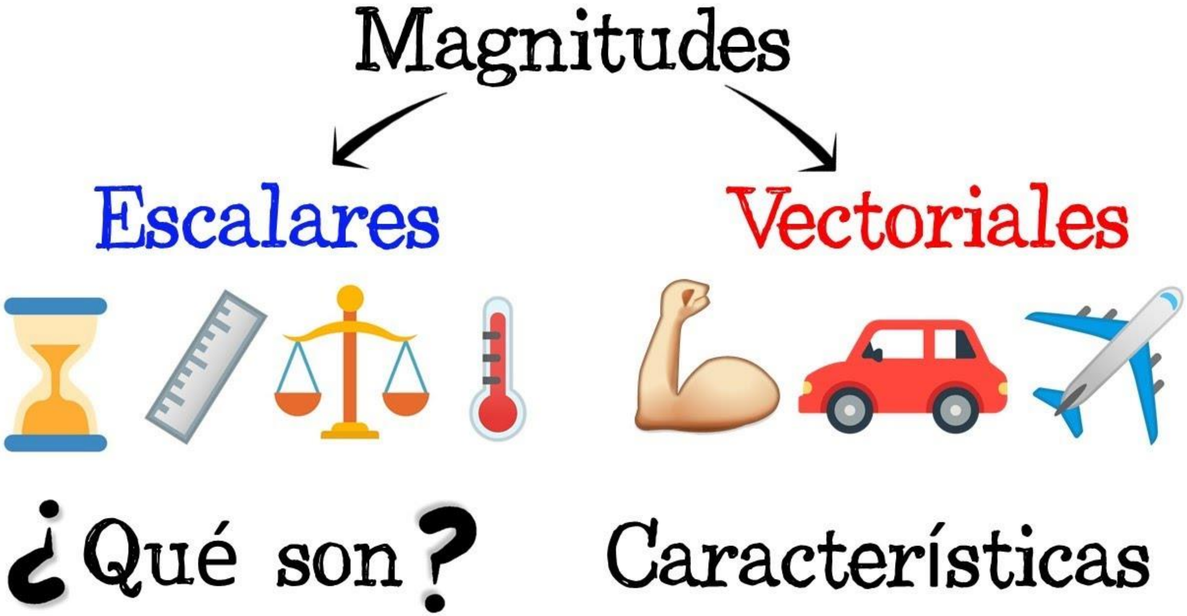
Introducción al análisis de datos con Python
Elementos de matemáticas

Profesor: Carlos Jiménez

Matrices y vectores: Teoría



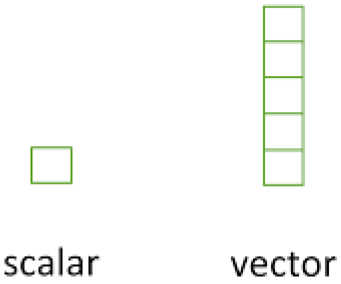
Matrices, vectores y más allá

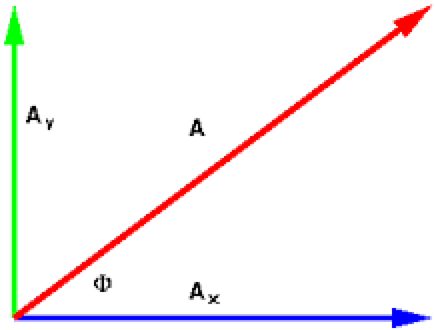
Las variables o datos se guardan en cantidades que en general se llaman tensores. Casos particulares de estos son las matrices, vectores y escalares.

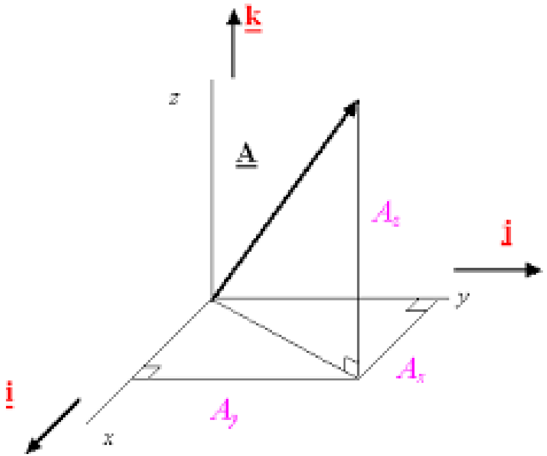
Escalar: Cantidad que requiere de un solo valor para quedar específica

Vector: Requiere más de una cantidad para quedar específico

En 1, 2 y 3 dimensiones los vectores se pueden visualizar y graficar en nuestro espacio

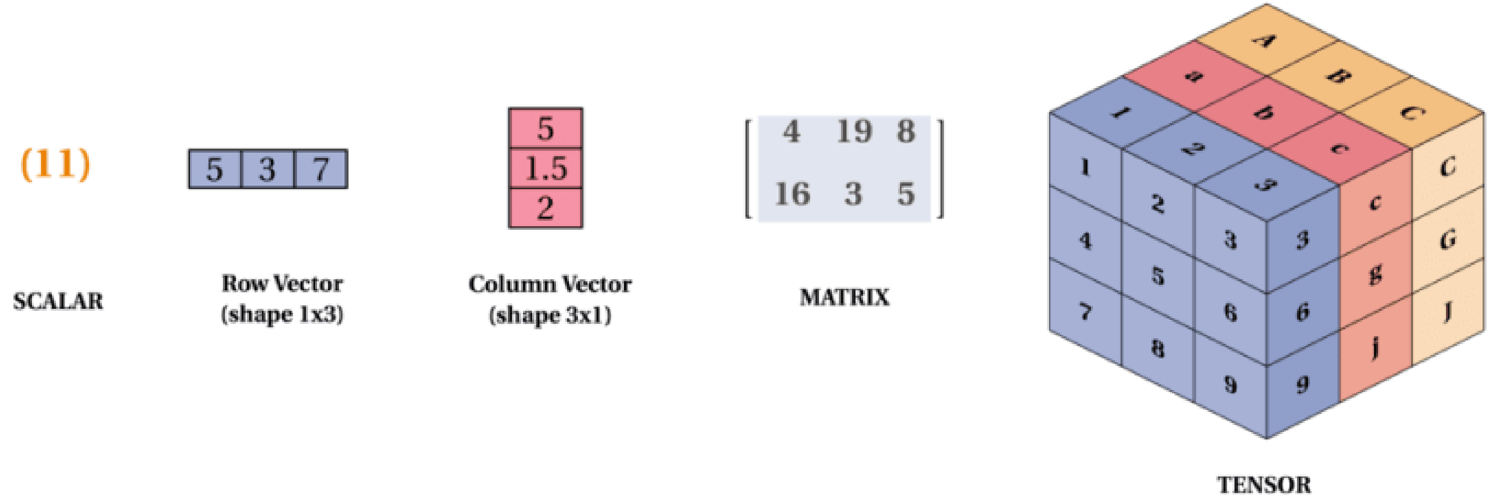






¿Qué es una matriz?

