

Maestría en Ciencia de Datos
Probabilidad y Estadística para Ciencia de Datos

Comportamiento de las Medias de una Distribución Binomial

Profesora: Ana Delia Olvera Cervantes
Alumno: Melchor Nolasco Cosijoeza

29 de noviembre de 2024

practica-clase-20

November 29, 2024

1 Comportamiento de las medias de 50 variables

El objetivo de esta práctica es observar el comportamiento que tienen muestras aleatorias(5,10,15,30,35,40,50) tomadas de un conjunto de 50 variables con 20 elementos una con distribución binomial.

```
[25]: import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom
import random
```

1.1 Generación de 50 variables con 20 elementos en cada una con distribución binomial

```
[26]: n = 20
p = 0.16
matrix = []
```

```
[27]: for i in range(50):
    binomial_samples = np.random.binomial(n=20, p=p, size=n)
    print(binomial_samples)
    matrix.append(binomial_samples)
```

```
[ 7  6  9 11  5  7  9  8  8 10  7 11  8 10  9 10 11  5  8 12]
[ 7  7 10  6  5  9 10  7  4  7  6  7  6  9 11  9  4  7  7 11]
[ 6  7 10 12 11  5  7  8  5 10  4 10 10  3  7  6  7  8  5  7]
[ 6  4  9  6  8 12 10 11  7  9  4  6  8 14 10 10 10 13 10  6]
[11 14  7  7  7  3  7  8  5  8  5  7  6  4 11  9  5  7 10 12]
[11  8  9  9  4  9  8  9  5  5  8  9  6  8  8 16  5 10 10 13]
[ 8  6 10  6  9  6  8  6  6  6 12  7  7 11  7  8  4  7  5 11]
[ 9  5  6 12  5  9  5  7  8  8  7  8  6  8  3  3  6 14  6 15]
[ 8  5  8 10  8  5  7 13 11 11 10  5 14  8  7  9  4  5  8 10]
[11  9  6 10  7  9  8  2  7  4 11 12  8  9  3  8  7  8  8 10]
[ 9 10 10  7  7  8  7  8  8  7 12 10  9 12  9 10 11 10  7 11]
[ 9  7 10 10 10  8  3 10  6  7  4 10  8  8  8 15  6  6 12 10]
[10 11  8  4 10  9  6  6  8  8  7 10  9  8 10  7  6 11  5 10]
[12  6  5  8  8  5  9  8 11 10  7  8  7  4  3  6  5  6  7  5]
[ 7 11 10  4  6 11  9  7  9  7  9 10  6  6 14  8 10  5  5 10]
```

```
[ 5  7  8  6  9  6 11  9  9  6  8  8 12  9  7  7 11 10 12  7]
[10 11  6 12  9  7  6  5  6  9  7  9  5  8  9  6  8 10 14  4]
[11  5 10  7  8  6  5  6 11 10  5 11  7  6 13 11  5  4  6  5]
