

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.

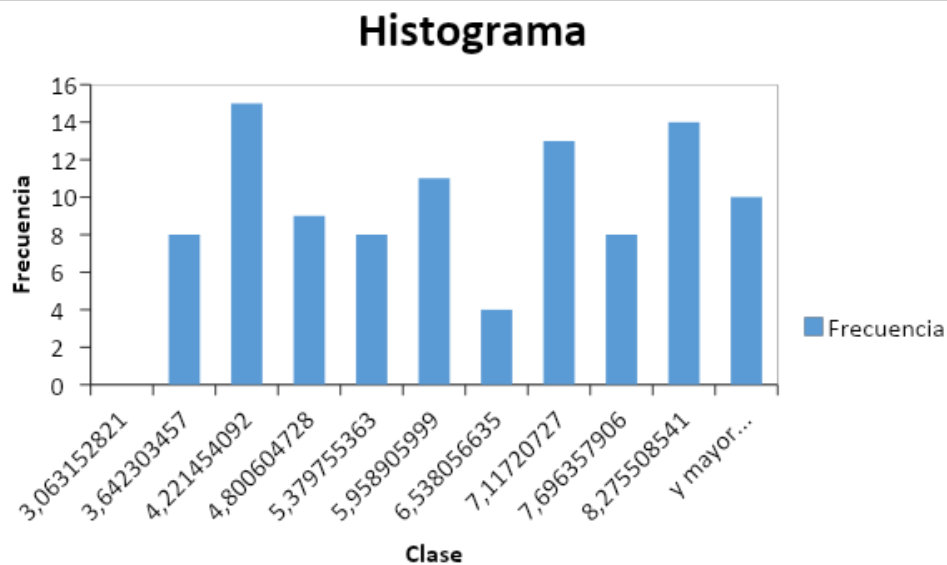
Distribución Uniforme (continua)

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.

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



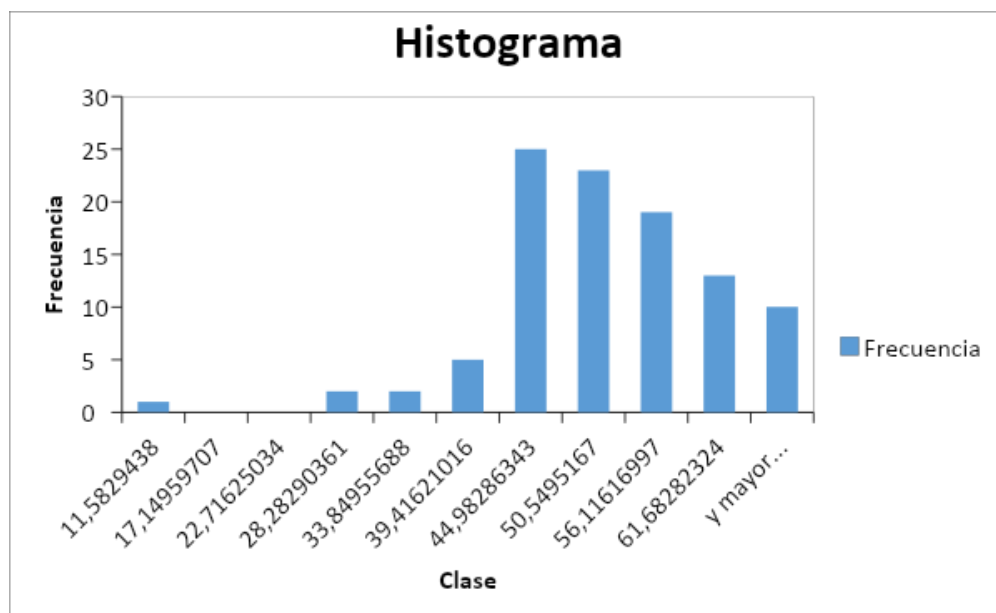
Promedio	6,522
Desvest	1,765
Mediana	6,636
Q3	8,093
Q1	5,120
Rango intercuartílico	2,973

Distribución Normal (continua)

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

- Media: 50
- Desviación Estándar: 10

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



Promedio	49,131
Desvest	9,533
Mediana	48,565
Q3	55,471
Q1	43,070
Rango intercuartílico	12,401

Distribución Exponencial (continua)