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$


(11)
5 3 7
5
1.5
2
4 19 8
16 3 5
SCALAR
Row Vector
(shape 1x3)
Column Vector
(shape 3x1)
MATRIX
TENSOR


Scalar
Vector
Matrix
Tensor
1
1
2
1 2
3 4
1 2
3 2
1 7
5 4

Suma y resta de matrices

$$A = \begin{pmatrix} 2 & 1 \\ -3 & 6 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 3 \\ 4 & -1 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 2 & 1 \\ -3 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 3 \\ 4 & -1 \end{pmatrix} = \begin{pmatrix} 2+1 & 1+3 \\ -3+4 & 6+(-1) \end{pmatrix} = \begin{pmatrix} \mathbf{3} & \mathbf{4} \\ \mathbf{1} & \mathbf{5} \end{pmatrix}$$

$$A - B = \begin{pmatrix} 2 & 1 \\ -3 & 6 \end{pmatrix} - \begin{pmatrix} 1 & 3 \\ 4 & -1 \end{pmatrix} = \begin{pmatrix} 2-1 & 1-3 \\ -3-4 & 6-(-1) \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{-2} \\ \mathbf{-7} & \mathbf{7} \end{pmatrix}$$

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 0 & 0 \\ 5 & 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 2+1 & 0+0 & 1+1 \\ 3+1 & 0+2 & 0+1 \\ 5+1 & 1+1 & 1+0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 2 \\ 4 & 2 & 1 \\ 6 & 2 & 1 \end{pmatrix}$$

$$A - B = \begin{pmatrix} 2-1 & 0-0 & 1-1 \\ 3-1 & 0-2 & 0-1 \\ 5-1 & 1-1 & 1-0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -2 & -1 \\ 4 & 0 & 1 \end{pmatrix}$$

Matrices

Las matrices se pueden multiplicar, dando como resultado otra matriz. Para ello es necesario que el número de columnas de la primera sea igual al número de filas de la segunda.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 & 1 \\ 2 & 2 & 3 \\ -2 & -3 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 4 \\ 0 & 2 & 5 \\ 1 & 3 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} 1+0+1 & 0-2+3 & 4-5+0 \\ 2+0+3 & 0+4+9 & 8+10+0 \\ -2+0-1 & 0-6-3 & -8-15+0 \end{pmatrix} =$$

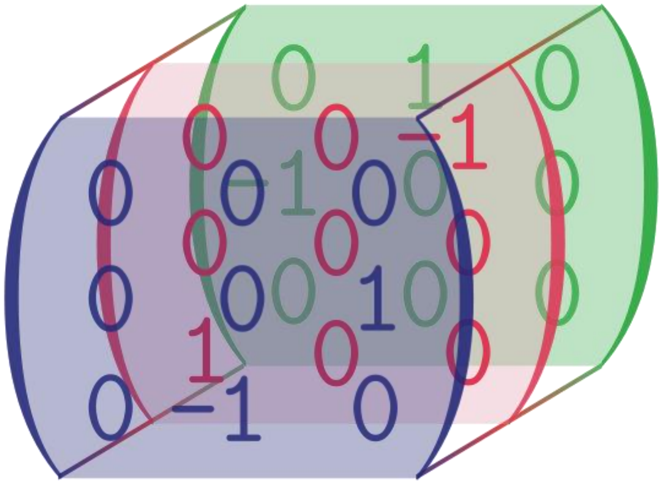
$$= \begin{pmatrix} 2 & 1 & -1 \\ 5 & 13 & 18 \\ -3 & -9 & -23 \end{pmatrix}$$

Tensor: Arrreglo de arreglos

Siguiendo en esta misma línea, existe otro objeto matemático más allá de las matrices llamado tensor, En este curso no entraremos en detalles técnicos acerca de las características de un tensor, pero este contiene mucha más información, y como caso particular contiene a los escalares, vectores y matrices. Todas estas cantidades se pueden representar por medio de subindices y superindices

Ejemplo: Tensor de Levi-Civita

$$\epsilon_{\mathrm{ijk}} =$$



$$\epsilon_{ijk} = \begin{cases} +1 & \text{si } (i,j,k) \text{ es } (1,2,3), (2,3,1) \text{ o } (3,1,2) \\ -1 & \text{si } (i,j,k) \text{ es } (3,2,1), (1,3,2) \text{ o } (2,1,3) \\ 0 & \text{de otro modo } i=j \text{ o } j=k \text{ o } k=i \end{cases}$$

Debido a sus propiedades, los “array” resultan mucho más útiles para construir matrices cuando programamos, pues las listas de listas no se dejan operar muy bien.

Suma y multiplicación de matrices en python

```
from matplotlib import pyplot as plt
import numpy as np

#Multiplicacion y suma
z = np.array([[1, 3], [-1, 10]])
w = np.array([[2, 4], [1, -3]])
print("z es: ")
print(z)
print("w es: ")
print(w)

suma = w + z
print("w + z es: ")
print(suma)

result = np.dot(z,w)
print("z por w es: ")
print(result)
```

```
z es:
[[ 1  3]
 [-1 10]]
w es:
[[ 2  4]
 [ 1 -3]]
w + z es:
[[3 7]
 [0 7]]
z por w es:
[[  5  -5]
 [  8 -34]]
```

Creación y modificación de matrices, desde luego lo mismo aplica para vectores

```python
# Construccion y modificación de matrices

z = np.zeros((3,4))
print(z)
for i in range(3):
    for j in range(4):
        z[i][j] = i**2-j

print("z modificado es: ")
print(z)
print("El elemento z[1,3] es: ", z[1,3])
print(z[1][3])
```

```
[[0. 0. 0. 0.]
 [0. 0. 0. 0.]
 [0. 0. 0. 0.]]
z modificado es:
[[ 0. -1. -2. -3.]
 [ 1.  0. -1. -2.]
 [ 4.  3.  2.  1.]]
El elemento z[1,3] es:  -2.0
-2.0
```

Mas operaciones

```python
import numpy as np
import pandas as pd
from numpy import linalg as LA

m1 = np.array([[1, 3],[4, -2]])
m2 = np.array([[-1, 0],[5, 3]])


#Transpuesta
t = m1.transpose()
print("La matriz es: ")
print(m1)
print("Y su transpuesta es: ")
print(t)

# Autovectores y autovalores
print("La matriz es: ")
print(m1)
eigenval, eigenvec = LA.eig(m1)
print("Sus autovalores son: ")
print(eigenval)
print("Sus autovectores son: ")
print(eigenvec)
```

```
La matriz es:
[[ 1  3]
 [ 4 -2]]
Y su transpuesta es:
[[ 1  4]
 [ 3 -2]]
La matriz es:
[[ 1  3]
 [ 4 -2]]
Sus autovalores son:
[ 3.27491722 -4.27491722]
Sus autovectores son:
[[ 0.79681209 -0.49436913]
 [ 0.60422718  0.86925207]]
```

Para los tensores aplica lo mismo, solo que es más difícil (para un humano) visualizar los elementos

```
w = np.ones((3,4,2))
print(w)
for i in range(3):
    for j in range(4):
        for k in range(2):
            w[i][j] = i**2 - j + 2*k

print("w modificado es: ")
print(w)
print("El elemento w[2,0,1] es: ", w[2][0][1])
```

```
[[[1. 1.]
  [1. 1.]
  [1. 1.]
  [1. 1.]]

 [[1. 1.]
  [1. 1.]
  [1. 1.]
  [1. 1.]]

 [[1. 1.]
  [1. 1.]
  [1. 1.]
  [1. 1.]]]
w modificado es:
[[[ 2.  2.]
  [ 1.  1.]
  [ 0.  0.]
  [-1. -1.]]

 [[ 3.  3.]
  [ 2.  2.]
  [ 1.  1.]
  [ 0.  0.]]

 [[ 6.  6.]
  [ 5.  5.]
  [ 4.  4.]
  [ 3.  3.]]]
El elemento w[2,0,1] es:  6.0
```