[ 7  6 10 10 13  5  6  9  6  5  8 11  4  6  4 10  5  7 11  6]
[12  6  3  5 11 11  8  9  8  6  4  8  5  8  7  3 10  9 10  7]
[10 12  5  6  9  9  8 11 11  7  6  5  8 12 10  8 12 10  4  9]
[ 7  8  8 11  5  7  3  4  7  7 10  7 10  5 13 10  8  8  7  9]
[ 7 10  6  9  7 10  5 15 12  6 10  7  9  9  8  7  7  7  5  9]
[ 5  6  5  6  7  9  7  8  7  8  5  7 11  5  9 12  4  9 10  6]
[ 6  5  7  8  6  7  7  4  5  9  5  5  9 13  4 10  8  4 16  7]
[ 9  8  4  8 11  2  7  6  8 11 11  3  8 10  9  6 12  6 13  8]
[10  9  7  0  9  7 13  8  6  5  7  8  8  9  9  8  4 10  3  7]
[ 5  9  7  7  7  5  9  9 12 16  4  9  6  9  8  6  8  8 10  9]
[13  7  6  8 10 11 11 10  6 12  9  8  5  9  9  7  5  9  9 10]
[ 4  3  8  7  8 14  7  9  5  3 12  9 11 12  7  6  9 12  6  7]
[ 5 10  1  7 10  7  9  4  8 11  7  5  9 10 11  7  4 12  6  8]
[ 9  3  7  5  8  9 10 10 11 10  5  8  6  5  8  2  9  7 13  8]
[ 9  7  5  3  8 11  9  6  7  9  8 11 12  8 13  6  7  5 10 13]
[ 5 13  5 11  8  7  9 12 11  4  7  6 10  6 13  6 10  7 10  8]
[ 7 12  3  5 12  7 11  7  9  7 10  3  6  2  5  6  6  8 16 10]
[ 9 10  7 10  3 14  9 15  4  3  7  7 11 12  9  8  8  9  4  7]
[10  6  3  6  7 11 13 10  1 10  5 10  8  5 10 12  8  9  5  5]
[ 6  7  6  6 11  7  7 10  3  7  9 11  6 11  8 11  8  9  7  8]
[ 3  7  8 13  7  5  7  7  7  9  8  5  6  5  8  7 13 12 10 14]
[ 5 11 12 11  6  9  7  4  4 13  9  9  8  5  6  8 11  8  9  5]
[10  6  5  4  8 12  8  7 12  8  8 11 15  8 12 12  9  6  8  7]
[ 8 10 13  7  9  7 10  8  9  8 13  6  7  4  9 10  2 10  8 15]
[ 7 13  7 12  9  9  7  7  6  6  8  8  8  7 13  9 14  4 10  7]
[ 6  7  8  9  8  6  8  4  6  8  6  3  5  7  6 10  8 10  8  5]
[10  2  7  5  7  8  5  8 15  5  9  5  9 10  5  9 10 11 10  6]
[10  4  5 12  6  5 10 11 10  9  6 15  8  7 12  8  8  6 10  9]
[ 7  9  5  5  9  6 13  2  8  9  4  9  7 11  6 10 10 10  6  6]
[ 4  9 11 11 10 11  6 13  7  9  7  9  8  9  4 10  4 12  7  8]
[12  4  9  7  7  8  4  5  6 10  7  8  6  7  9 17 11  9  5  4]
[ 7 12 12  9  8 10  5  9  8 10 10  7 13  4  6  5  9  7 10  5]
```

```
[28]: print("Conjunto de {} con {} elementos".format(len(matrix),len(matrix[0])))
```

