

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)

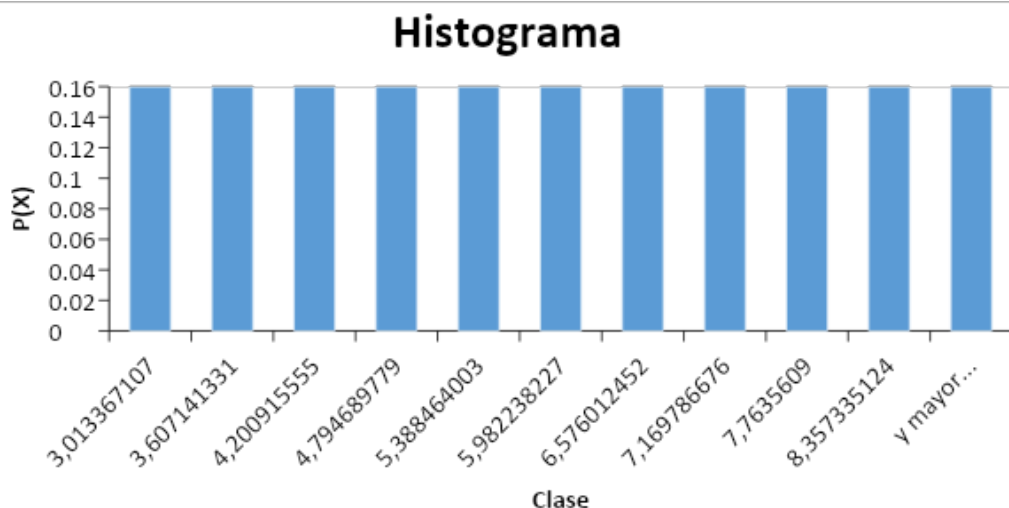
Crea una muestra de 100 observaciones que provenga de una distribución de probabilidad uniforme entre 3 y 9.

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	MEDIANA	DT	Q3	Q1	IQR
6,18	6,13	1,70	7,90	4,83	3,06



NOTA: He debido calcular el valor máximo de probabilidad para hacer el histograma con la fórmula $1/b-a$. Me queda una duda de la correspondencia entre el histograma de frecuencias y el histograma de probabilidad ¿qué relación guardan uno con el otro? En el tema no aparece nada al respecto y se generan los histogramas de frecuencia como si fueran los de probabilidad. **Ruego aclaración.**



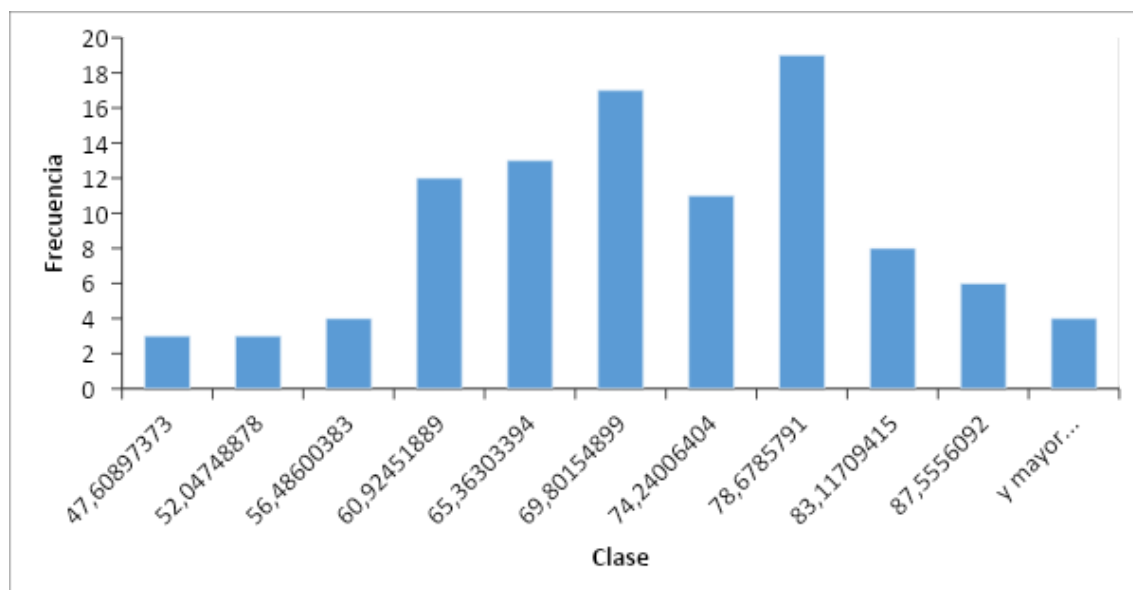
Distribución Normal (continua)

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 ☺)

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



MEDIA	MEDIANA	DT	Q3	Q1	IQR
69,08	69,50	10,15	75,44	62,52	12,93

Distribución Exponencial (continua)