Dada una matriz se puede tomar determinado elementos de ella con el recurso " : "

```python
import numpy as np

v = np.array(([1, 2, 3, 4],
      [5 ,6, 7, 8],
      [9, 10, 11, 12],
      [13, 14, 15, 16]))
print("La matriz v es: ")
print(v)
print("La matriz v sin la primer línea es: ")
print(v[1:, :])

print("La matriz v sin la última línea es: ")
print(v[:-1, :])

print("La matriz v desde la columna 1 hasta la columna 2 es: ")
print(v[:, 1:3])

print("La matriz reducida de v es: ")
print(v[1:3, 1:3])
```

```
La matriz v es:
[[ 1  2  3  4]
 [ 5  6  7  8]
 [ 9 10 11 12]
 [13 14 15 16]]
La matriz v sin la primer línea es:
[[ 5  6  7  8]
 [ 9 10 11 12]
 [13 14 15 16]]
La matriz v sin la última línea es:
[[ 1  2  3  4]
 [ 5  6  7  8]
 [ 9 10 11 12]]
La matriz v desde la columna 1 hasta la columna 2 es:
[[ 2  3]
 [ 6  7]
 [10 11]
 [14 15]]
La matriz reducida de v es:
[[ 6  7]
 [10 11]]
```

Podemos operar matrices cambiando, es bueno hacer una copia para no perder los datos, este proceso se podria hacer tambien por componentes, pero resulta más complicado y costoso, como lo veremos más adelante

```python
u = np.ones((5,5))
u[1:4,1:4] = 2
u[2:3,2:3] = 3
un = u.copy()
print("La matriz un es: ")
print(un)

u[1:,1:] = un[1:,1:] - un[:-1,1:]

print("La matriz u modificada es: ")
print(u)
```

```
La matriz un es:
[[1. 1. 1. 1. 1.]
 [1. 2. 2. 2. 1.]
 [1. 2. 3. 2. 1.]
 [1. 2. 2. 2. 1.]
 [1. 1. 1. 1. 1.]]
La matriz u modificada es:
[[ 1.  1.  1.  1.  1.]
 [ 1.  1.  1.  1.  0.]
 [ 1.  0.  1.  0.  0.]
 [ 1.  0. -1.  0.  0.]
 [ 1. -1. -1. -1.  0.]]
```

```
un[1:,1:]                 un[:-1,1:]                  u [1:,1:]

[[1. 1. 1. 1. 1.]         [[1. 1. 1. 1. 1.]           [[ 1.  1.  1.  1.  1.]
 [1. 2. 2. 2. 1.]     -    [1. 2. 2. 2. 1.]     =      [ 1.  1.  1.  1.  0.]
 [1. 2. 3. 2. 1.]          [1. 2. 3. 2. 1.]            [ 1.  0.  1.  0.  0.]
 [1. 2. 2. 2. 1.]          [1. 2. 2. 2. 1.]            [ 1.  0. -1.  0.  0.]
 [1. 1. 1. 1. 1.]]         [1. 1. 1. 1. 1.]]           [ 1. -1. -1. -1.  0.]]
```

Probabilidad

La probabilidad es una medida de la certidumbre de que ocurra un evento. Su valor es un número entre 0 y 1, donde un evento imposible corresponde a cero y uno seguro corresponde a uno.

Experimento: Una operación que puede producir algunos resultados bien definidos pero que no se puede predecir cuál de ellos se obtendrá, se llama un experimento aleatorio

Espacio muestral: Todos los resultados posibles de un experimento en su conjunto, forman el Espacio de la muestra.

Resultado: Cualquier elemento posible del espacio muestral S de un experimento aleatorio se llama Resultado.

Suceso: Cualquier subconjunto del espacio muestral S se llama un Evento (denotado por E). Cuando se produce un resultado que pertenece al subconjunto E, se dice que ha ocurrido un suceso.

Ejemplo:

Experimento: Lanzar una moneda al aire

Espacio muestral: { cara, sello}

Suceso: Caer cara

Resultado: Cae sello

No ocurrio el evento

La probabilidad la denotamos con la letra P

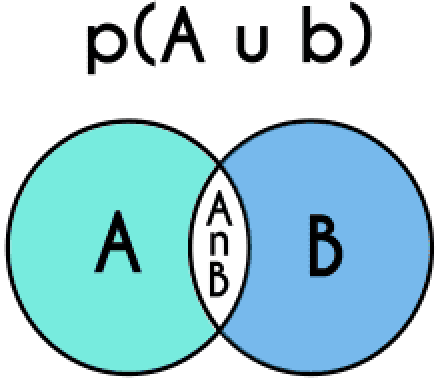
Sea $P(A)$ la probabilidad de que ocurra el evento A, entonces la probabilidad de que no ocurra A, $P(NA)$, es:

$$P(NA) = 1 - P(A)$$

Regla de la adición

Si dos eventos A y B son mutuamente excluyentes (Independientes)
Entonces la probabilidad de que caigan uno o el otro es:

P(A U B) = P(A) + P(B)

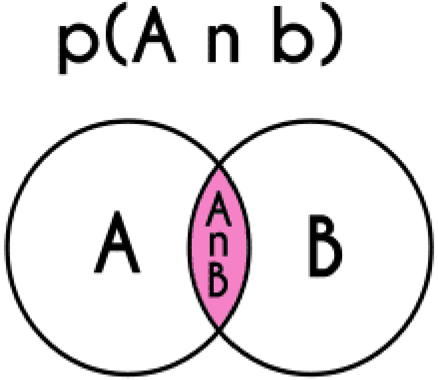


Dos eventos A y B son dependientes, si la ocurrencia de uno de ellos afecta la ocurrencia del otro. Para eventos dependientes, la regla de la multiplicación establece que:

$$P(A \cap B) = P(A) \times P(B|A)$$

$$\mathrm{P}(A \cap B) = P(B) \times P(A|B)$$

siendo $P(B|A)$ la probabilidad de que ocurra B habiéndose dado o verificado el evento A.



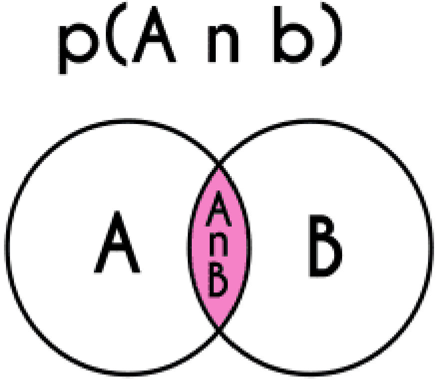
Eventos independientes

Dos eventos A y B son independientes, si la ocurrencia de uno de ellos no afecta la ocurrencia del otro, es decir, cuando **los eventos A y B no están relacionados**. Para eventos independientes, la regla de la multiplicación establece que:

$$P(A \cap B) = P(A) \times P(B)$$

Esto se debe, a que en los eventos independientes, la ocurrencia de un evento, no afecta a la ocurrencia del otro:

$$P(A|B) = P(A) \quad \wedge \quad P(B|A) = P(B)$$



Probabilidad condicionada

Sea un espacio probabilístico y un suceso **B** perteneciente al *Algebra de Boole*, tal que $P(B) \neq 0$, entonces se define la probabilidad de que ocurra **A** si antes ha ocurrido **B**, como:

$$P(A/B) = \frac{\mathbf{P}(\mathbf{A} \cap \mathbf{B})}{\mathbf{P}(B)} \qquad P(B) \neq 0,$$

Regla de Laplace

La Regla de Laplace establece que:

La probabilidad de ocurrencia de un suceso imposible es 0.
La probabilidad de ocurrencia de un suceso seguro es 1, es decir, P(A)=1.
Para aplicar la regla de Laplace es necesario que los experimentos den lugar a sucesos equiprobables, es decir, que todos tengan o posean la misma probabilidad.

La probabilidad de que ocurra un suceso se calcula así:

$$P(A) = \frac{N^{\underline{o}}\ de\ casos\ favorables}{N^{\underline{o}}\ de\ casos\ totales}$$

Ley de los Grandes Números.

