CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES

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)

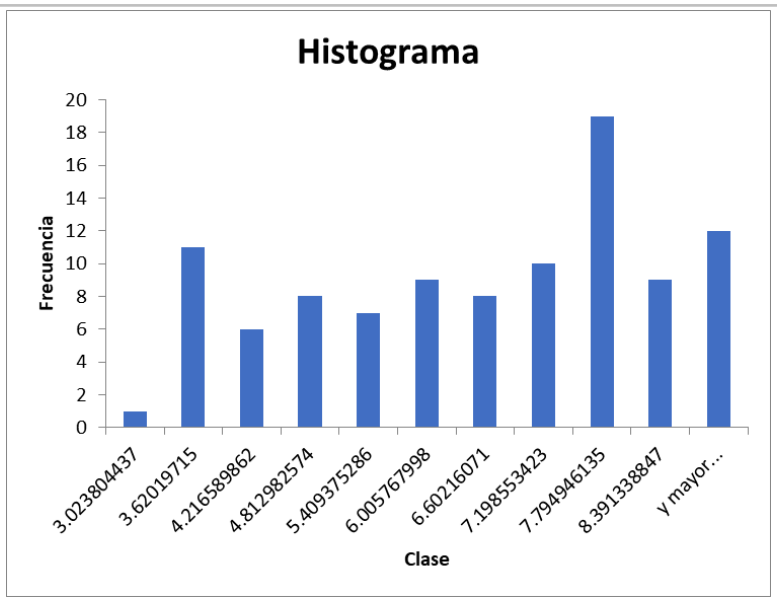
Crea una muestra de 100 observaciones que provenga de una distribución de probabilidad uniforme entre 3 y 9.

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.

media	6.23047273
mediana	6.53798029
desviación estándar	1.76866196
IQR	3.04156629

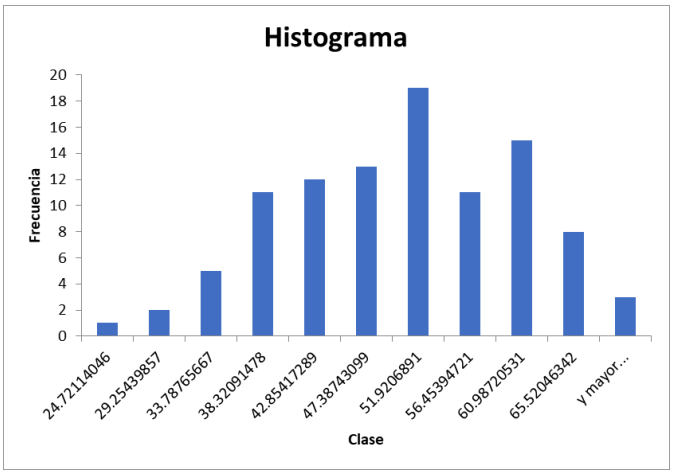


Distribución Normal (continua)

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 ☺)



	48.353936
media	9
	48.549856
mediana	2
desviación estándar	10.233319
	16.466657
IQR	4

Distribución Exponencial (continua)

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

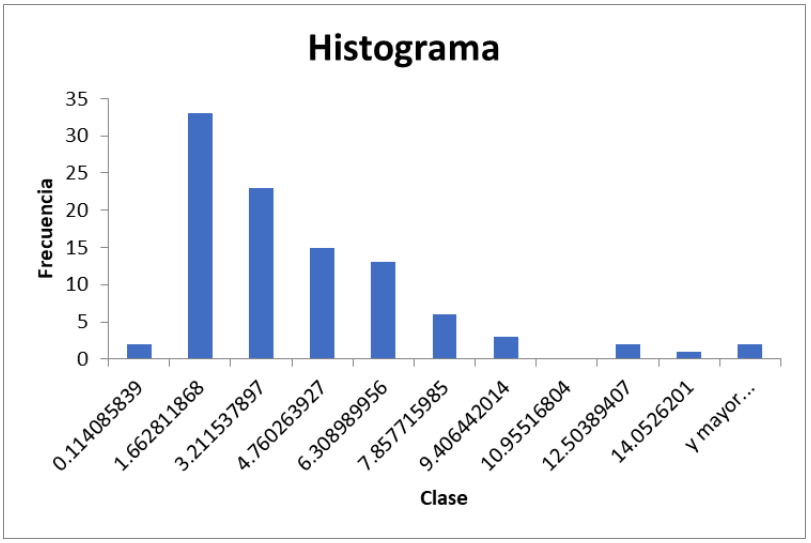
- Landa: 3

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.



	2.6888911
media	5

	1.8467879
mediana	9
	2.4949638
desviación estándar	5
	2.8511661
IQR	5

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

En graficas que no son en forma de montaña es mejor usar el rango IQR y la mediana porque los valores atípicos afectan el promedio y la desviación estándar. Cuando tienen la forma de campana se puede confiar en estas últimas.

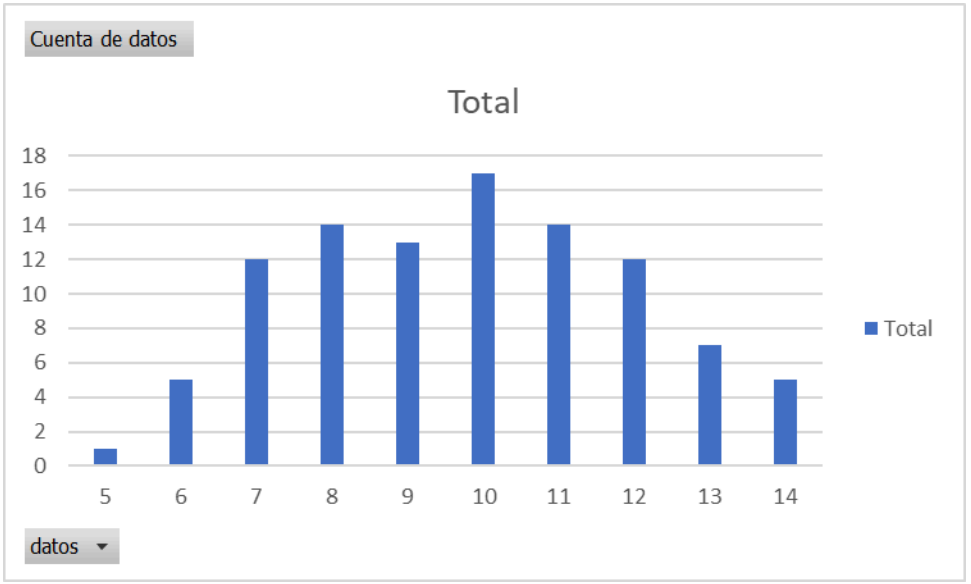
Distribución binomial (discreta)

¿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



Tiene forma de campana

¡Bien hecho!

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

¡Enhorabuena!