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.

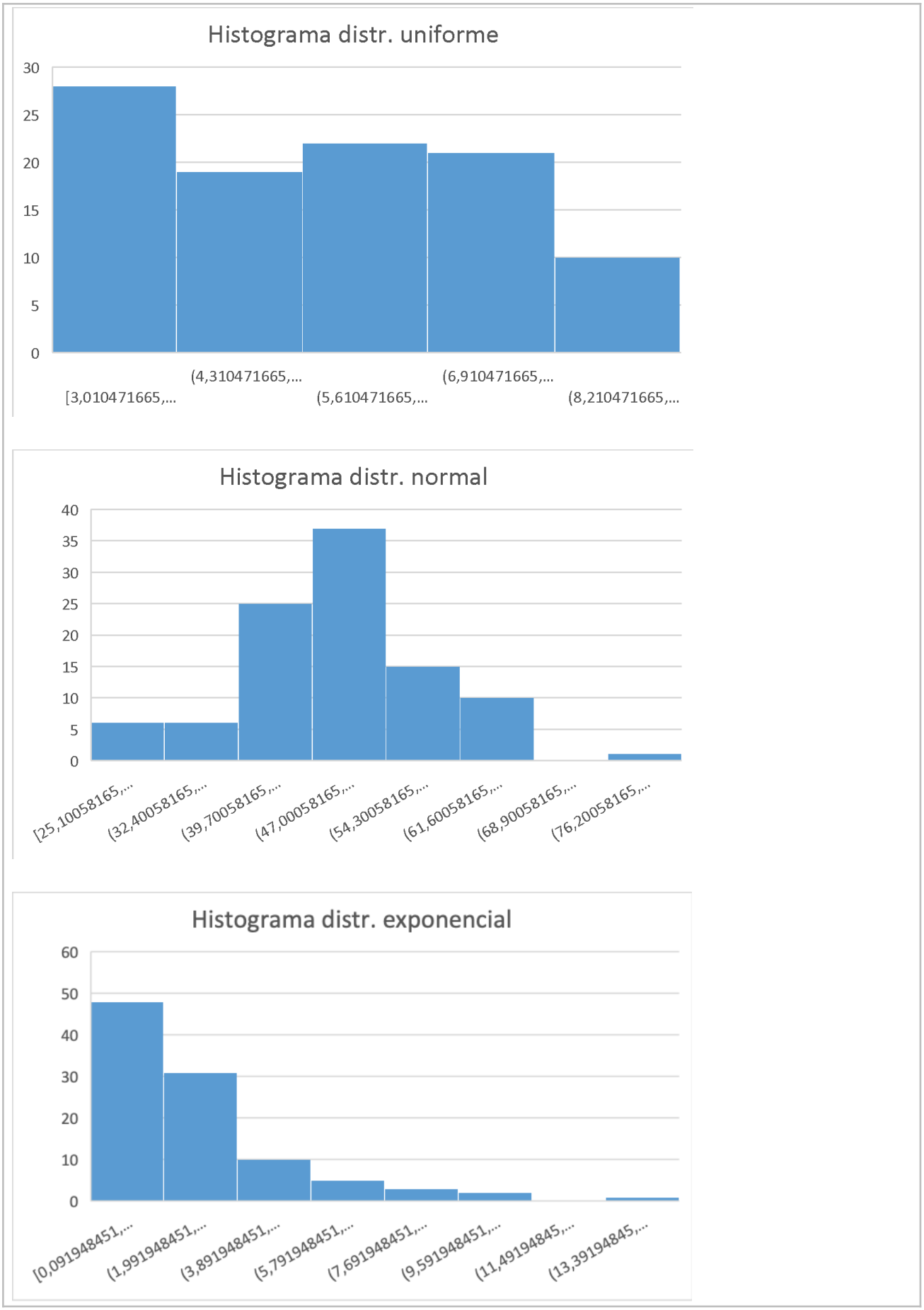
Distribución Uniforme (continua)

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.