La Ley de los Grandes Números nos dice que si no conocemos la probabilidad de un suceso en un experimento aleatorio, debemos hacer tantas veces el experimento que al hacer un estudio estadístico de los resultados, la frecuencia relativa de cada suceso llega un momento que se estabiliza.

Ejemplo:

Se lanza una sola vez un dado,

a) encuentre la probabilidad de que caiga 3
b) Encuentre la probabilidad de que caiga 5 o 6
c) Encuentre la probabilidad de que caiga 5 y 6
d) Encuentre la probabilidad de que caiga uno de estos valores, 1, 2, 3, 4, 5, o 6

Ejemplo:

Se lanza una sola vez un dado,

a) encuentre la probabilidad de que caiga 3
b) Encuentre la probabilidad de que caiga 5 o 6
c) Encuentre la probabilidad de que caiga 5 y 6
d) Encuentre la probabilidad de que caiga uno de estos valores, 1, 2, 3, 4, 5, o 6

E = {1, 2, 3, 4, 5, 6} espacio muestral, tamaño=6

a) p(3)= 1/6

b) p(5o6) = p(5) + p(6) = 2/6

c) p(5 y 6) = 0

d) p(All) = 1

<u>Ejemplo:</u> Sabiendo que al lanzar un dado ha salido un número par, hallar la probabilidad que este número haya sido un dos:

Ejemplo: Sabiendo que al lanzar un dado ha salido un número par, hallar la probabilidad que este número haya sido un dos:

A = {2}   B = {2, 4, 6} $A \cap B$ = {2}

$$P(A \cap B) = \frac{1}{6} \qquad P(B) = \frac{3}{6}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Ejemplo: De una urna que contiene 9 bolas rojas y 5 negras, se extraen sucesivamente 2 bolas. Calcular la probabilidad de los siguientes sucesos:

a) Que las dos sean negras
b) Que las dos sean rojas
c) Que la primera se roja y la segunda negra
d) Que la segunda se roja sabiendo que la primera fue negra

La solución en cada apartado es la siguiente.

a) Sea N_1: Sacar la 1ª Negra
N_2: Sacar la 2ª Negra

$$P(N_1 \cap N_2) = P(N_1) \cdot P(N_2/N_1)$$
$$= 5/14 \cdot 4/13$$

b) Sea R1: Sacar la 1ª Roja
R_2: Sacar la 2ª Roja

$$P(R_1 \cap R_2) = P(R1) \cdot P(R_2/R_1) = 9/14 \cdot 8/13$$

c) Sea R_1: Sacar la 1ª Roja
N_2: Sacar la 2ª Negra

$$P(R_1 \cap N_2) = P(R_1) \cdot P(N_2/R_1) = 9/14 \cdot 5/13$$

d) Sea N_1: La 1ª es Negra
R_2: La 2ª es Roja

$P(R_2/N_1) = 9/13$ (quedan 13 bolas de las cuales 9 son rojas).

En Python la librería random permite generar número pseudoaleatorios de múltiples maneras

```
import numpy as np
import random
```

```
random.random()
```

```
0.4725058944595895 7
```

```
random.randrange(2, 10, 3)
```

```
5
```

```
random.randint(2, 12)
```

```
4
```

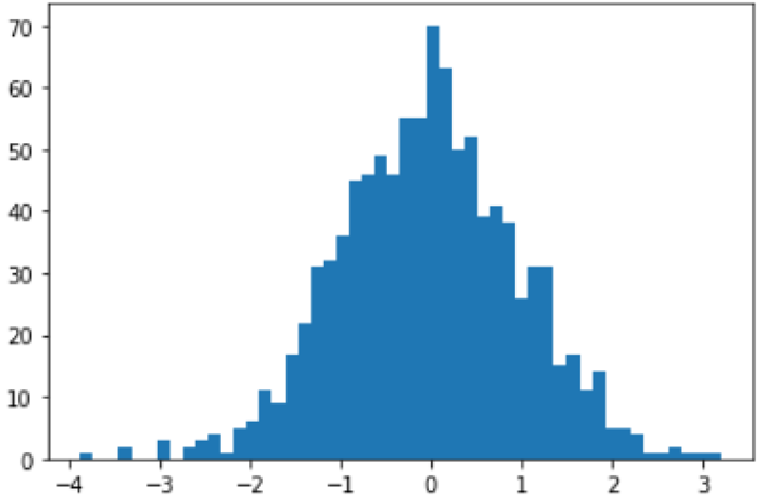
```
L = ["Manzana", "Naranja", 5, 10, "Pera"]
random.choice(L)
```

```
10
```

```
#help(random)
```

```
x = []
for i in range(1000):
    x.append(random.gauss(mu=0.0, sigma=1.0))
```

```
plt.hist(x, bins=50)
plt.show()
```

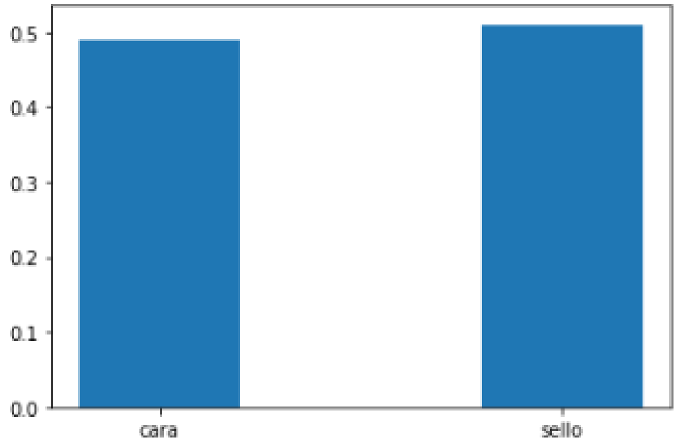


Ejemplo:
Simulemos el lanzamiento de una moneda

Ejemplo:
Simulemos el lanzamiento de una moneda

```
import numpy as np
import random
import matplotlib.pyplot as plt
```

```
N = 100
result = {"cara":0, "sello":0}
for i in range(N):
    x = random.choice(["c", "s"])
    if x == "c":
        result["cara"] = result["cara"] + 1/N
    else:
        result["sello"] = result["sello"] + 1/N

result["cara"] = round(result["cara"], 2)
result["sello"] = round(result["sello"], 2)
plt.bar(list(result.keys()), list(result.values()), width=0.4)
plt.show()
```



Reto:

Simula el resultado en el lanzamiento de dos dados:

a) Para 1000 lanzamientos haz una gráfica de las probabilidades. ¿Cuál es la probabilidad de sacar como resultado un 5?

Reto:

Simula el resultado en el lanzamiento de dos dados:

a) Para 1000 lanzamientos haz una gráfica de las probabilidades. ¿Cuál es la probabilidad de sacar como resultado un 5?

b) Supongamos que los dados están cargados de tal manera que sacar un 6 en cada dado es dos veces mas probable que sacar otro valor. ¿Cómo se ven las probabilidades?

