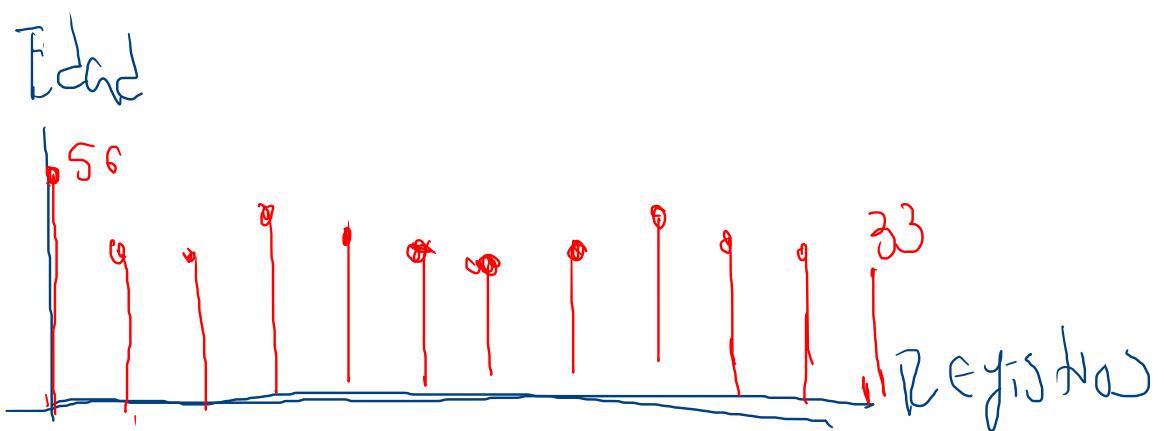
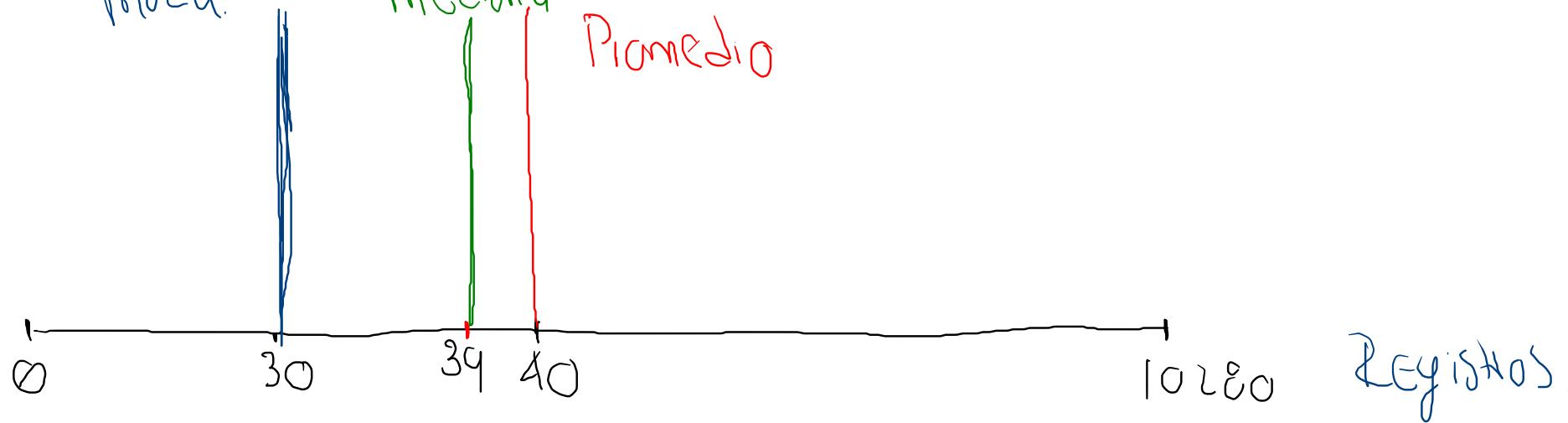
mediana

1, 2, 2, 3 | 3, 4 | 5, 6, 8, 9
1 2 3 4 | 5 6 | 7 8 9 10 → P25

moda: el dato que más se repite.

$$\frac{3+4}{2} = \frac{7}{2} \Rightarrow 3.5 \text{ mediana}$$

Distribución de datos 'Edad'



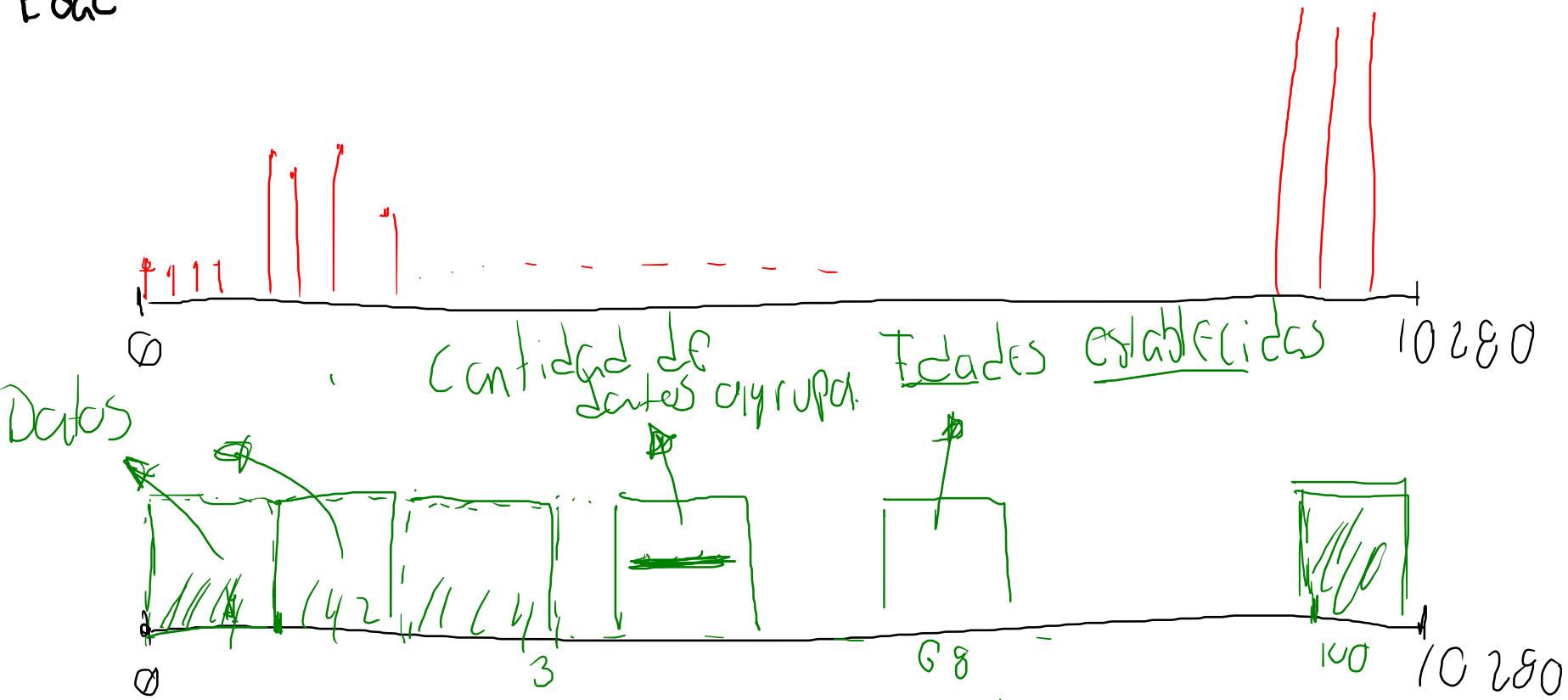
H → P_{-i}, P_{-c}, M_{-sys}, S_{-c}.