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

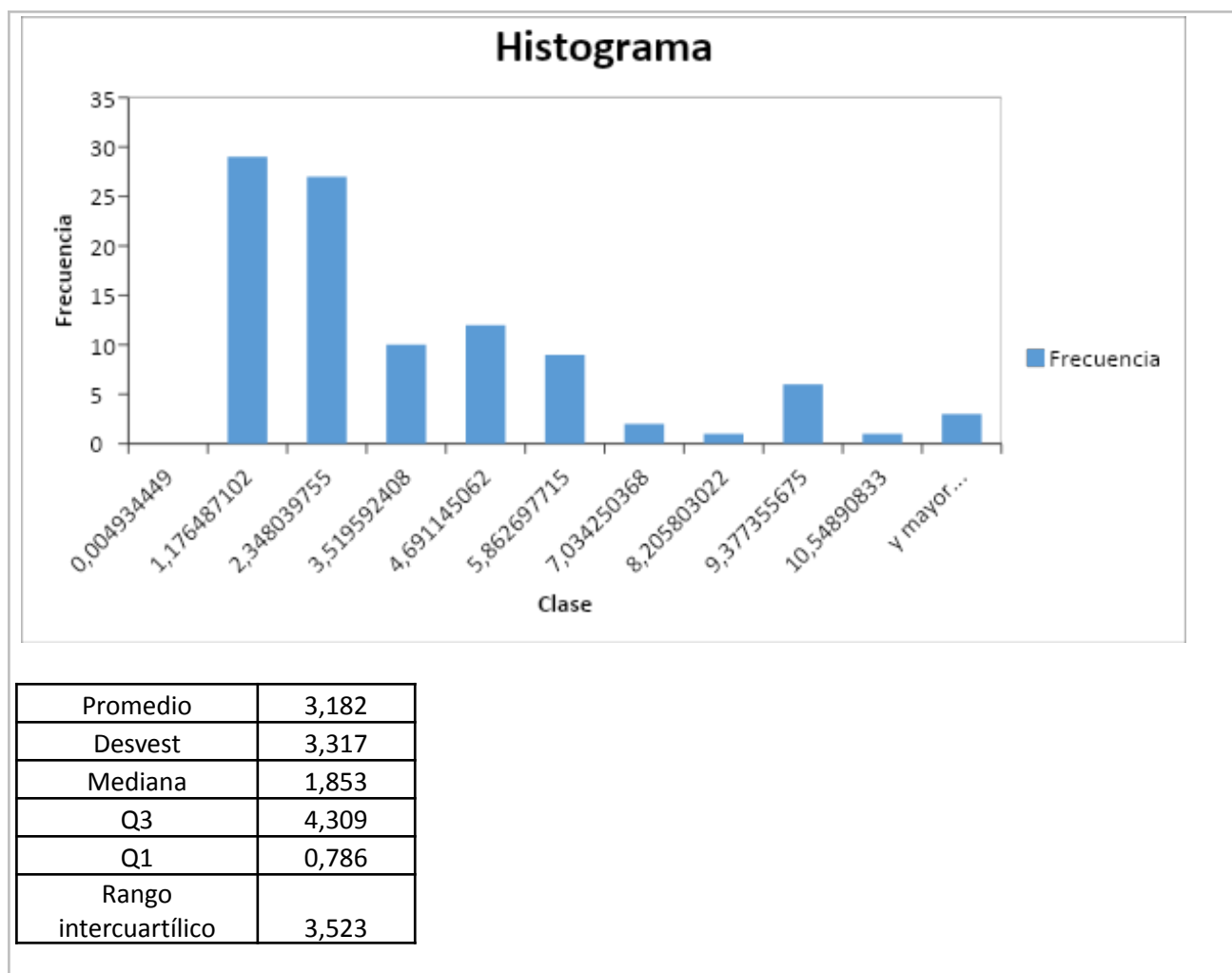
- Landa: 3

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

CREANDO MUESTRAS A PARTIR DE DISTRIBUCIONES



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

La distribución normal puesto que tiene esa forma de montaña, se parte de un valor medio y a partir de él se conoce la dispersión de los datos.

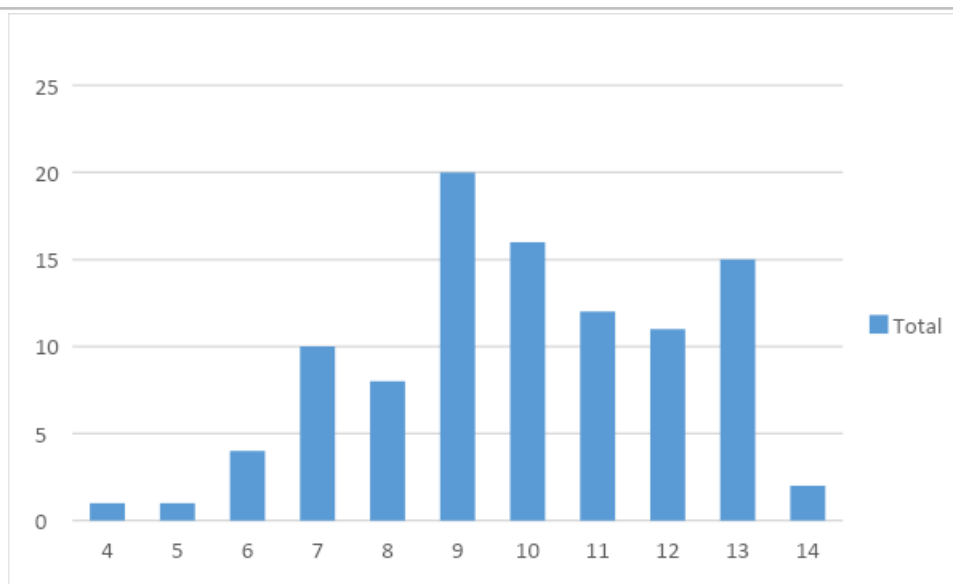
Distribución binomial (discreta)

¿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 Distribución binomial
4	1
5	1
6	4
7	10
8	8
9	20
10	16
11	12
12	11
13	15
14	2
Total general	100

Es una forma similar a la de la distribución normal, con forma de dos montaña.

¡Bien hecho!

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

¡Enhorabuena!