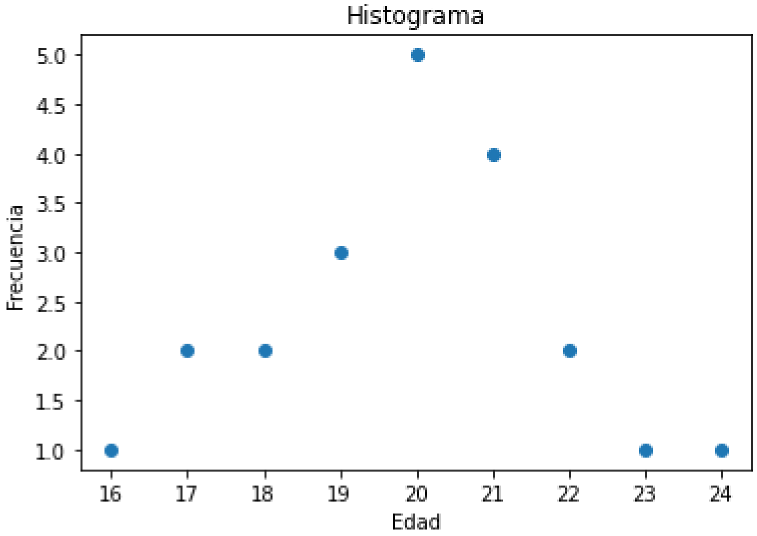
Estadística

La estadística es una ciencia que estudia la variabilidad, recolección, organización, análisis, interpretación, y presentación de los datos, así como el proceso aleatorio que los genera siguiendo las leyes de la probabilidad

Cuando tenemos un conjunto de datos de una sola variable se puede hacer un estudio del mismo por medio de su estadística. Los datos se agrupan por sus frecuencias y se realiza un histograma. En el ejemplo hacemos un análisis de la edad de los estudiantes del curso

Edad
19
18
17
17
19
21
22
18
16
20
21
22
24
21
20
20
23
21
19
20
20

Edad	Frecuencia
16	1
17	2
18	2
19	3
20	5
21	4
22	2
23	1
24	1



Para conjunto de datos más grandes y que estén relacionados se puede notar que estos oscilan alrededor de un valor central. Por ejemplo, si vamos a una tienda y preguntamos al azar por el precio de 20 artículos, estos no tienen realmente una relación, pero si vamos a 20 tiendas y preguntamos por el precio de una libra de arroz, si existe una relación.

¿Crees que hay una relación entre las edades de los estudiantes de este curso?

¿Qué sucede si incluimos la edad del profesor?

La varianza (y la desviación estandar) es una medida de la relación existente entre un conjunto de datos que representan una variable.

La varianza es el cuadrado de la desviación, pero las dos representan una medida de dispersión

Varianza
Desviación Estándar

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

¿Qué es la media?

La media, también conocida como promedio, es el valor que se obtiene al dividir la suma de un conglomerado de números entre la cantidad de ellos.

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Ventas
10
7
4
6
8
10
10
9

$$\overline{X} = \frac{10+7+4+6+8+10+10+9}{8}$$

$$\overline{X} = \frac{64}{8}$$

$$\overline{X} = 8$$

¿Qué es la mediana?

La mediana es un conjunto es un valor que se encuentra a la mitad de los otros valores, es decir, que al ordenar los número de menor a mayor, éste se encuentra justamente en medio entre los que están por arriba.

Ejemplo de Mediana

- La cantidad de valores es impar

Si se tienen los valores: **9,5,4,2,7**, se ordenan: **2, 4, 5, 7, 9**. El elemento de en medio es el **5**, ya que se encuentra dos valores por encima y dos valores por debajo.

- La cantidad de valores es par

Si se tienen los valores **9,5,4,2**, se ordenan: **2,4,5,9**. En este caso se toman los dos valores centrales **5** y **4**, la mediana es el promedio de ambos:

¿Qué es la moda?

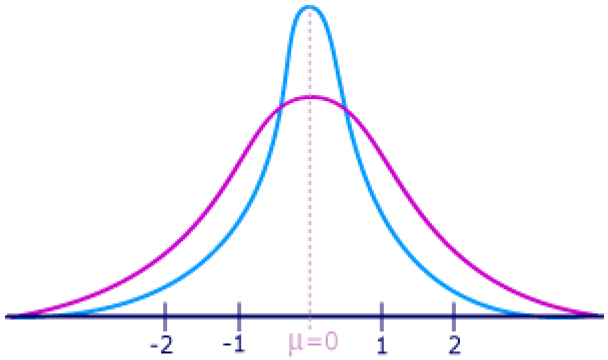
La moda es el valor que aparece más dentro de un conglomerado. En un grupo puede haber dos modas y se conoce como bimodal, y más de dos modas o multimodal cuando se repiten más de dos valores; se llama amodal cuando en un conglomerado no se repiten los valores.

Moda	2,5,5,7,9,10	$\hat{M} =$ 5
Bimodal	2,3,3,5,7,8,9,9	$\hat{M} =$ 3,9
Multimodal	2,3,3,5,7,7,8,9,9	$\hat{M} =$ 3,7,9
Amodal	2,4,5,7,9	$\hat{M} =$ No

La desviación estándar σ nos da información de que tan dispersos están los datos desde el valor promedio, de alguna manera nos habla de la relación entre los datos. La varianza es el cuadrado de la desviación

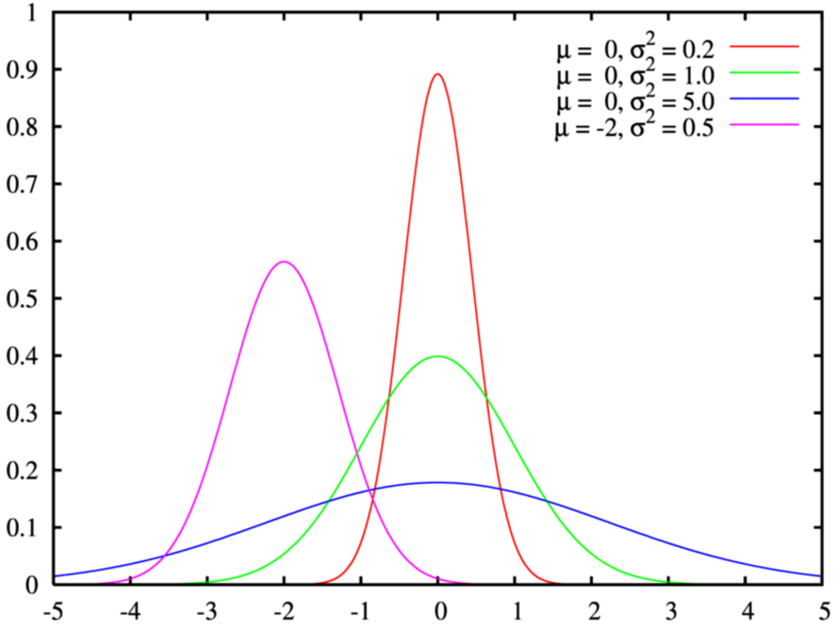
$$\sigma = \sqrt{\frac{\sum_{k=1}^{n} (d_k)^2}{n-1}}$$