M → P_{-i}, P_{-c}, M_{-sys}, S_{-c}.

P_{-i} = [30]

Percentil = En 100 intervalos

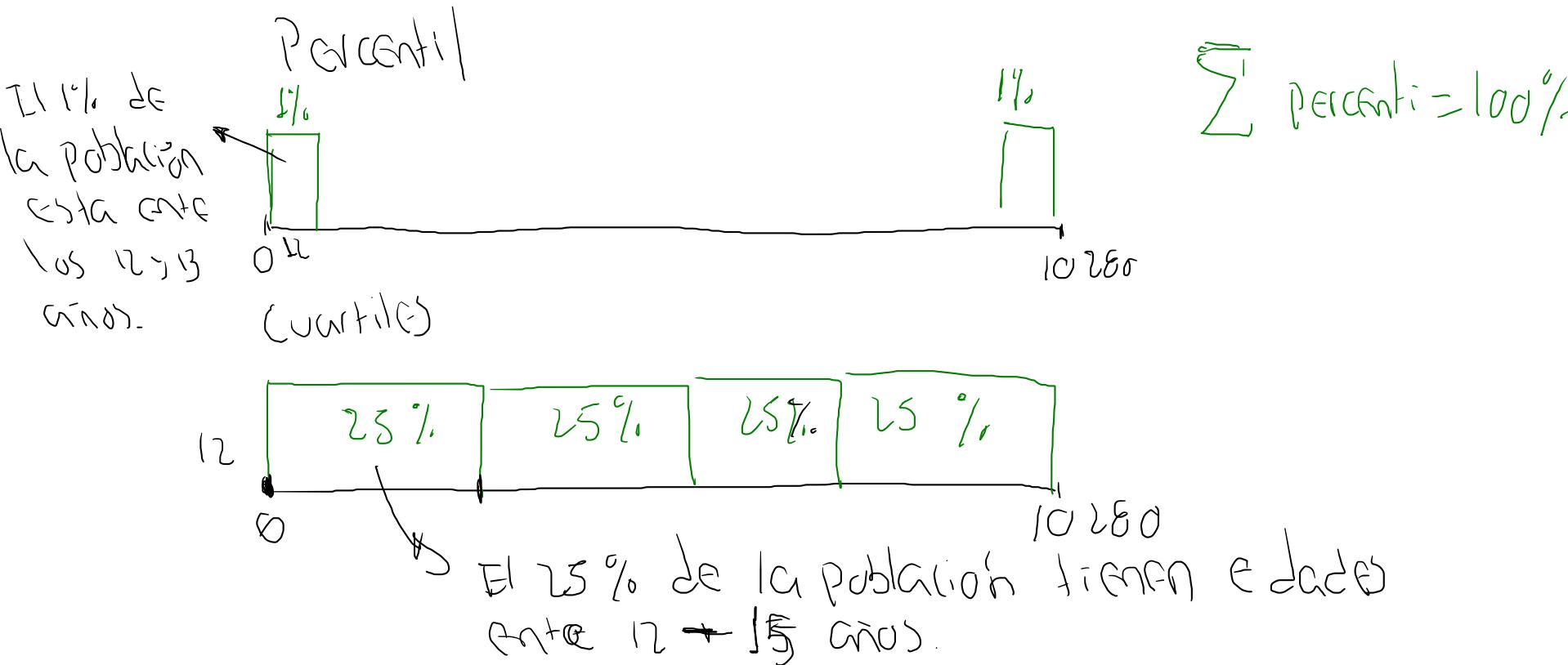
Edad



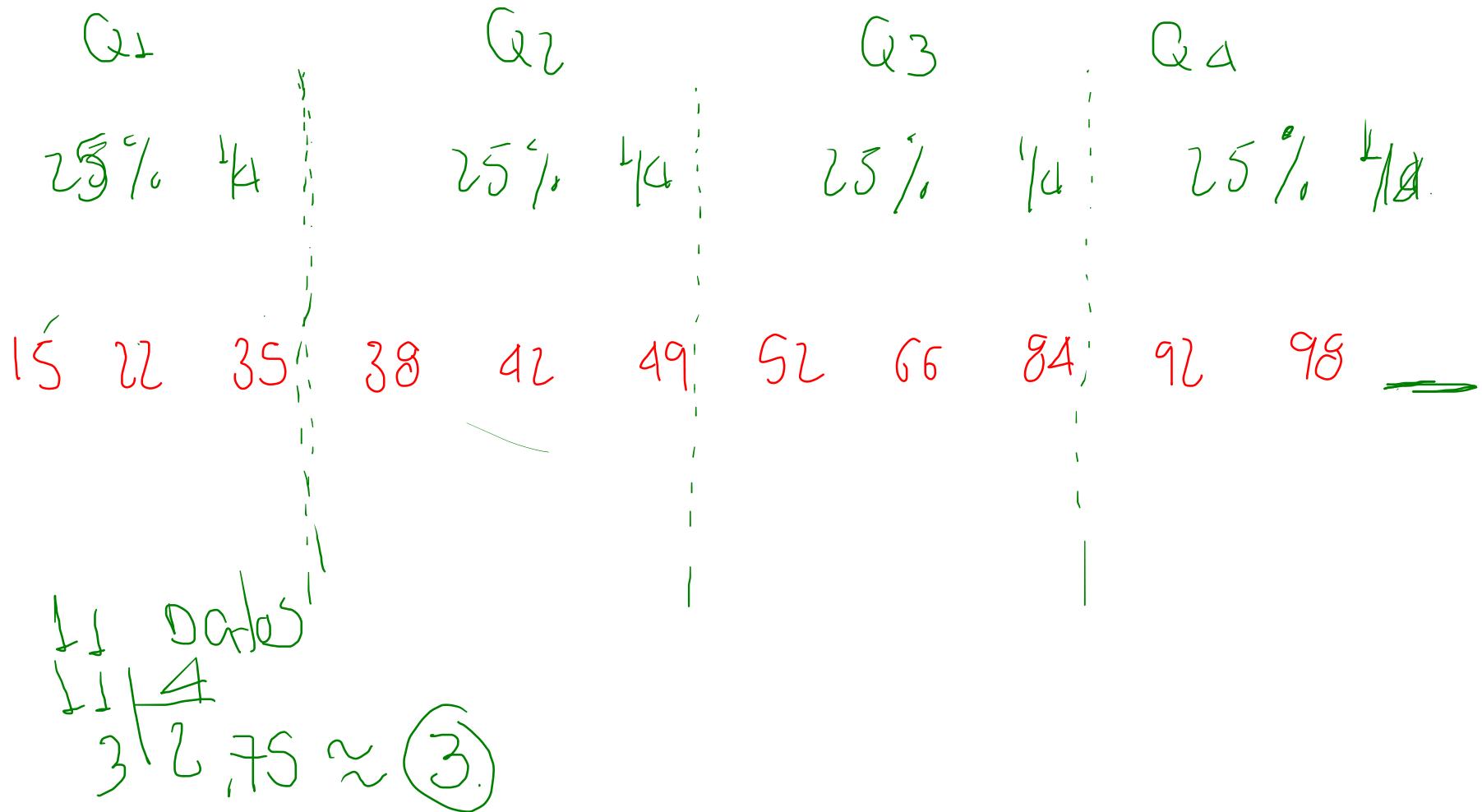
$$i = \frac{P}{100} n$$

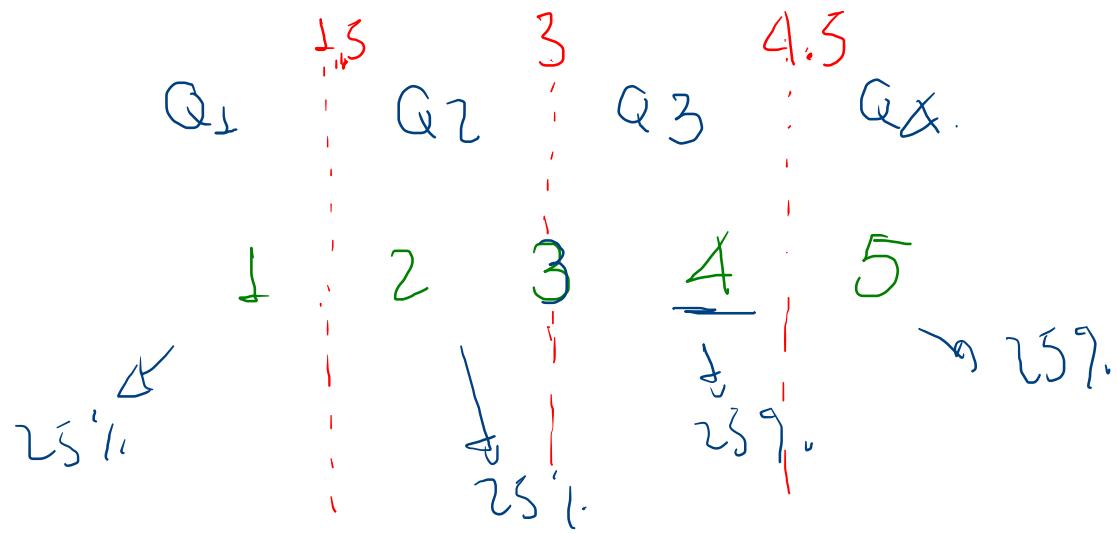
i = intervalo a Encuentrar
P = Percentil asignado
n = cantidad de datos agrupados.
100 = La cantidad de intervalos

Edades



Quartiles





D1

Q Q1

Percentil para "edad" → Lista de edades de data.xlsx

