Cuando trabajamos el análisis inferencial (calculamos un modelo predictivo o aplicamos un test) el primer paso es saber de qué distribución proviene la muestra de datos que tenemos.

Es decir, encontrar la forma de la distribución poblacional de nuestras variables.

En este ejercicio vas a hacer precisamente lo contrario.

A partir de las distribuciones conocidas vas a generar variables que sigan estas distribuciones.

Esto te ayudará mucho a captar los conceptos de distribución de probabilidad y de muestra y población.

Para ello vas a utilizar la generación de números aleatorios a partir de distribuciones de probabilidad conocidas.

¡A por ello!

Abre un Excel y genera las muestras.

El número de observaciones será 100 para todas las variables generadas.

Te he facilitado un Excel de ejemplo para que entiendas cómo estoy creando las muestras.

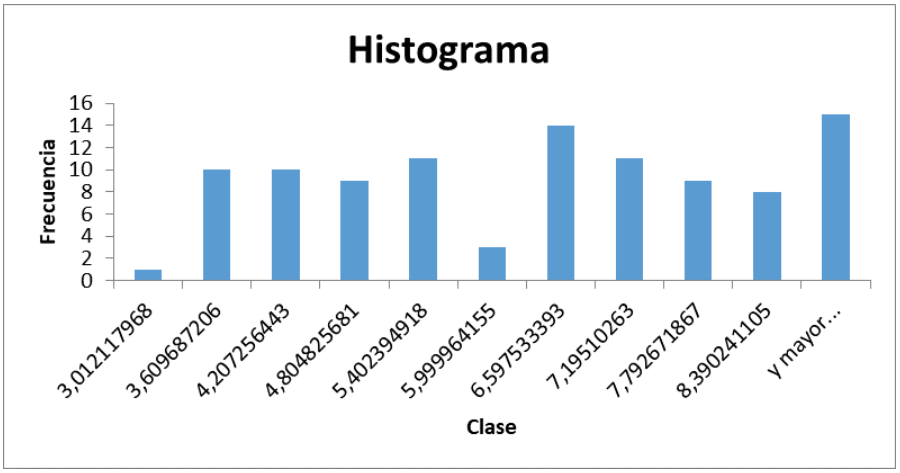
Distribución Uniforme (continua)

Crea una muestra de 100 observaciones que provenga de una distribución de probabilidad uniforme entre 3 y 9.

Calcula la media y la mediana y la desviación estándar y IQR (rango intercuartílico).

Dibuja el histograma de esta muestra.

Copia el histograma y los valores de centralidad y dispersión.



Promedio	5,944
Desvest	1,686
Mediana	6,251
Q3	7,274
Q1	4,454
Rango intercuartílico	2,820

Distribución Normal (continua)

Crea una muestra de 100 observaciones que provenga de una distribución de probabilidad normal con:

- Media: 50
- Desviación Estándar: 10

(Este ejercicio es parecido a la hoja de trabajo anterior, hazlo si quieres ☺)

Distribución Exponencial (continua)

Crea una muestra de 100 observaciones que provenga de una distribución de probabilidad exponencial con:

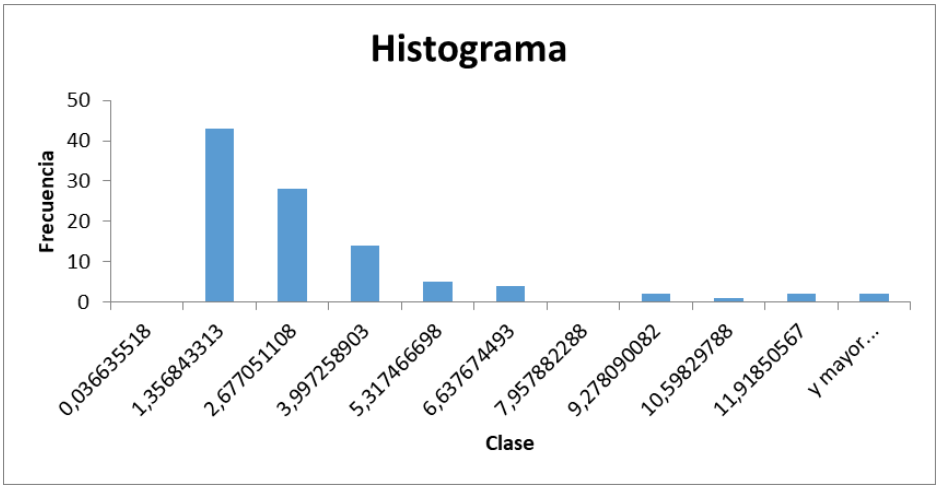
- Landa: 3

Inspírate del Excel y el vídeo Tech Zen para calcular estas distribuciones.

Dibuja el histograma y calcula la centralidad y la dispersión:

- Media y desviación estándar
- Mediana y rango intercuartílico

Copia el histograma y los valores de centralidad y dispersión.



Promedio	3,264
Desvest	3,143
Mediana	2,570
Q3	5,009
Q1	0,797
Rango intercuartílico	4,212

¿Cuál crees que es mejor medida de centralidad y dispersión?

Dado que la distribución no es normal, quizás sería más conveniente utilizar la mediana como medida de centralidad y el rango intercuartílico como medida de dispersión.

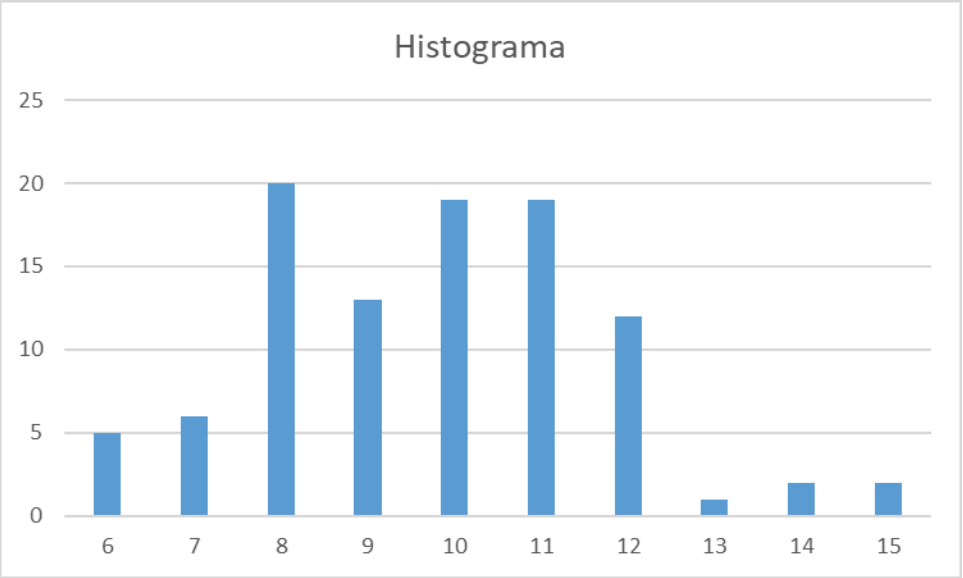
Distribución binomial (discreta)

¿Cuántas mujeres aparecen en 100 muestras de 20 observaciones recogidas de una población con una probabilidad del 50% de ser mujer?

Es decir, crea una muestra aleatoria de una distribución binomial de n = 20 y probabilidad de éxito del 50%.

Después puedes resumir los resultados en una tabla de frecuencias y construir el gráfico de barras. ¿Qué forma intuyes?

Copia el diagrama de barras



Etiquetas de fila	Cuenta de 12
6	5
7	6
8	20
9	13
10	19
11	19
12	12
13	1
14	2
15	2
Total general	**99**

Se observa una forma similar a la de la distribución normal, pero con una especie de picos y valles en la zona central

¡Bien hecho!

¡Estás profundizando en las probabilidades y el concepto de muestra y población!

¡Enhorabuena!