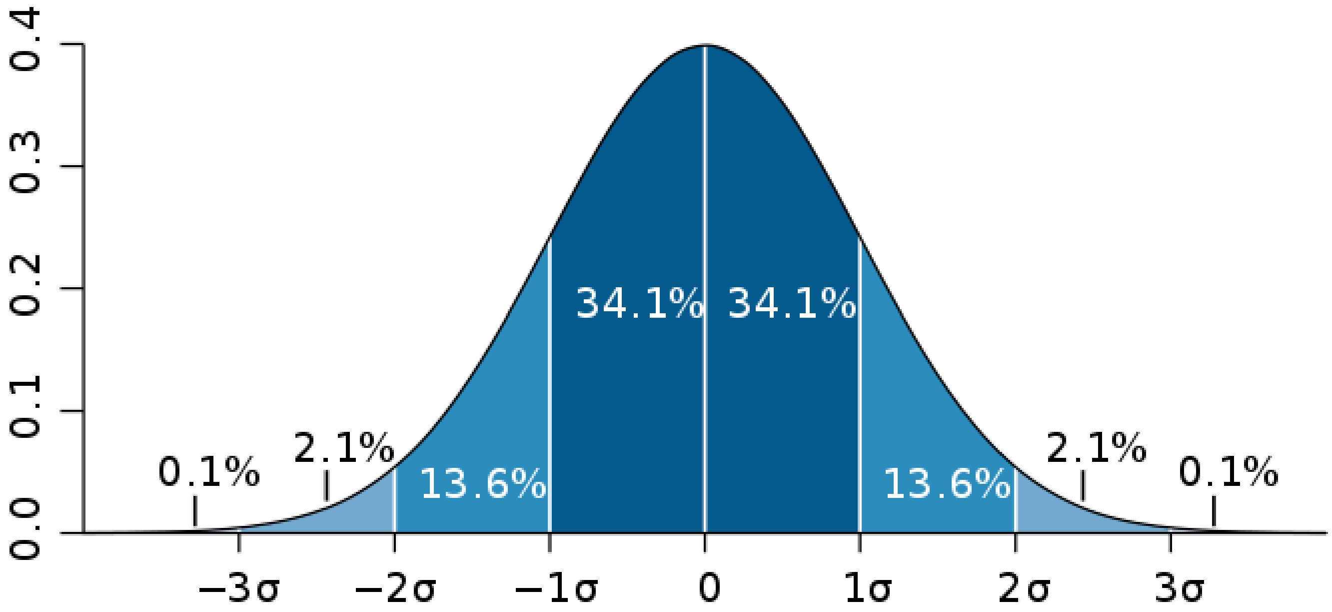
$$d_k = \overline{X} - X_k$$



En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria. También puede decirse que tiene una relación estrecha con las distribuciones de frecuencia. Entre las distribuciones de frecuencia más conocidas está la distribución normal o de gauss

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$





La desviación estandar encaja un cuadrado en la distribución normal. Aproximadamente un 68.2% de los datos se ubican entre el valor medio y mas y menos la desviación estandar.

Una empresa fabrica clavos de una misma referencia en una maquina (no perfecta). La longitud de una muestra de los clavos se muestra en la tabla. Esta empresa vende en distintas ciudades, el número de cajas vendidas en mayo se muestra en la tabla. Haga un análisis de los datos

Longitud (mm)
12,1
12,3
1,3
11,8
11,4
11,6

Ciudad	Ventas (cajas)
Bogotá	120
Medellin	80
Cali	10
Cucuta	300
Leticia	20
Ibague	1200

Ejercicio

Para el conjunto de datos que se muestran, haga un histograma,

[2, 10, -4, 2, 6, 2, 4, 6, 8, 6, 4, -2, 4, 0, 4, 8, 0, 10, 4, 4]

- Determine el valor de la media y la desviación estándar
- Normalice los datos en términos de probabilidades, calcule nuevamente la desviación y la media y haga una gráfica de una distribución normal

Ajuste lineal de un conjunto de dos datos.

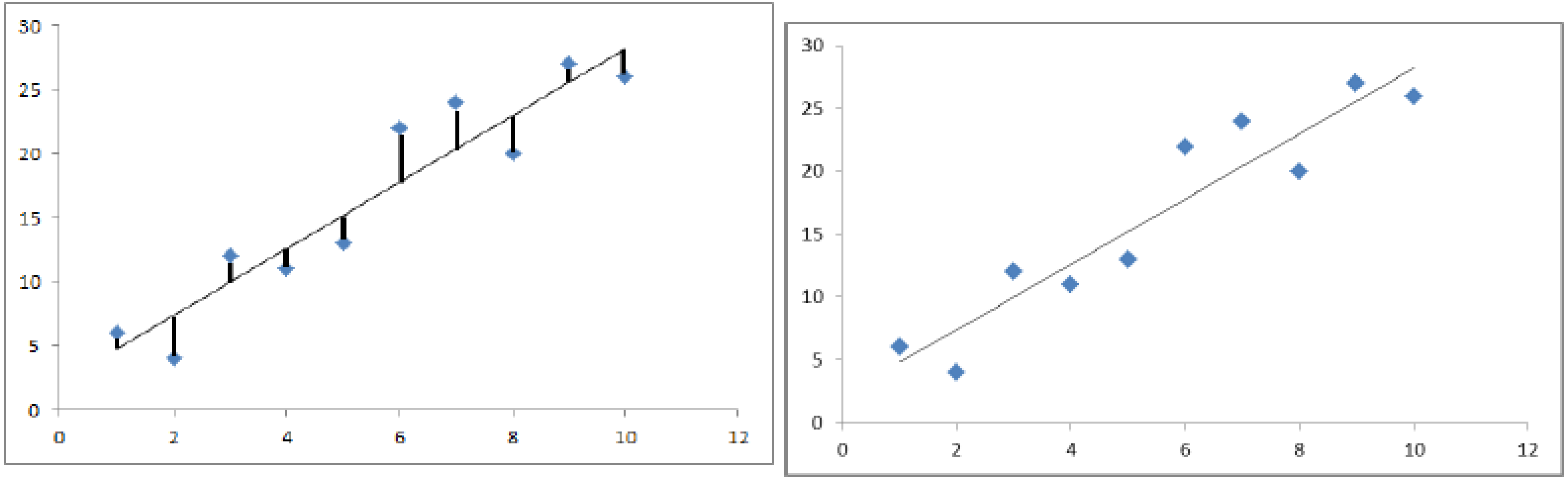


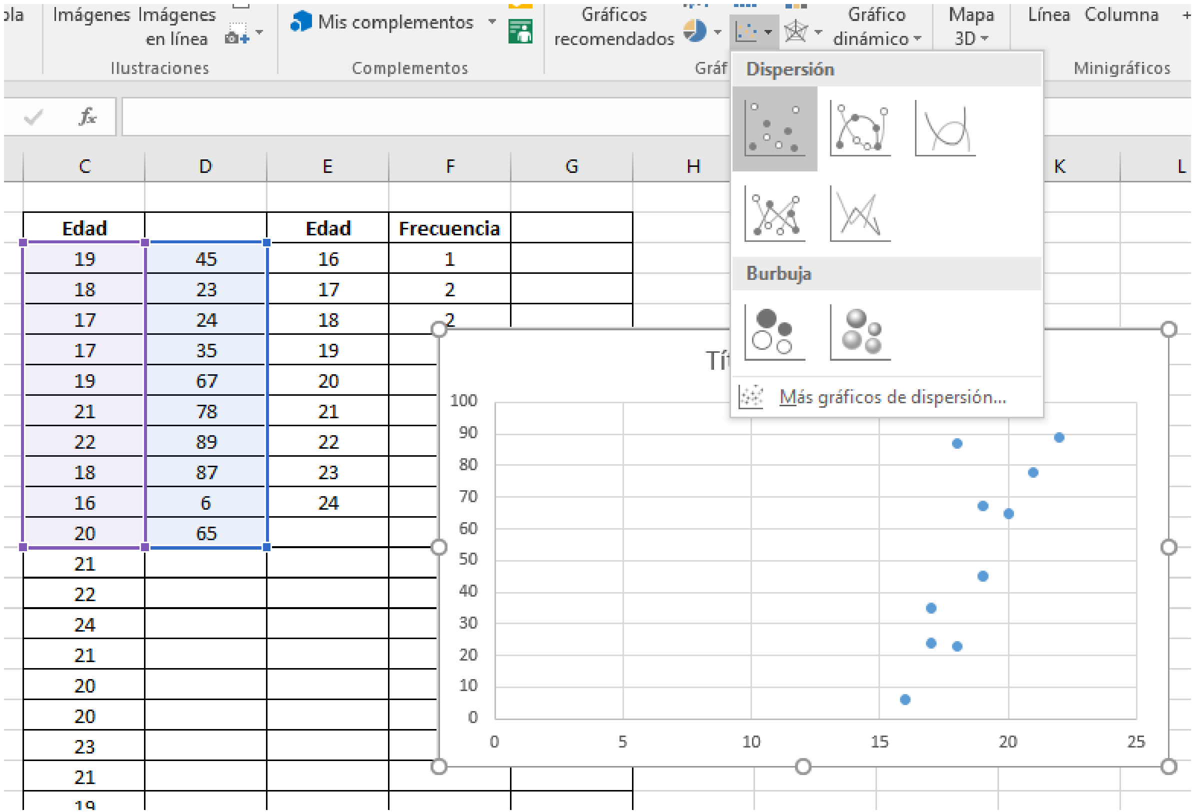
Figura 2.2 (a) Ajuste de la recta óptima trazada usando el criterio de que la suma de los cuadrados de las distancias verticales de los puntos a ella, sea mínimo. (b) La recta trazada por mínimos cuadrados posee sus respectivos parámetros A y B calculados por las ecuaciones