1% de los registros está entre 12 - 16 años

1% de los registros está entre 74 - 98 años

- 8 personas de 12 años pertenecen al 1% de los registros.
- 92 personas de 12 a 15 años son el 1% de los registros.
- 178 personas de 18 años son el 2% de los registros

4) limites de los Intervalos

16 - 1

17 - 1

18 - 2

19 - 1

.

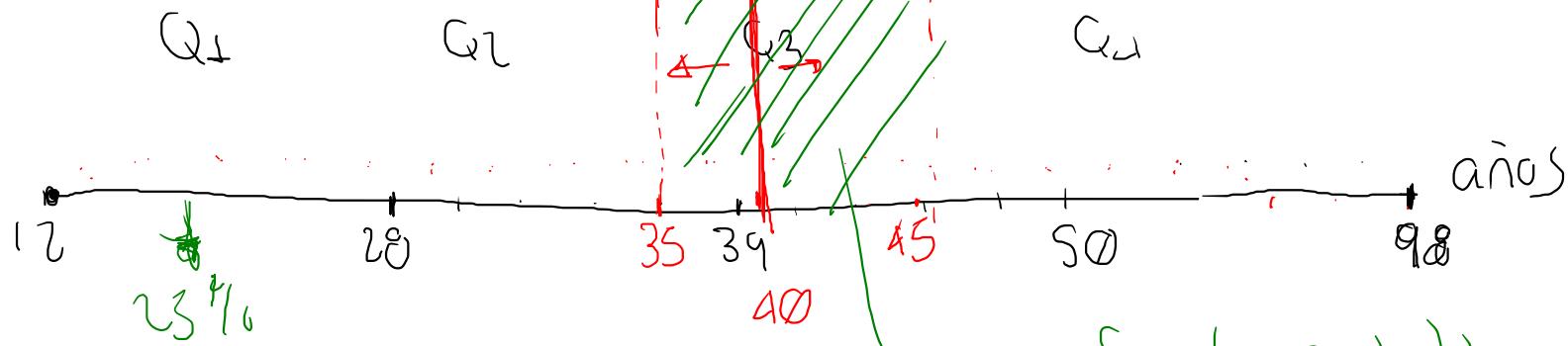
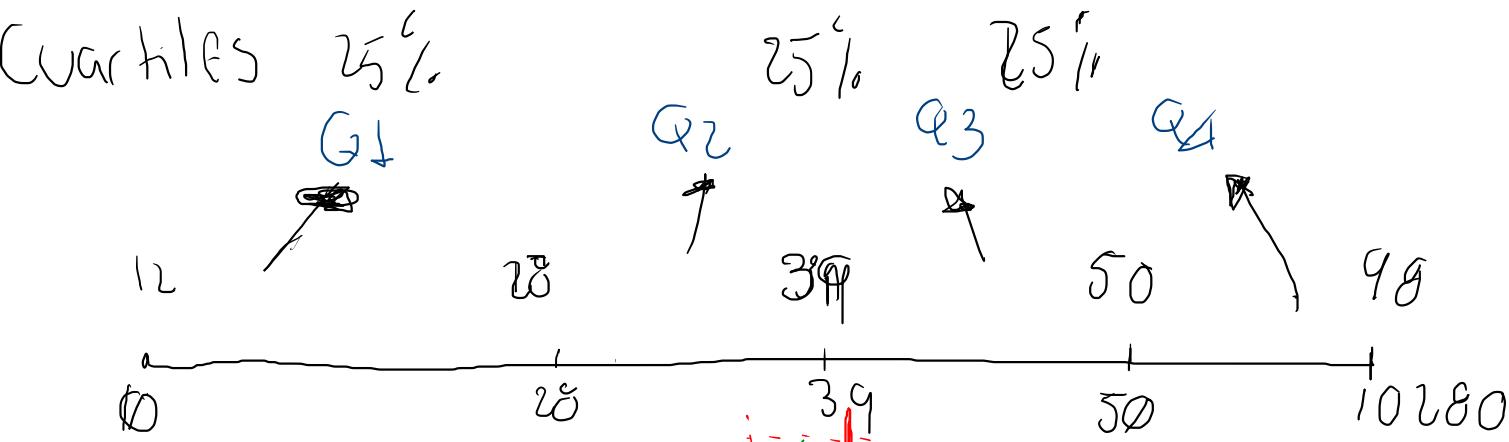
.