Conjunto de 50 con 20 elementos

```
[29]: # meanVariables calcula el promedio de las muestras y retorna un vector de tamaño n.
def meanVariables(sampleX, size):
    accumulator = 0
    result = []
    for j in range(len(sampleX[0])):
        for i in range(len(sampleX)):
            accumulator += sampleX[i][j]
```

```

        accumulator += sampleX[i][j]
    mean = accumulator / size
    result.append(mean)
    accumulator = 0
return result

```

```

[30]: # createGraphics crea un histograma del conjunto de datos que recibe
def createGraphics(data, size):
    fig, ax = plt.subplots()
    plt.hist(data, density=True)
    plt.title(f"Distribución binomial con {size} muestras", size=14)
    plt.xlabel("Muestras")
    plt.ylabel("Densidad")

```

1.2 Tomamos muestras de manera aleatoria, calculamos su promedio y graficamos su histograma

```

[31]: subsetVariables = []
sampleSize = [5, 10, 15, 30, 35, 40, 50]
for i in sampleSize:
    print(f"Muestras de {i} X")
    for j in range(i):
        randomIndex = random.randint(0, 50 - 1)
        subsetVariables.append(matrix[randomIndex])
    mean = meanVariables(subsetVariables, size=i)
    createGraphics(mean, size=i)
    subsetVariables = []
    print(mean)