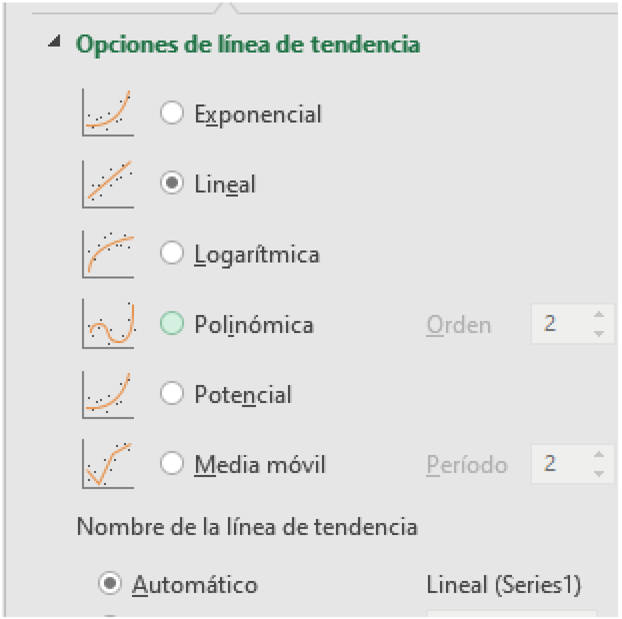
$$Y = A + BX \qquad \mathrm{A} = \frac{\bar{X}\sum_{\mathrm{i}=1}^{n} X_i Y_i - \bar{Y}\sum_{i=1}^{n} X_i^2}{n\bar{X}^2 - \sum_{\mathrm{i}=1}^{n} X_i^2} \qquad \mathrm{B} = \frac{n\bar{X}\bar{Y} - \sum_{\mathrm{i}=1}^{n} X_i Y_i}{n\bar{X}^2 - \sum_{\mathrm{i}=1}^{n} X_i^2}$$

Se define la ***correlación*** como una medida de la covariabilidad (variabilidad mutua entre las cantidades en estudio), que no depende de la escala. La correlación entre dos variables aleatorias está dada por:

$$R = \frac{Cov(X,Y)}{s_x s_y} = \frac{\overline{XY} - \bar{X}\bar{Y}}{s_x s_y}$$

Desde luego es posible ajustar un conjunto de datos a otro tipo de gráficas, el proceso y la idea es la misma al ajuste por mínimos cuadrados o a procesos como el ajuste de Lagrange. Esto se puede hacer fácilmente en excel





Método Montecarlo

Este método hace uso del azar utilizando números pseudoaleatorios para generar simulaciones cercanas a la realidad

Ejemplo: Calcular la probabilidad de sacar un 7 al lanzar dos dados

Podemos organizar el espacio muestral en su forma reducida, asociando una tabla de frecuencias

$$E_{red} = \{2,3,4,5,6,7,8,9,10,11,12\}$$

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

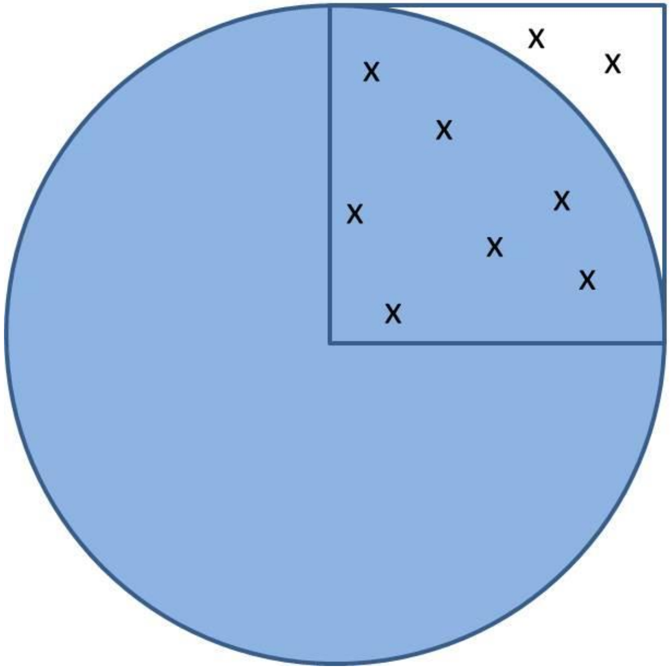
$$P(7) = \frac{6}{36} = 0.167$$

El tamaño del espacio muestral es 36

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

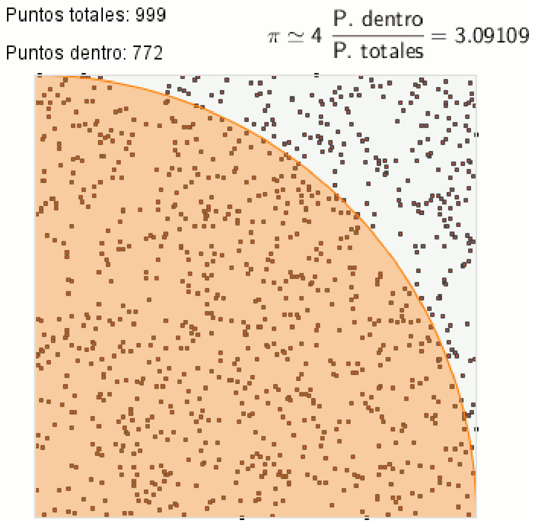
El método montecarlo se puede utilizar para resolver múltiples problemas que no parecieran depender del azar

Hallemos por ejemplo el valor de pi, utilizando dicho método.



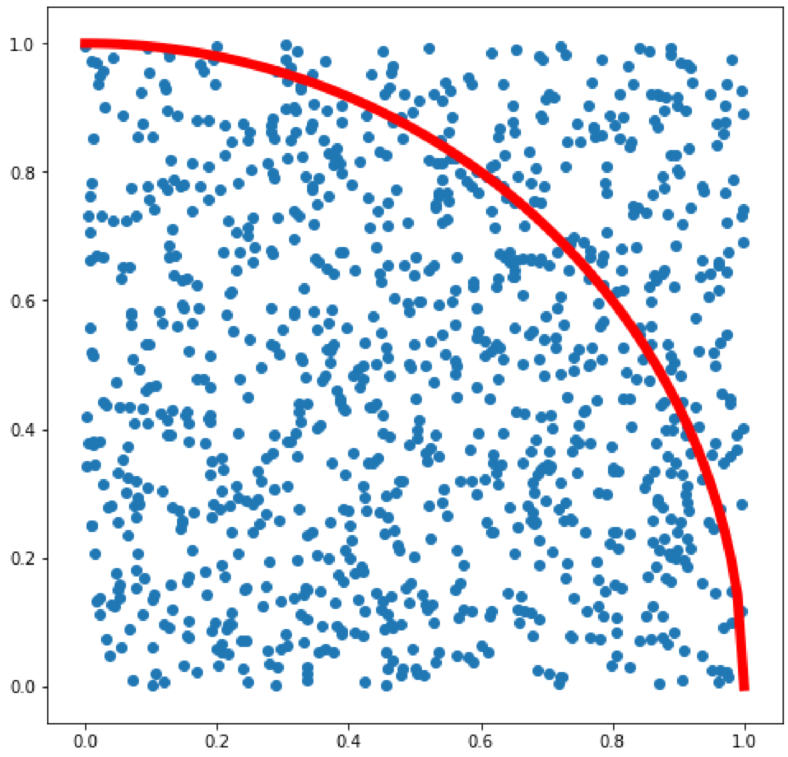
$$\pi \approx 4\,\frac{Dardos\ dentro}{Dardos\ totales}$$