50 - 2

16 \rightarrow 1

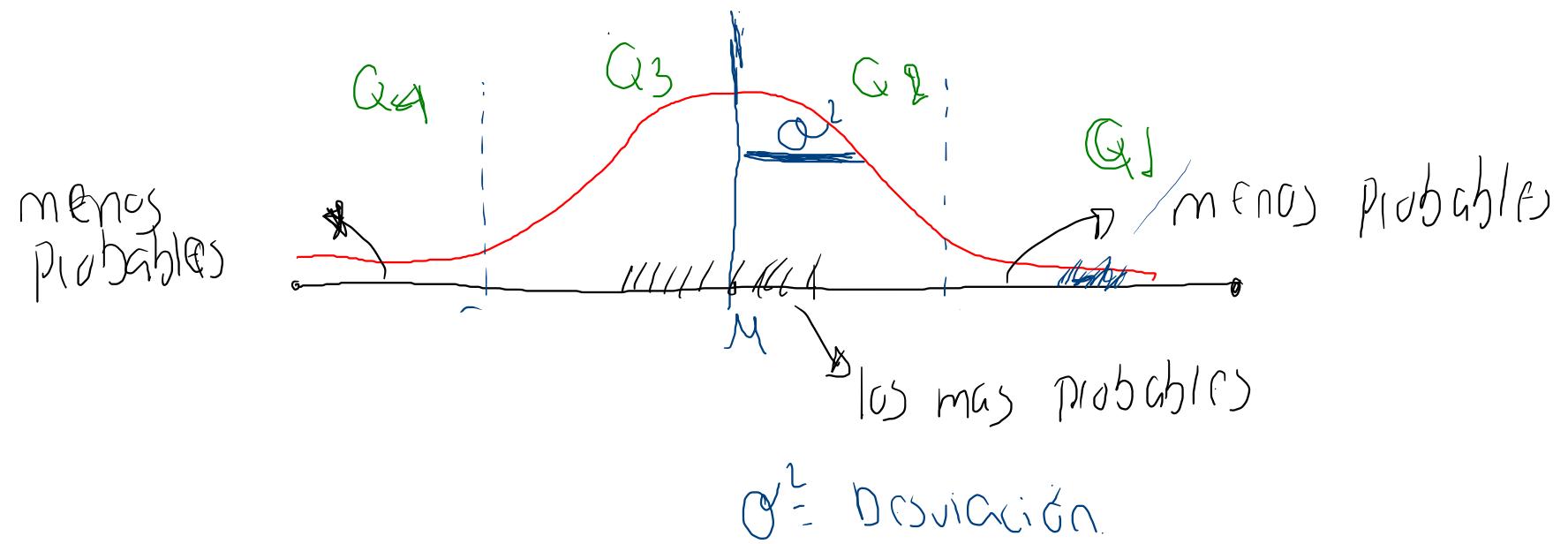
17 \rightarrow 2

- Las personas de 30 años son el 3% de los registros
- 270 personas tienen 30 años que corresponde al 3% de los registros.

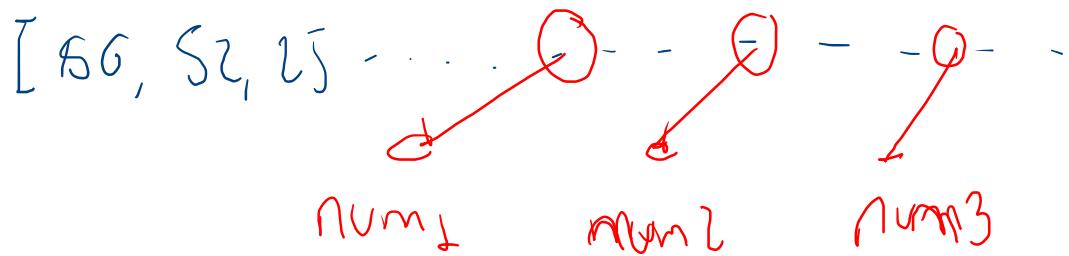


Promedio. { más probable en condic
un registro entre los
35 y 45 años}

Distribución Normal PDF = Función de distribución normal



Edge



10280
98]

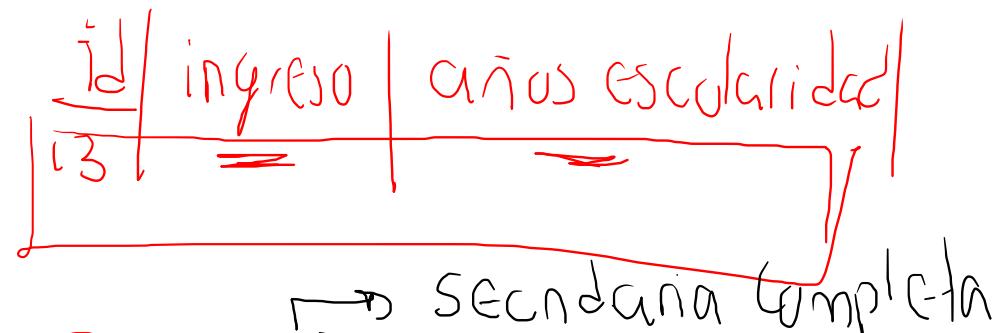
ANÁLISIS Ingreso mensual vs años de escolaridad.

Promedio → Rangos de años de escolaridad.

min(años) max(años)



Ingreso mensual



Secundaria Completa = ~~12~~

medio Superior y Superior = $[12 - 22]$

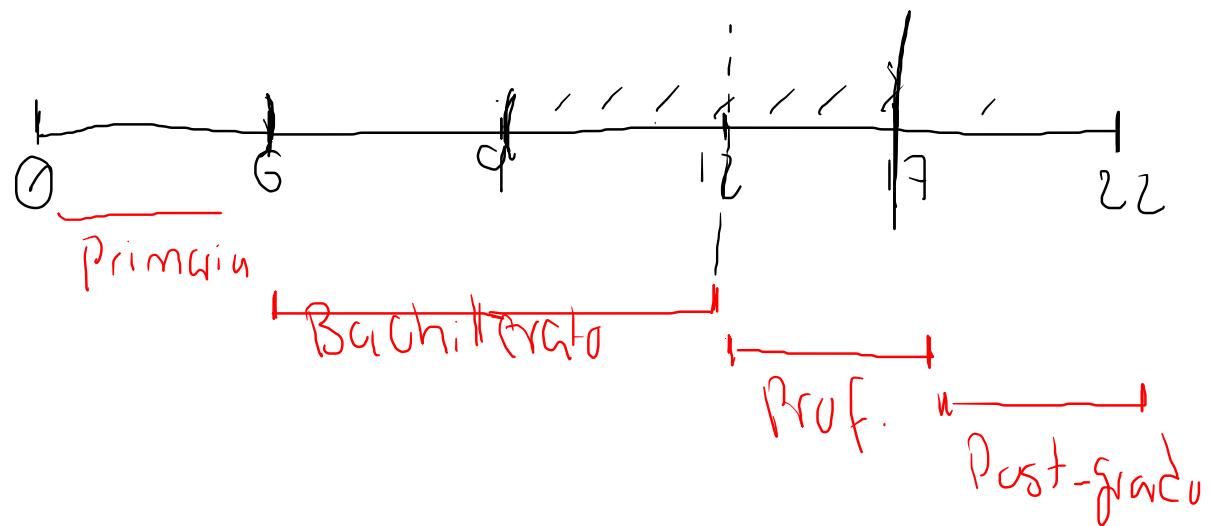
Primaria Completa = \textcircled{G}

0 - 5 : Primaria incompleta

6 : Primaria completa

6 - 9 : Secundaria completa.

9 - 22 : medio Superior y Superior



0 → "Prim" (1)
1 → "Prim" (2)
2 →
3
4
5
6
:
22

Desviación típica.

$$\sigma_1 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

x_i : es el dato

\bar{x} : promedio de la lista donde está el dato

n : cantidad de datos.

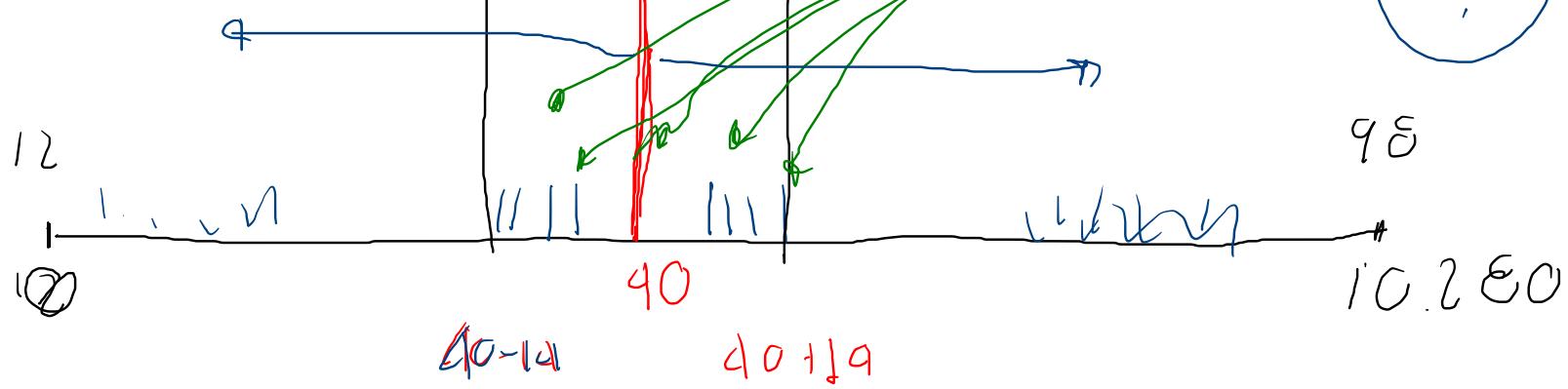
recoger la lista de edad , promedio edad