```

Muestras de 5 X

```
[7.0, 5.8, 9.4, 8.2, 7.6, 10.0, 5.8, 7.8, 7.4, 8.0, 8.8, 8.0, 8.4, 8.6, 6.2,
10.0, 9.0, 10.4, 7.4, 6.0]
```

Muestras de 10 X

```
[7.7, 7.6, 6.7, 6.9, 8.7, 9.4, 8.5, 9.3, 6.5, 8.6, 8.3, 8.4, 8.1, 9.8, 8.9, 8.8,
6.8, 9.3, 7.1, 8.9]
```

Muestras de 15 X

```
[7.8, 7.133333333333334, 7.6, 7.933333333333334, 9.066666666666666,
7.266666666666667, 8.6, 7.8, 7.333333333333333, 7.533333333333333,
7.333333333333333, 8.266666666666667, 7.266666666666667, 8.466666666666667,
7.466666666666667, 8.466666666666667, 7.2, 8.133333333333333, 8.666666666666666,
7.733333333333333]
```

Muestras de 30 X

```
[8.766666666666667, 8.433333333333334, 7.7, 8.366666666666667, 8.1,
7.633333333333334, 7.3, 6.833333333333333, 7.433333333333334, 7.466666666666667,
7.533333333333333, 8.733333333333333, 7.433333333333334, 7.533333333333333,
8.166666666666666, 8.533333333333333, 7.766666666666667, 7.566666666666666,
8.766666666666667, 8.0]
```

Muestras de 35 X

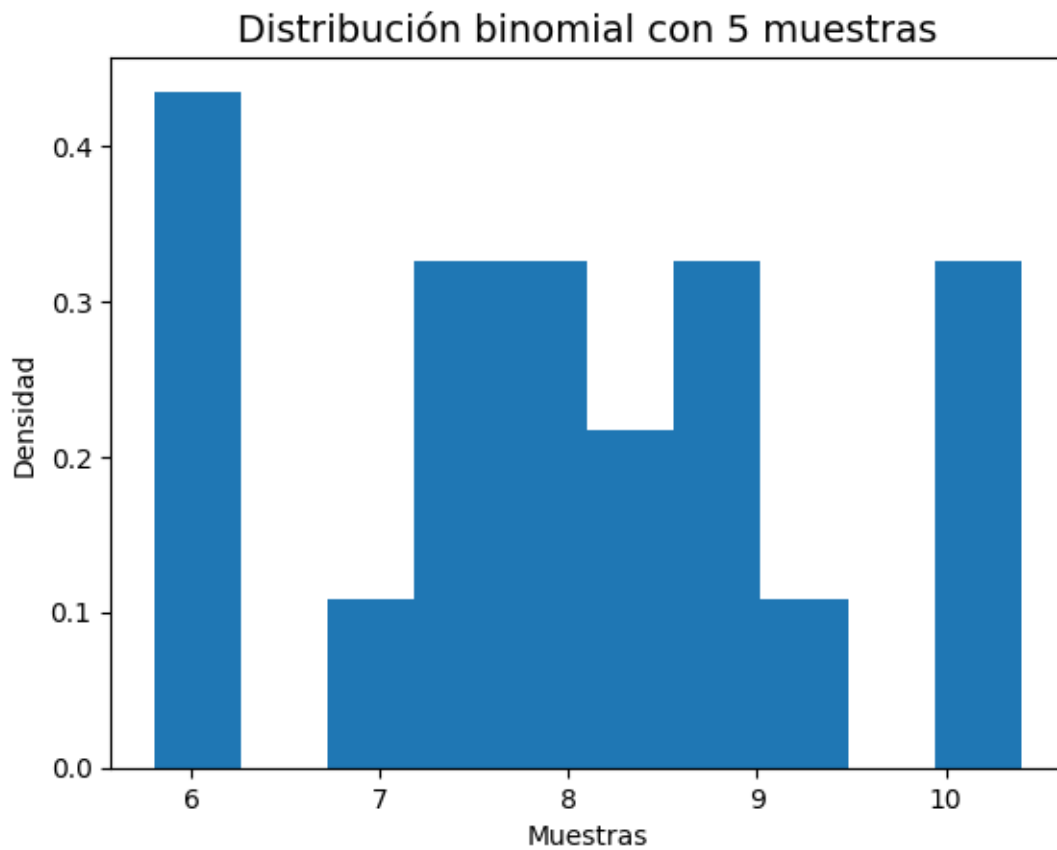
[7.6, 7.857142857142857, 7.542857142857143, 7.514285714285714,
8.314285714285715, 8.0, 8.028571428571428, 8.714285714285714, 8.257142857142858,
8.257142857142858, 7.685714285714286, 7.628571428571429, 8.285714285714286,
8.057142857142857, 7.628571428571429, 7.085714285714285, 8.114285714285714,
8.685714285714285, 8.2, 7.742857142857143]