Histograma distr. uniforme
30
25
20
15
10
5
0
[3,010471665,...
(4,310471665,...
(5,610471665,...
(6,910471665,...
(8,210471665,...
Histograma distr. normal
40
35
30
25
20
15
10
5
0
[25,10058165,...
(32,40058165,...
(39,70058165,...
(47,00058165,...
(54,30058165,...
(61,60058165,...
(68,90058165,...
(76,20058165,...
Histograma distr. exponencial
60
50
40
30
20
10
0
[0,091948451,...
(1,991948451,...
(3,891948451,...
(5,791948451,...
(7,691948451,...
(9,591948451,...
(11,49194845,...
(13,39194845,...

Característica		Distr. Uniforme	Distr. Normal	Distr. Exponencial
CENTRALIDAD	Media	5,81	49,61	2,69
	Mediana (cuartil 2)	5,71	49,71	2,09
DISPERSIÓN	Desviación Estándar	1,78	9,70	2,57
POSICIÓN del 50% de los datos más centrados	=Media-0.67*Desv. Estándar	4,61	43,11	0,96
	Cuartil 3	7,41	54,96	3,77
	=Media+0.67*Desv. Estándar	7,00	56,11	4,41
	Cuartil 1	4,07	44,23	0,74
POSICIÓN del 95% de los datos más centrados	=Media+1.96*Desv. Estándar	2,31	30,59	-2,36
	Percentil 97.5	8,90	68,70	9,96
	=Media-1.96*Desv. Estándar	41,39	243,30	54,04
	Percentil 2.5	3,12	25,96	0,15

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

Pues no lo sé, supongo que cada una refleja las características de la distribución de cada muestra y al ser diferentes distribuciones no se pueden comprar las medidas.

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.

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

Creo que la mejor medida es la de la distribución normal

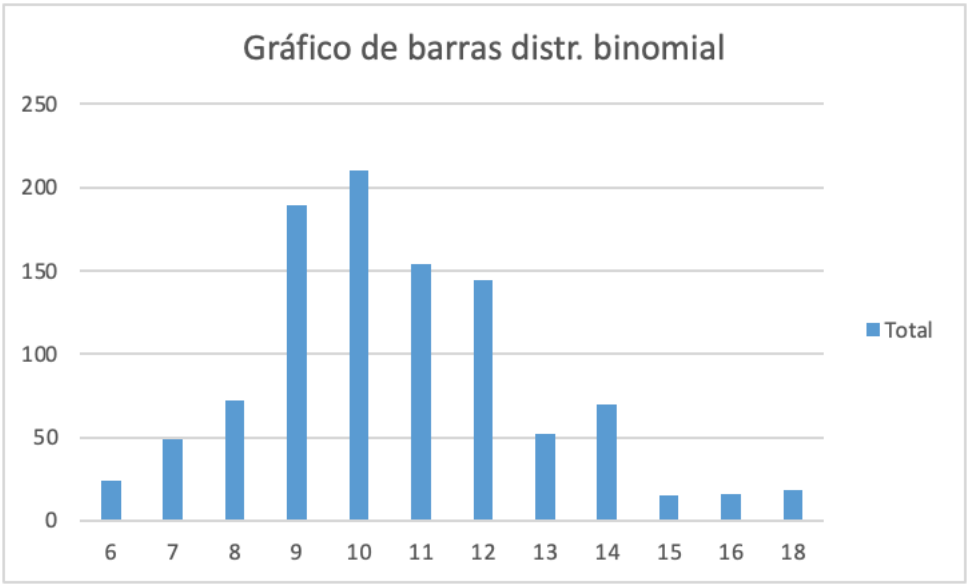
Distribución binomial (discreta)

¿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



Se intuye una forma de distribución normal

¡Bien hecho!

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

¡Enhorabuena!