$$\frac{\sum (x_i - \bar{x})}{\text{len(lista)}}$$

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

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

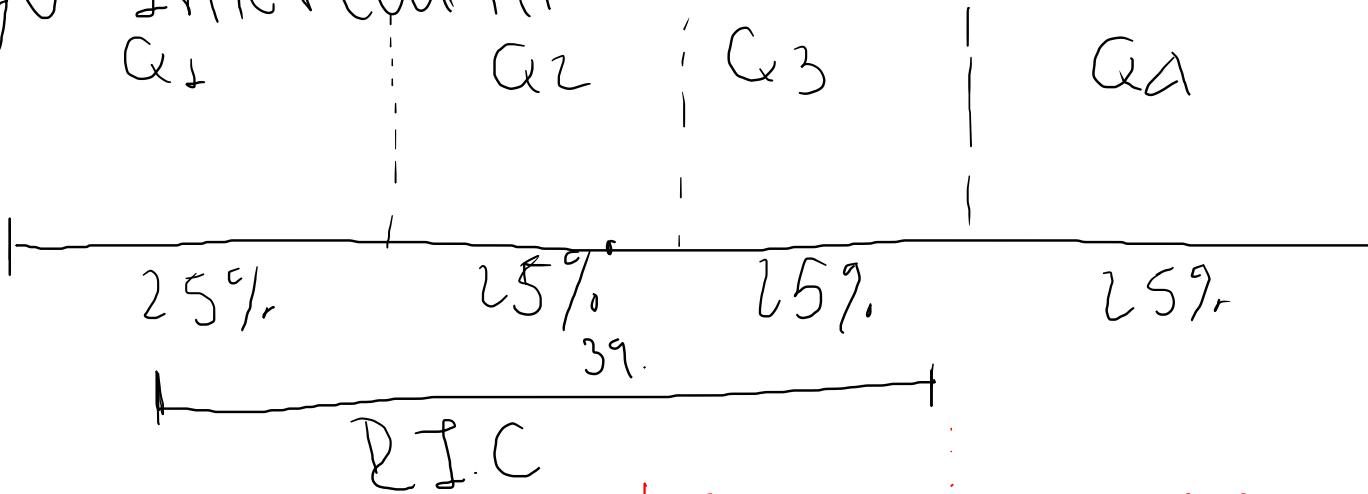
Variando



Probabilidad

TdG

Rango Inter cuartil



$Q_3 - Q_1$

↓ 2 3 4 5 6 7 8 9
6 9 9 10 10 12 12 14 14
I₇ ← : → dcr

$$J_7 = 6$$

$$dcr = 12$$

$$q \rightarrow 9$$

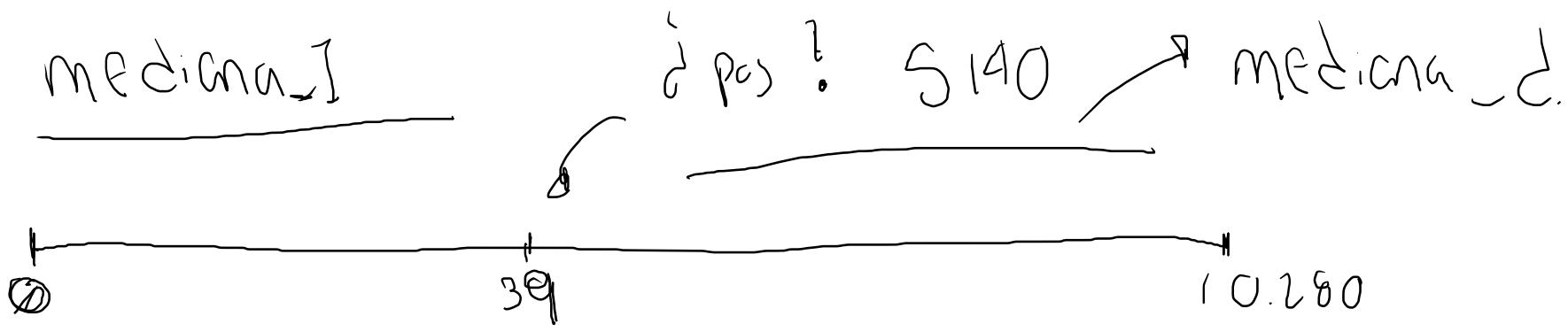
$$12 \rightarrow 13$$

$$10$$

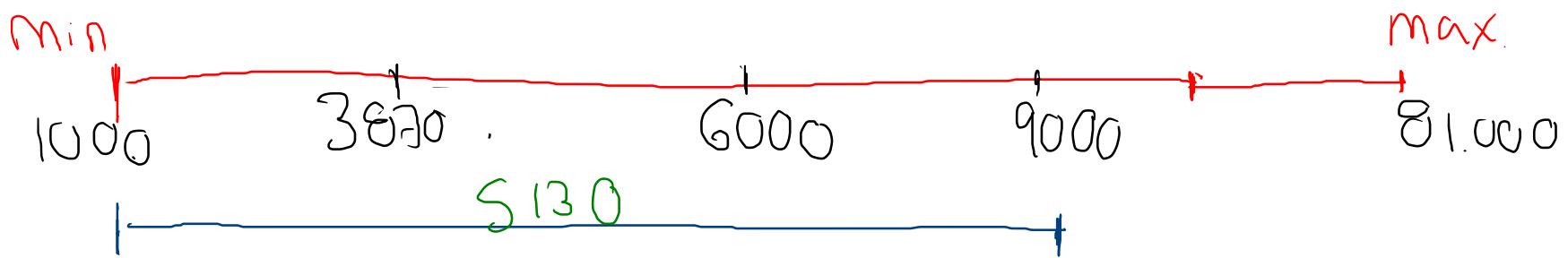
$$14$$

$$R.I.C = 13 - 9 = 4$$

$$\text{Zanjo} = J_7 - q = 8$$



RIC Inyego mensual



Distribución de Frecuencias

Cuántas VECES se repite un dato.

Poder calcular la probabilidad de ocurrir el dato

$$\text{Dato} = \{1, 2, 3, 4, 5, 6\}$$

$$\text{Experimento: } \{6, 6, 1, 2, 3, 4, 1, 2, 4, 5\}$$

Probabilidad de que salga 5

$$P(5) = \frac{3}{10} = 0.3 \quad P(3) = \frac{1}{10}$$

$$P = \frac{d}{m(\text{dato})}$$

$$P = \frac{1}{6} = 0.1666$$

$$P(5) = \frac{1}{10}$$

$$P(2) = \frac{2}{10}$$

FRECUENCIA RELATIVA

$$P(4) = \frac{1}{10}$$

$$P(1) = \frac{2}{10}$$

Frecuencia Absoluta = Las veces que se repite un dato

Frecuencia Relativa = $\frac{F_A}{N} = \frac{\text{Frecuencia absoluta}}{\text{cantidad de datos}}$

Frecuencia Acumulada = $f(6) + f(5) + f(4) + f(3) + f(2) + f(1)$
= Frecuencia absoluta.