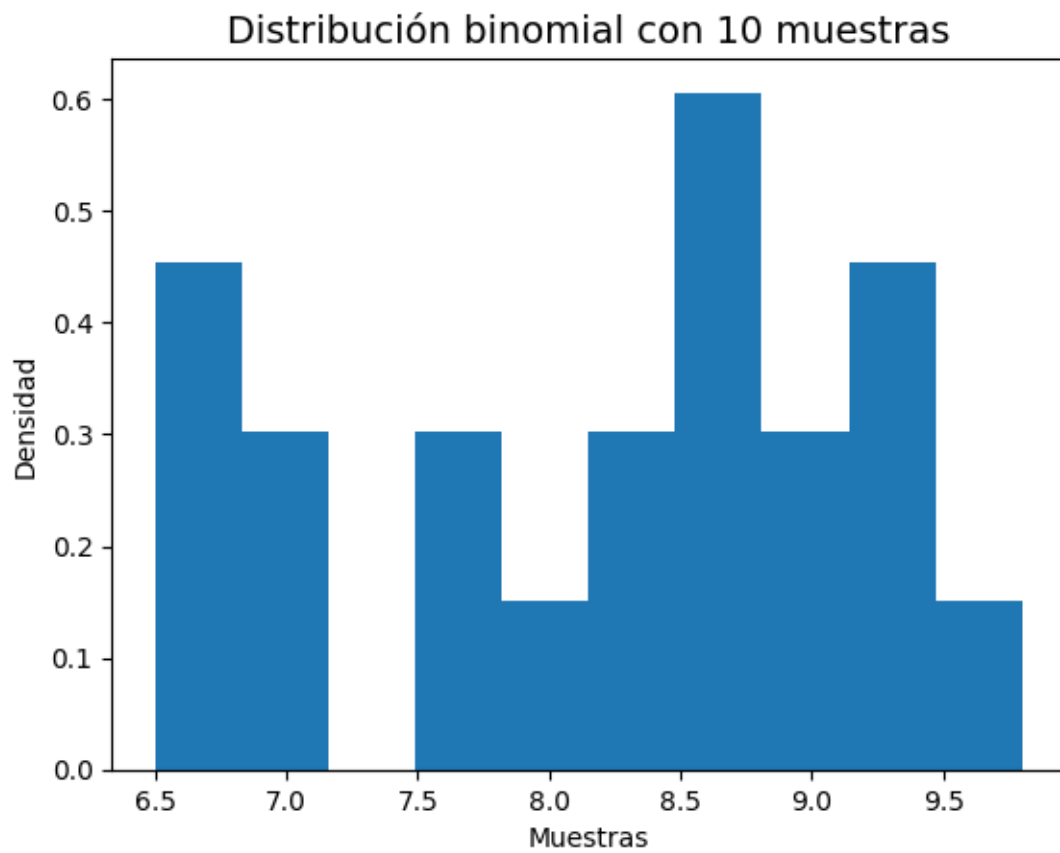
Muestras de 40 X

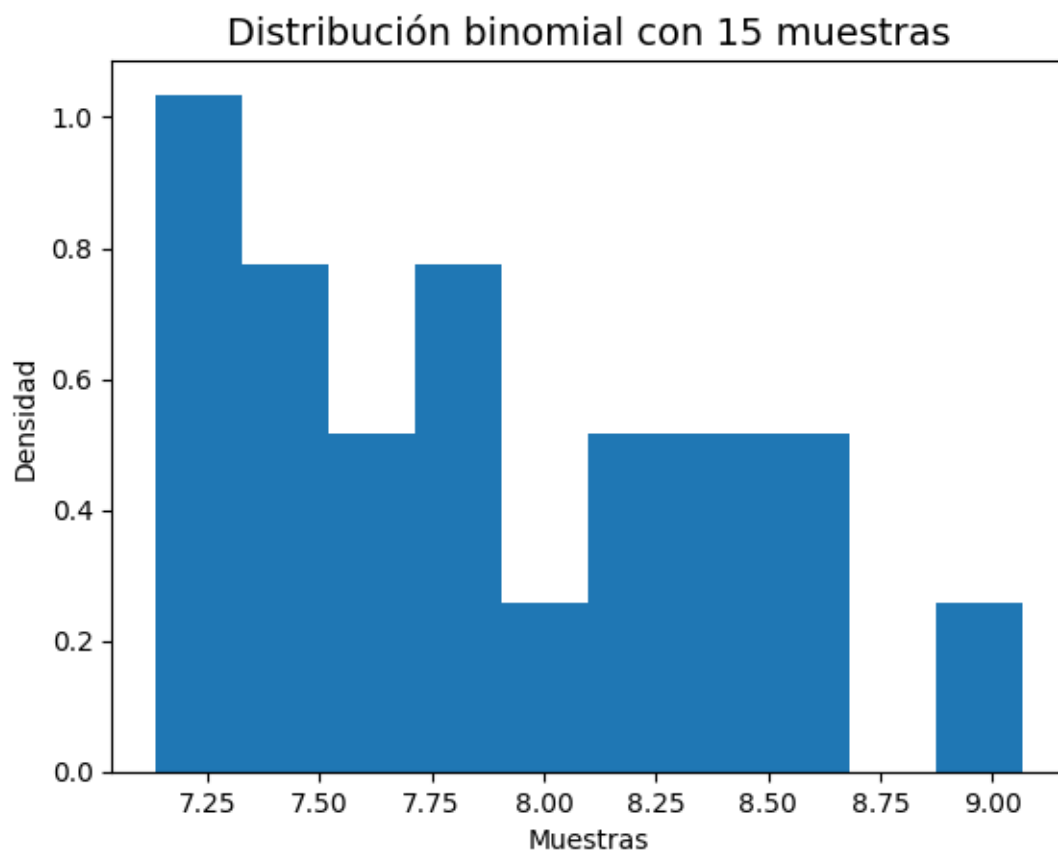
[8.2, 7.35, 7.55, 8.275, 7.9, 8.0, 7.55, 7.6, 7.4, 7.75, 8.125, 7.7, 8.2, 8.175,
7.65, 8.525, 8.2, 7.65, 8.975, 8.325]