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

- Landa: 3

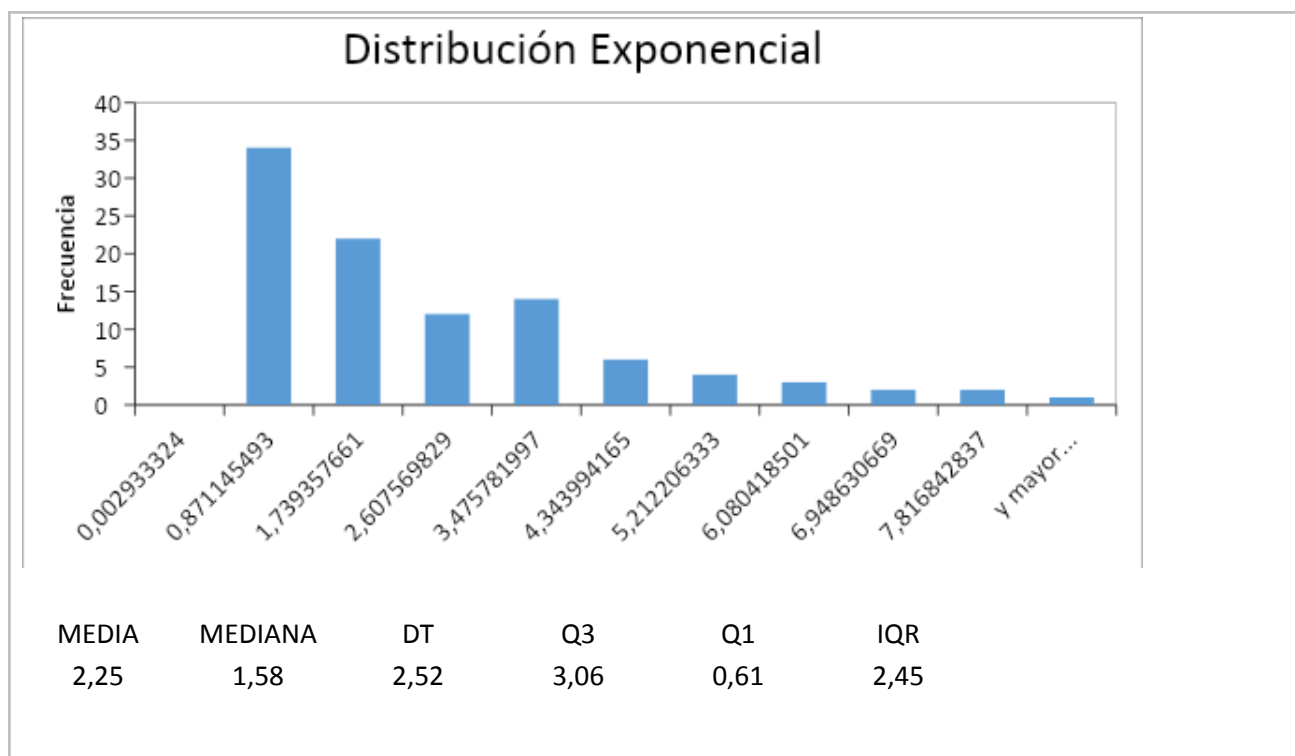
Inspírate del Excel y el vídeo Tech Zen para calcular estas distribuciones.

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.

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



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

La mediana como medida de centralidad y el recorrido intercuartílico como medida de dispersión (muy similar en este caso a la DT). La mediana deja la misma cantidad de observaciones a un lado y al otro de su valor, al disminuir la probabilidad a medida que aumenta X, la mediana reflejaría mejor la centralidad y, dado que la Mediana es Q2, la diferencia entre el Q3 y Q1, marcarían la dispersión en la distribución.

Distribución binomial (discreta)

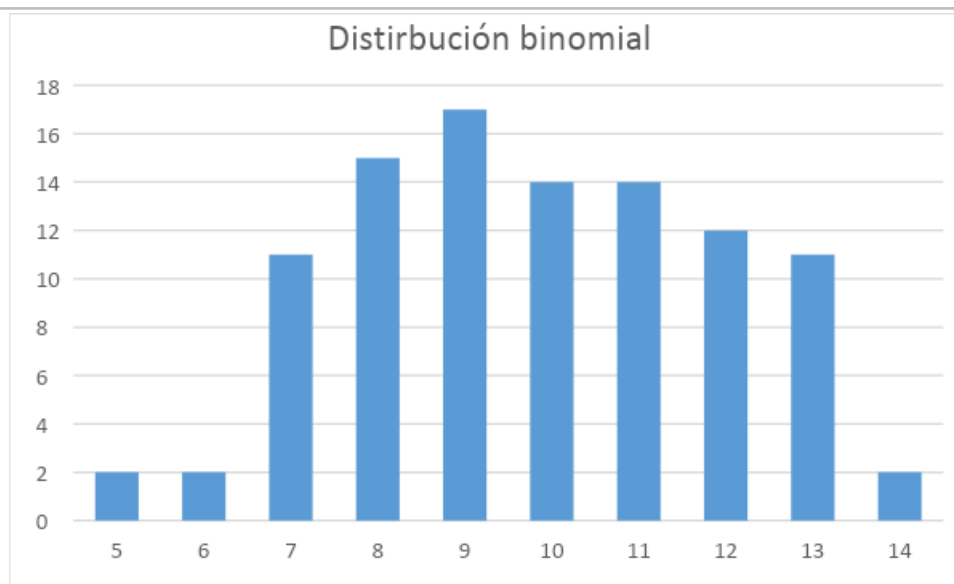
¿Cuántas mujeres aparecen en 100 muestras de 20 observaciones recogidas de una población con una probabilidad del 50% de ser mujer?

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

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



MEDIA	MEDIANA	DT	Q3	Q1	IQR
10,02	10,00	8,72	11,75	8,00	3,75

Guarda mucho parecido con la normal ¿no?

¡Bien hecho!

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

¡Enhorabuena!