[0 - 100%] Porcentaje

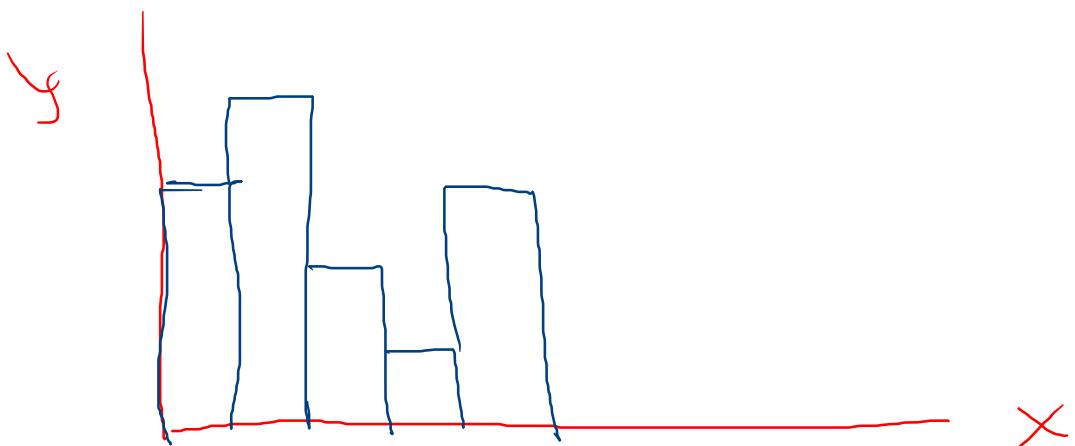
[0 - 1] Número que representa.

Datos Discretos = [1, 4, 3, 2, 5]