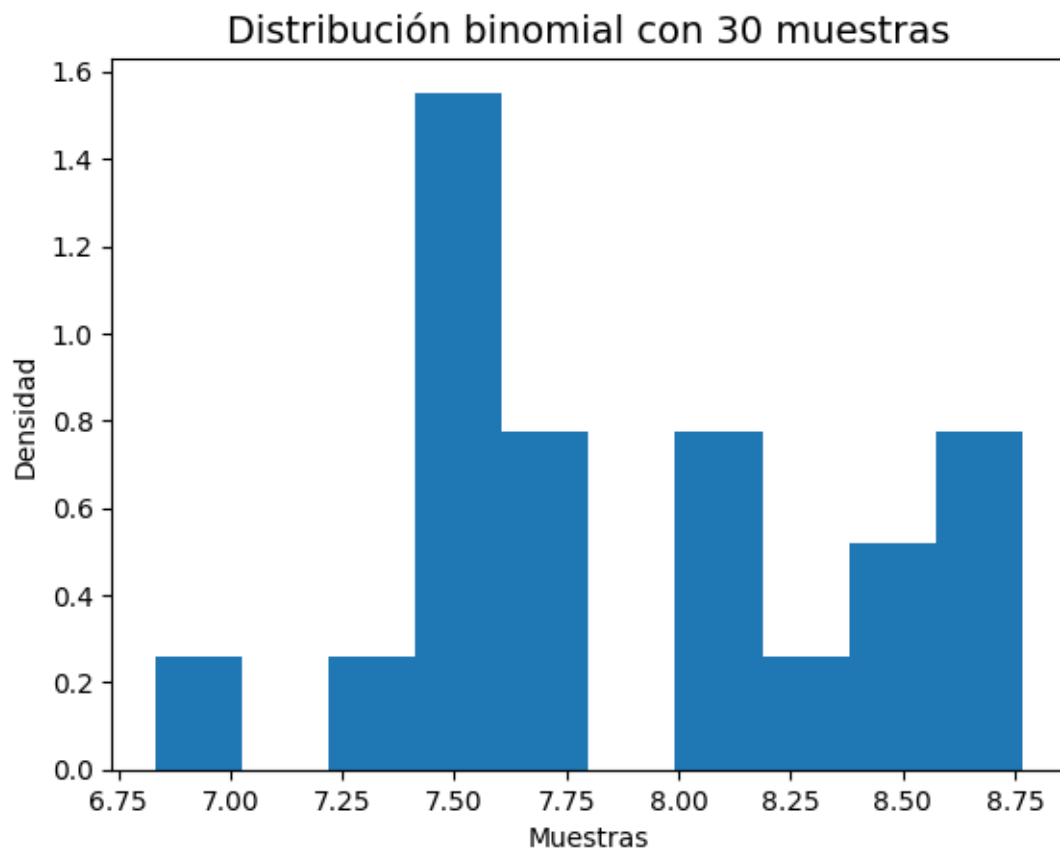
Muestras de 50 X

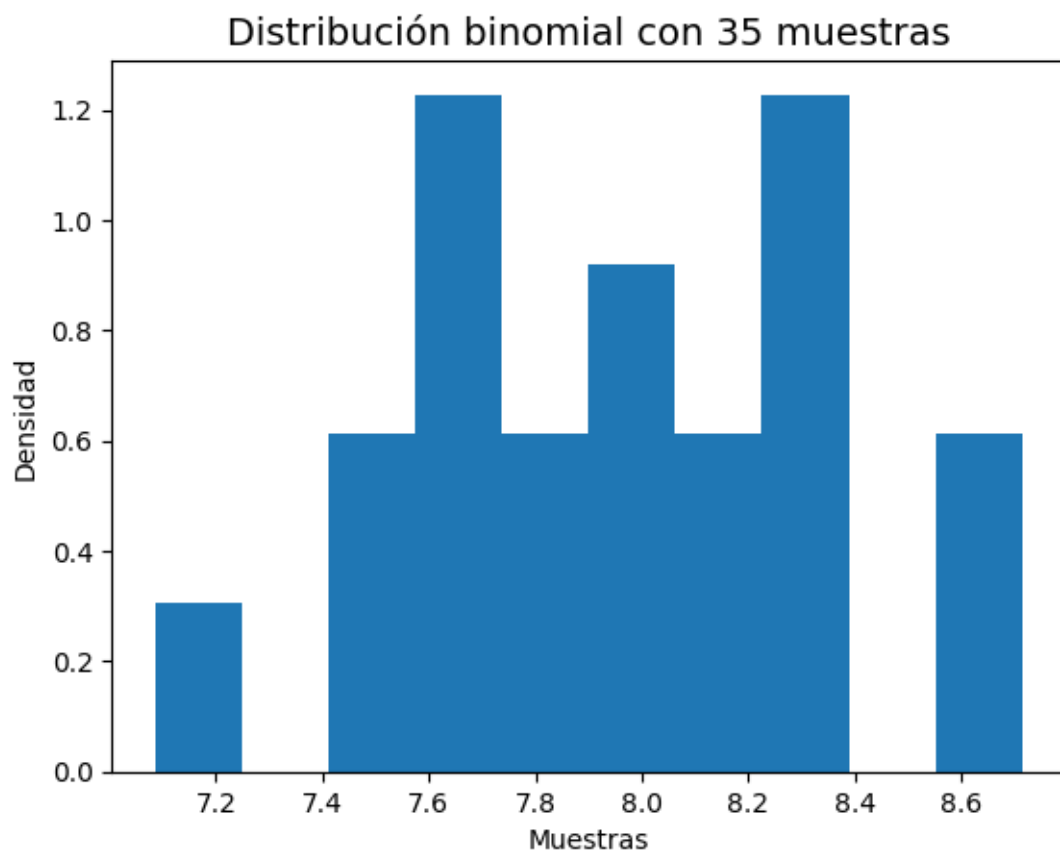
[8.22, 6.98, 8.34, 7.42, 7.52, 8.36, 7.52, 8.22, 8.2, 8.34, 7.64, 8.08, 8.44,
7.84, 8.38, 8.68, 7.36, 8.3, 8.24, 8.56]

