

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.

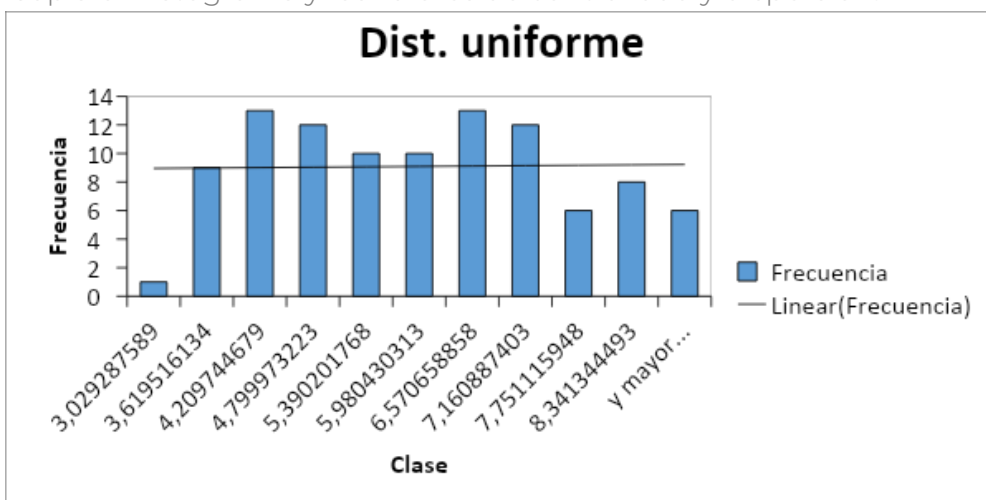
Te he facilitado un Excel de ejemplo para que entiendas cómo estoy creando las muestras.

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.



Columna1

Media	5,970808689
Mediana	5,961229606
Desviación estándar	1,775118326

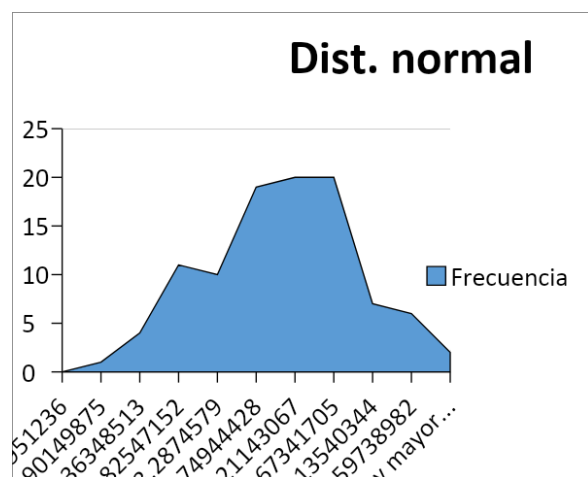
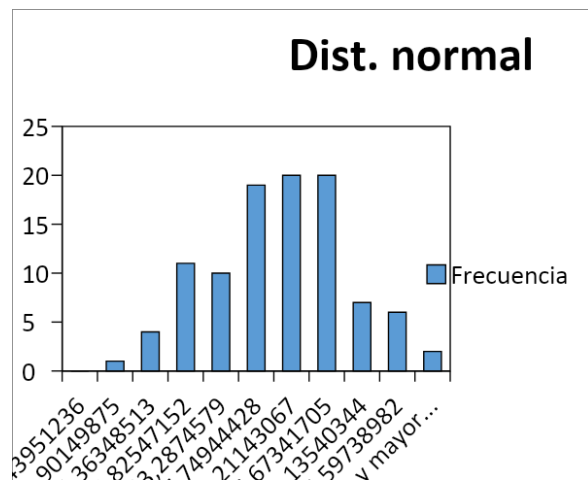
Q3	7,693059714
Q1	4,357238764
IQR	3,33582095

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

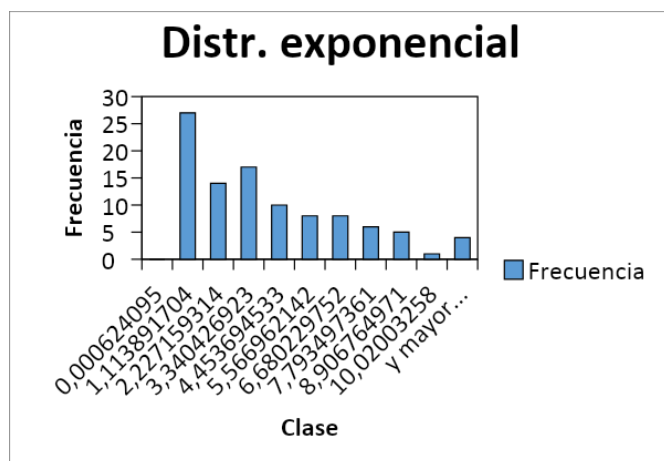
CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES

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.



Media 3,033121
 Mediana 1,99689276
 Desviación estándar 3,20382607
Q3 4,09155205
Q1 0,65569691
IQR 3,43585514

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

La distribución normal, ya que los valores están más concentrados próximos al valor central.

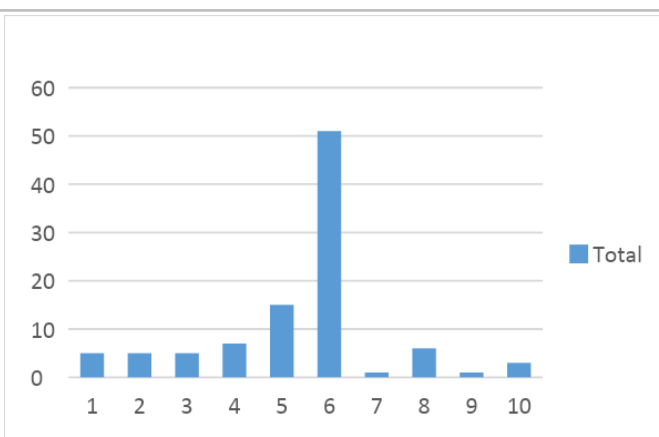
Distribución binomial (discreta)

¿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?

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



Etiquetas de fila	Cuenta de 4
1	5
2	5
3	5
4	7
5	15
6	51
7	1
8	6
9	1
10	3
Total general	99

La distribución es parecida a una distribución normal, con la mayoría de valores en la centralidad.

¡Bien hecho!

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

¡Enhorabuena!