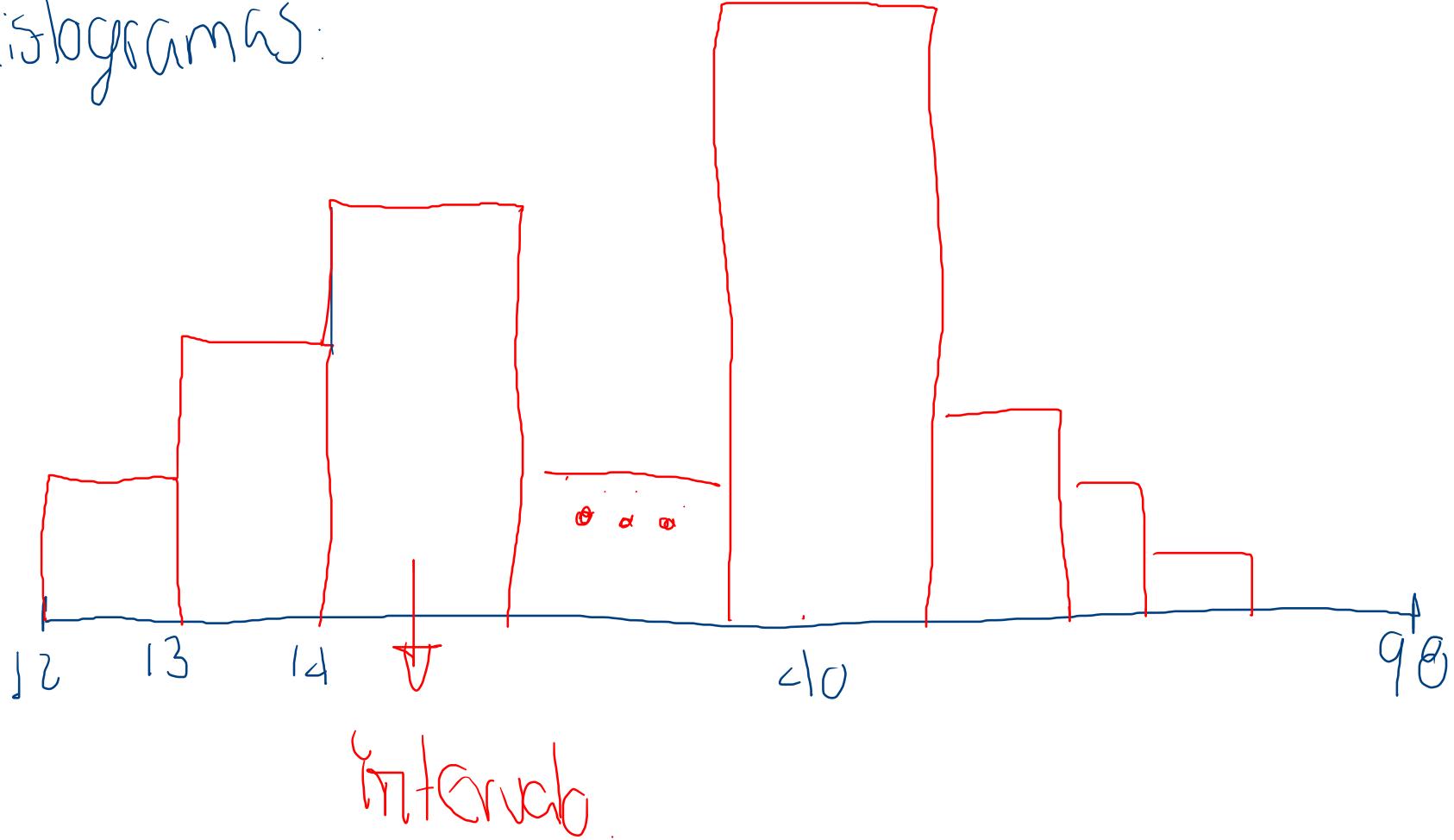
↑

MUESTRA

Grafico de Barras

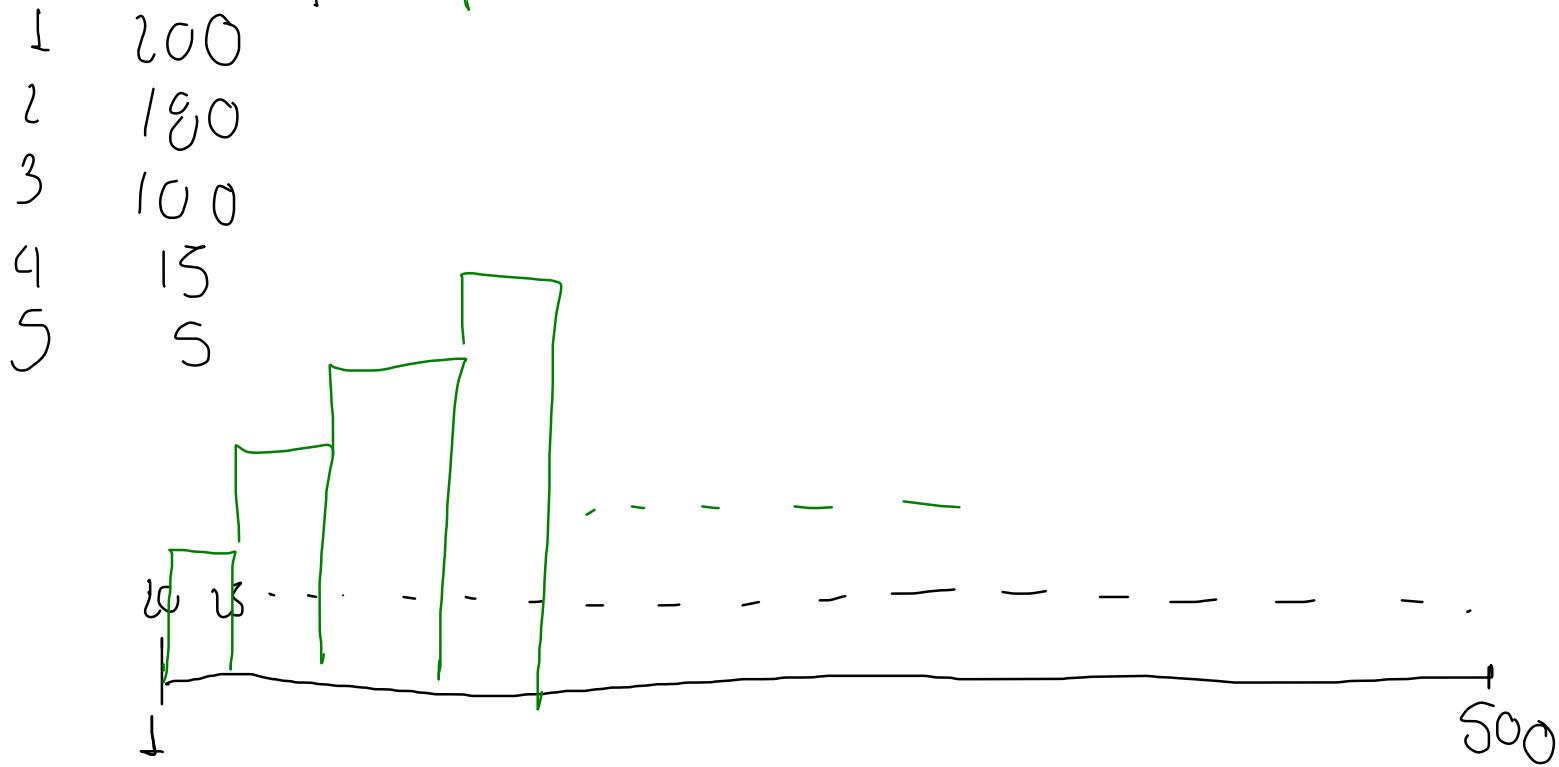


Histograms:

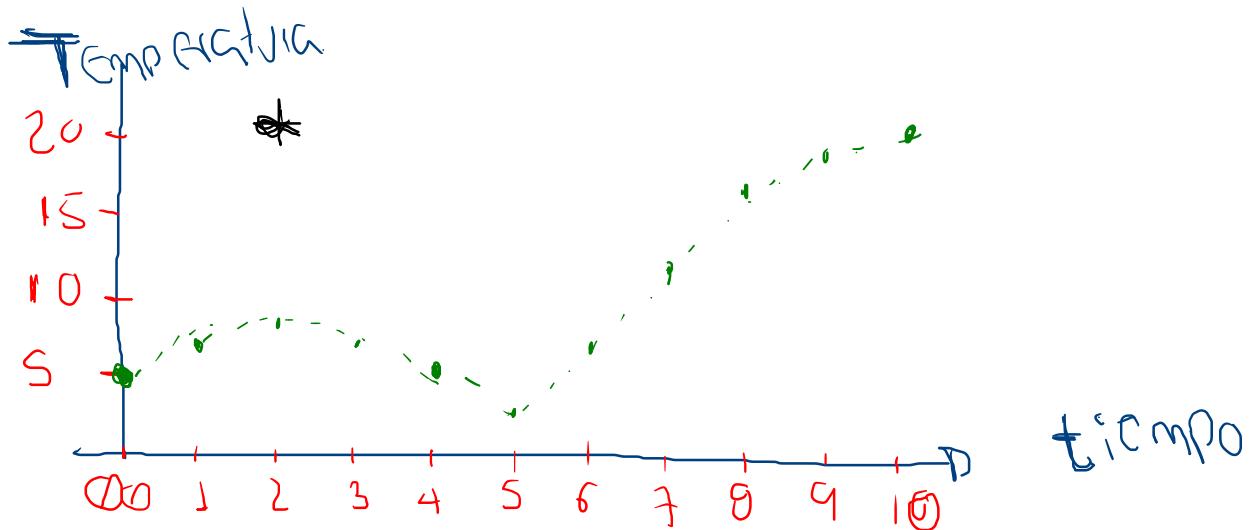
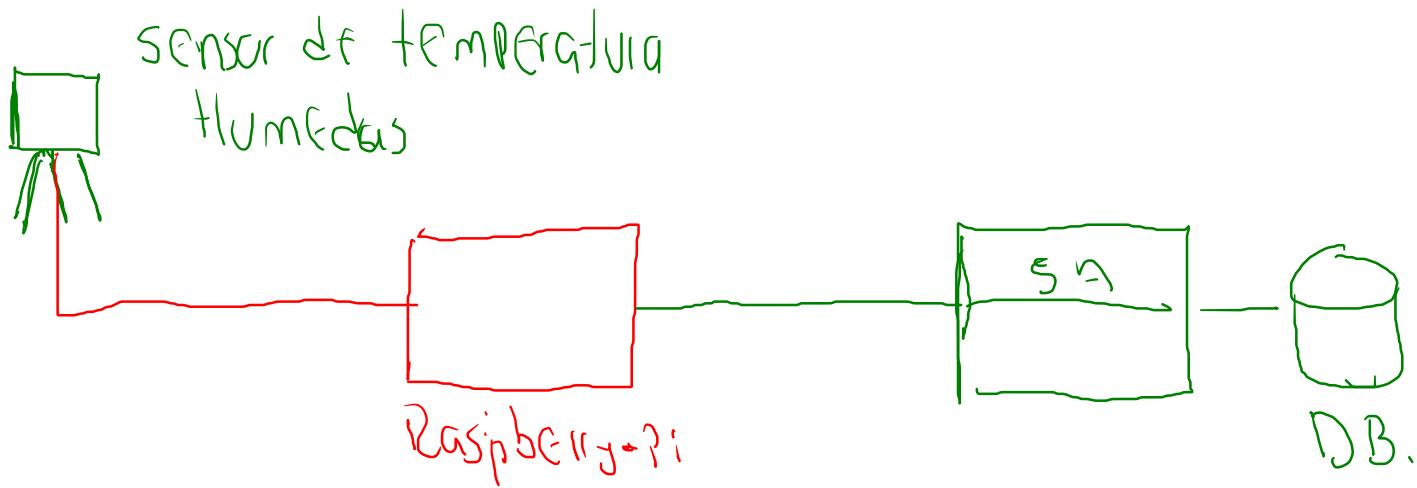


Muestra = 500 Estudiantes

f_i = FRECUENCIA ABSOLUTA



Valores atípicos



Error



Data atípico / Sensor no funciona bien



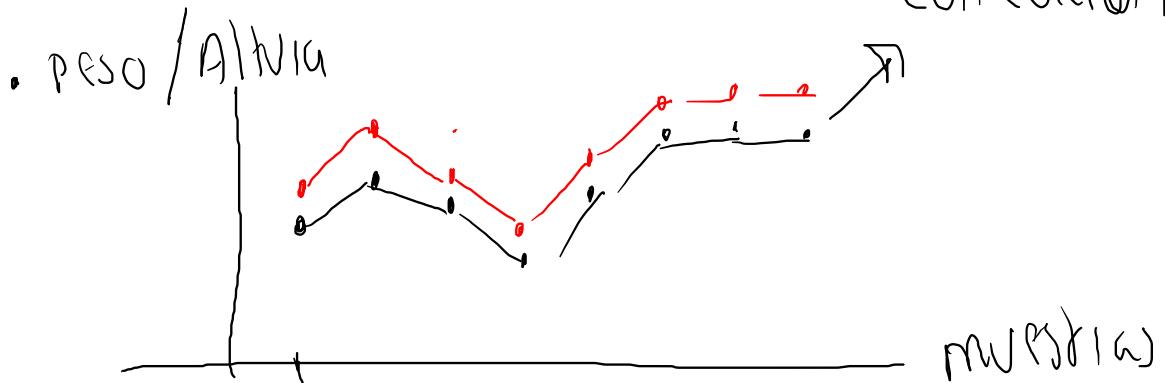
1784 → Error.