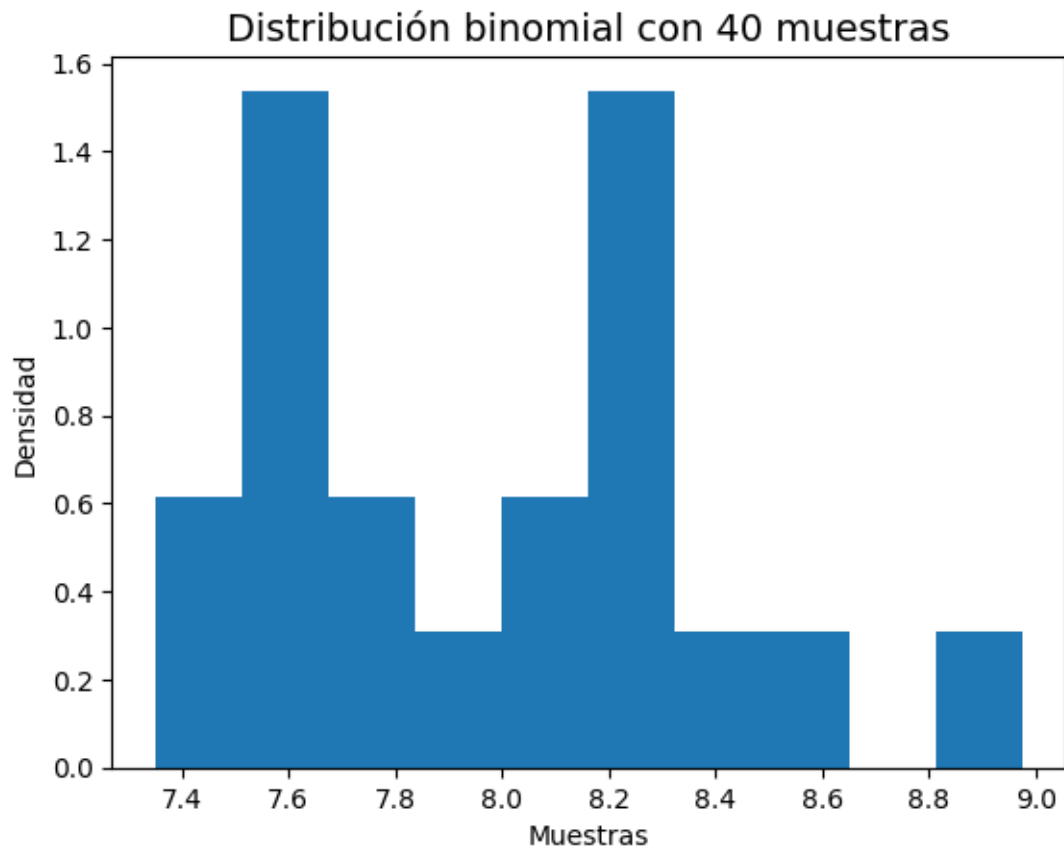


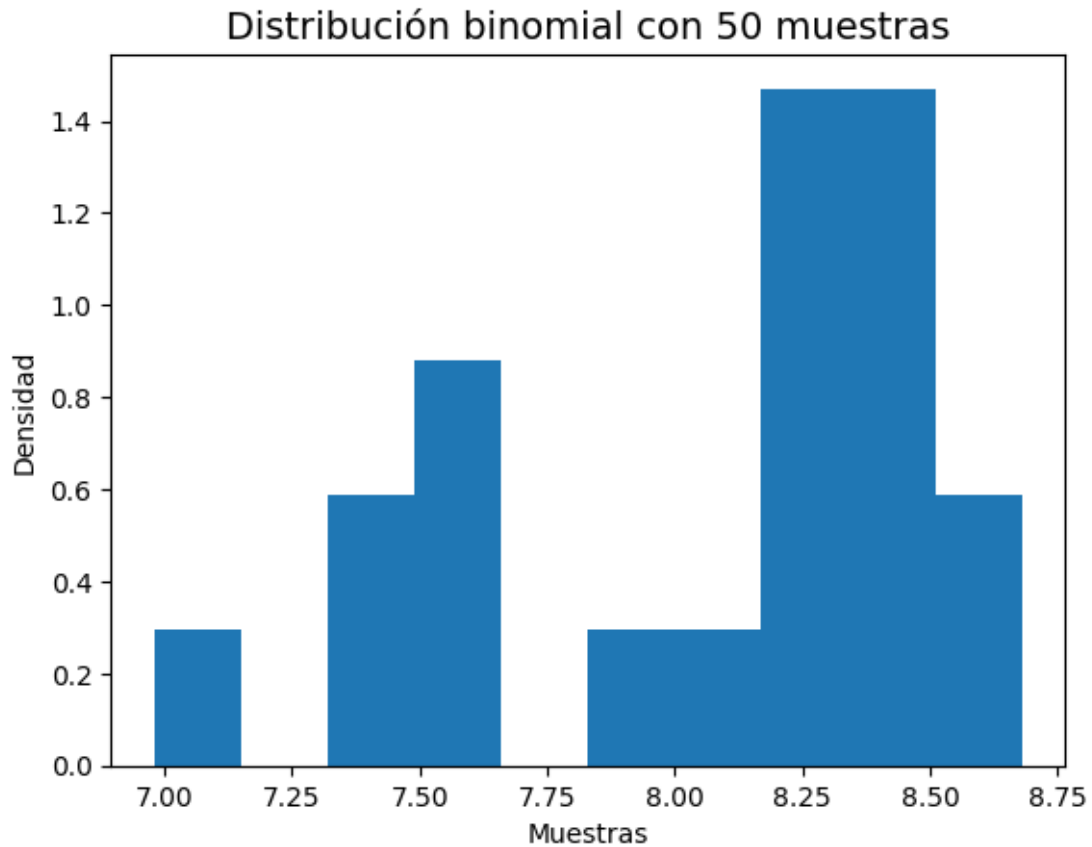












1.3 ¿Qué va sucediendo cuando la muestra de datos aumenta?

Los histogramas van cambiando mucho con el numero de muestras. Cada histograma tiene la información centrada de diferentes maneras y en diferentes lugares. En algunas partes va tomando la forma de una distribución normal.