Cuando se dibujan aleatoriamente puntos en un cuadrado unitario, estos se distribuirán uniformemente siempre y cuando el espacio muestral sea grande. El cociente entre el área del círculo y el cuadrado es pi/4 y este será proporcional al número de puntos dentro del cuarto de círculo respecto al número total de puntos



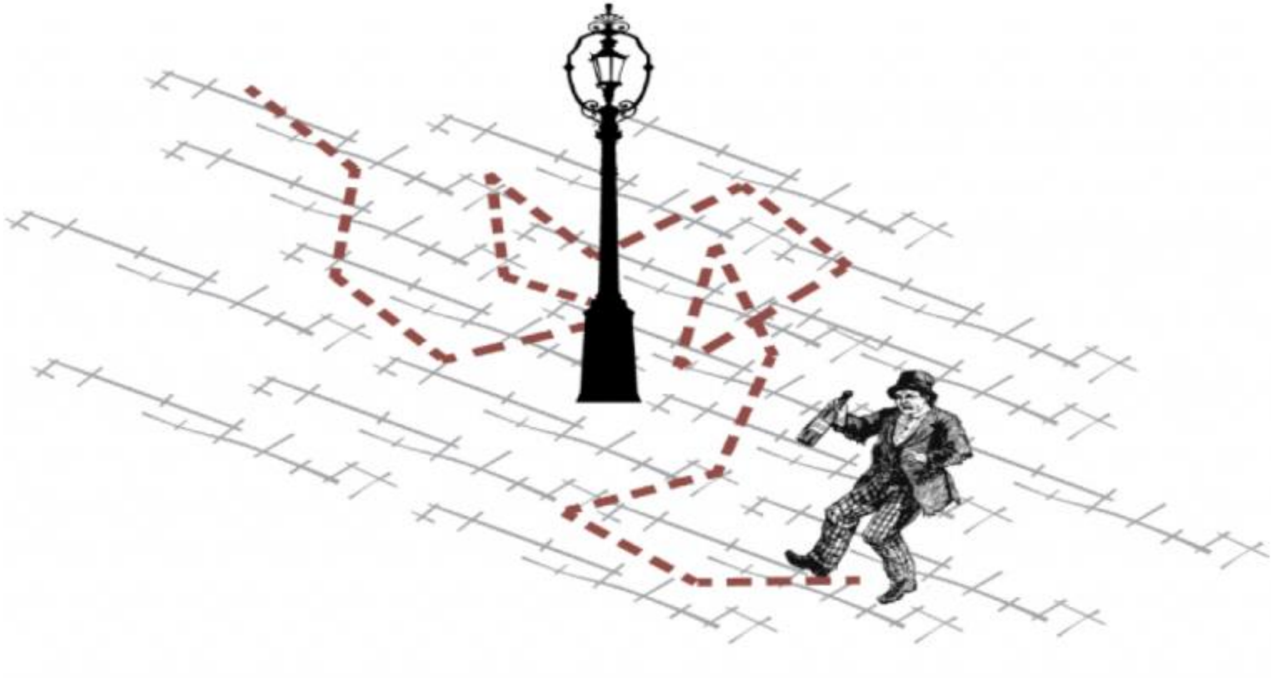
```python
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Thu Jun 23 08:26:23 2022

@author: carlos
"""
#import matplotlib.pyplot as plt
import random
import numpy as np
import matplotlib.pyplot as plt
N = 1000
c = 0
xx = []
yy = []
for i in range(N):
    x = random.random()
    y = random.random()
    l = np.sqrt(x**2+y**2)
    xx.append(x)
    yy.append(y)
    if l <= 1:
        c = c + 1
ppii = 4*c/N
error = abs(100*(np.pi-ppii)/np.pi)
f = lambda a: np.sqrt(1-a**2)
aa = np.linspace(0,1,100)
plt.rcParams["figure.figsize"] = (8,8)
plt.plot(aa, f(aa),color = 'red', linewidth=6)
plt.savefig('fig-pi.pdf')
print("Para un total de :",N, "puntos, el valor de pi es", ppii)
print("Con un error porcentual de ", "{:.2f}".format(error),"%")
plt.scatter(xx,yy)
```

```
Para un total de : 1000 puntos, el valor de pi es 3.224
Con un error porcentual de  2.62 %
```



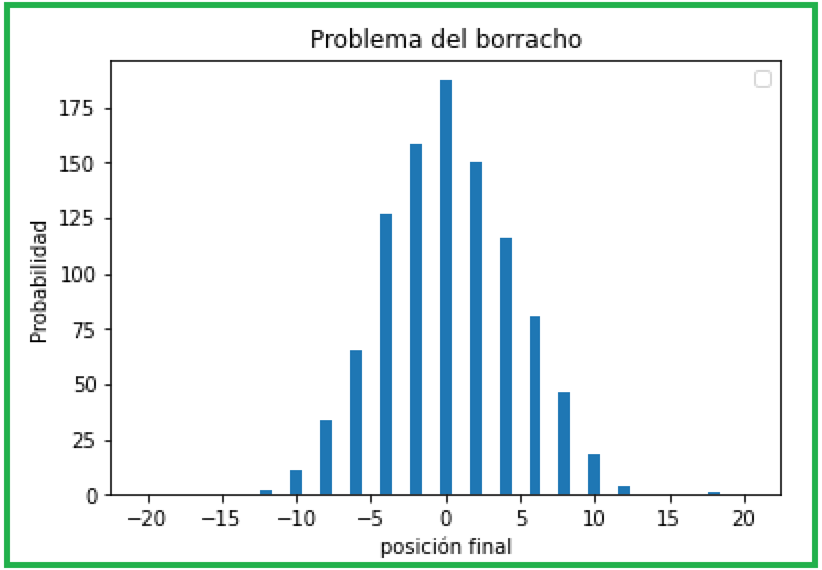
Problema del paso aleatorio

Un caminante puede dar aleatoriamente un paso hacia la derecha, o hacia la izquierda, ¿cuál será su posición final?



Este código simula problema en una dimensión

```
import matplotlib.pyplot as plt
import numpy as np
import random
#el numero de pasos es n, el experimento se realiza N veces
n = 20
N = 1000
L = [0]*(2*n+1)

for j in range(N):
    p = 0
    for i in range(1,n+1):
        x = random.randint(0,1)
        if x == 0:
            p = p + 1
        else:
            p = p - 1
    i += 1
    L[n+p] = L[n+p] + 1
j += 1
lista = []
for i in range(-n,n+1,1):
    lista.append(i)
plt.bar(lista, L)
plt.xlabel('posición final')
plt.ylabel('Probabilidad')
plt.title('Problema del borracho')
plt.legend()
plt.show()
```



Reto:

Escriba un código que generalice el problema del paso aleatorio pero ahora para dos dimensiones