Tablas de probabilidad

- Probabilidad INDEPENDIENTE \rightarrow sacar 2 cartas de una baraja
- Probabilidad DEPENDIENTE \rightarrow 1 suceso influye en el otro
P.E.: peso y estatura.



Correlación de variables

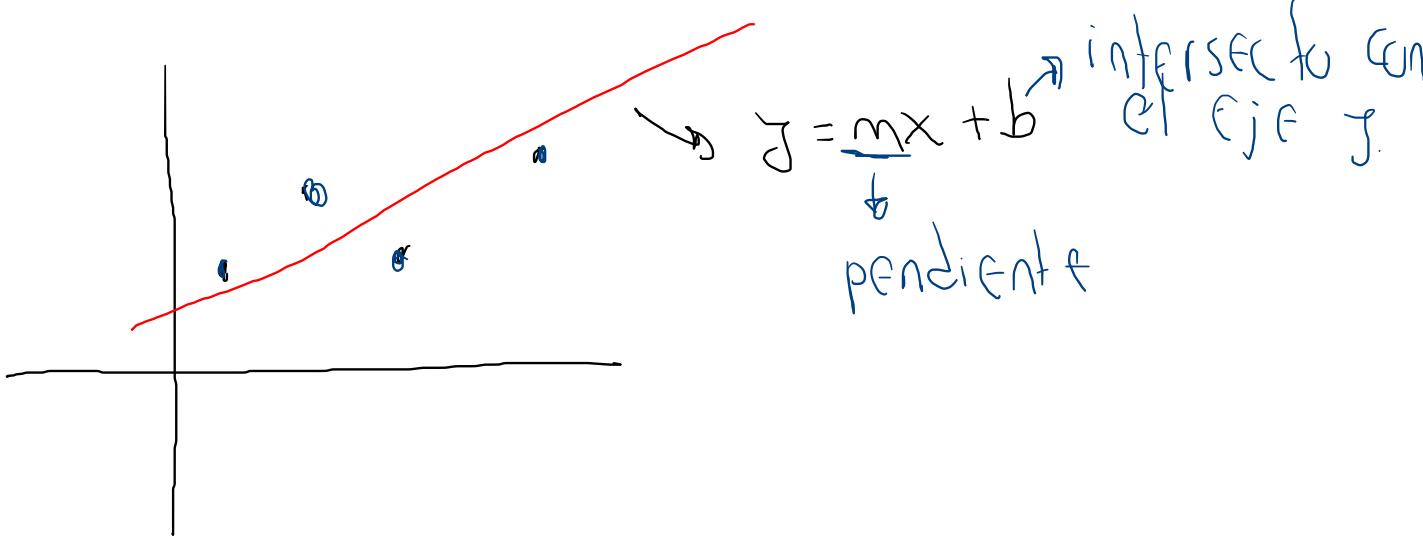


Correlación - (se approx -1)
Correlación + (se approxima a 1)

- ALTURA
- PESO

Regression.

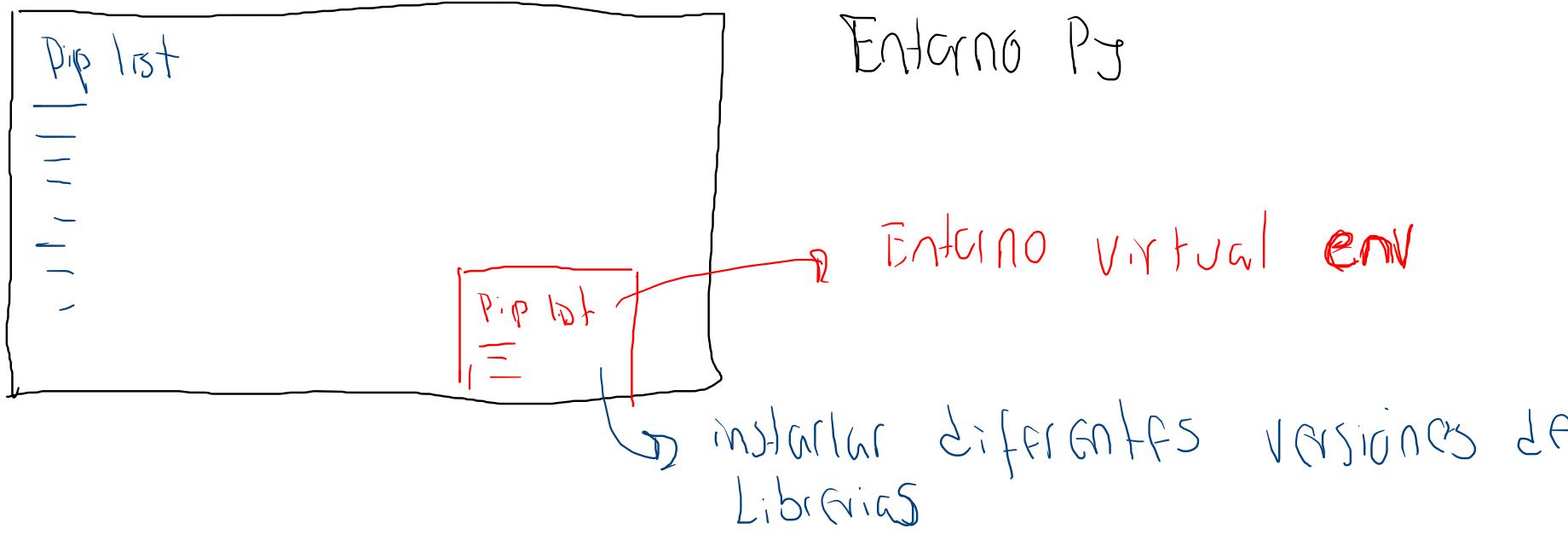
X	y
0.8	10
1.2	18
1.7	15



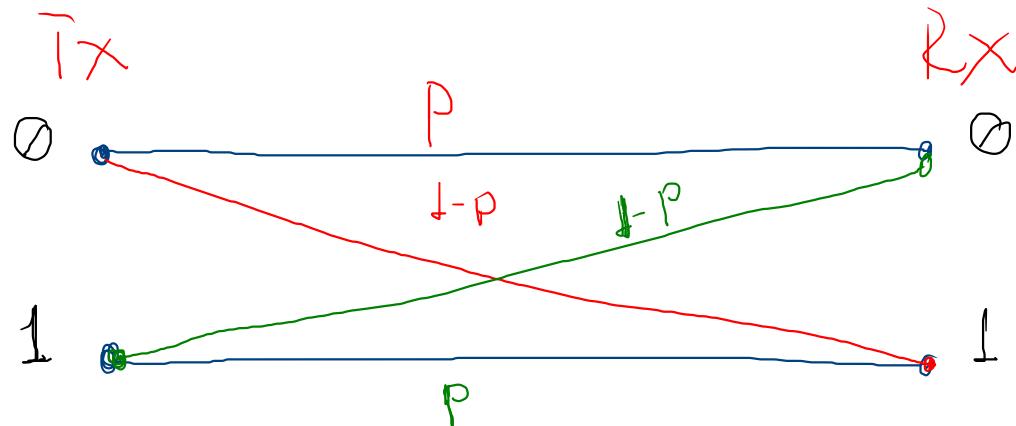
$$y = mx + b$$

intersect to con
el ejf y .
pendiente

Entornos Virtuales con Python



Tx Digital



0101010

$P_0(0/0)$ $P_0(0/1)$ \rightarrow El fstg
 $P_1(1/1)$ $P_1(1/0)$ \rightarrow Alula
Depend del
contrario