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.

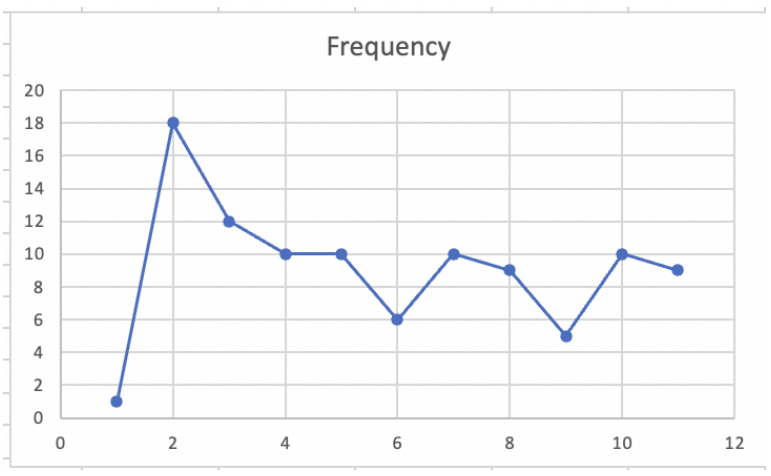
Distribución Uniforme (continua)

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.

Bin	*Frequency*
3,06563744	1
3,65402249	18
4,24240753	12
4,83079258	10
5,41917763	10
6,00756267	6
6,59594772	10
7,18433276	9
7,77271781	5
8,36110285	10
More	9



Mean	5,633037263
Median	5,390221045
Standard Deviation	1,796563079
Cuartil 1	3,950330569
Cuartil 3	7,032277422
IQR	3,081946853

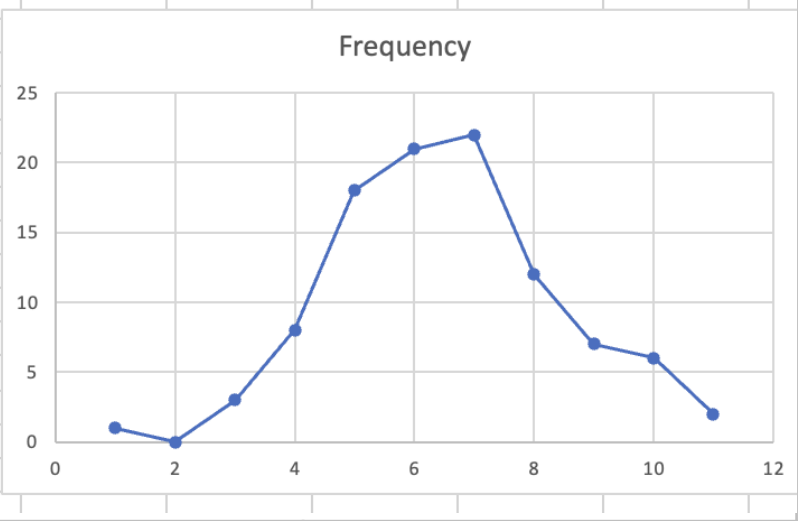
Distribución Normal (continua)

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 ☺)

Bin	Frequency
19,46233518	1
25,29079925	0
31,11926331	3
36,94772738	8
42,77619145	18
48,60465552	21
54,43311959	22
60,26158366	12
66,09004773	7
71,9185118	6
More	2



Mean	49,1540321
Median	47,8934066
Standard Deviation	10,8882222
Cuartil 1	41,8102243
Cuartil 2	55,1873528
IQR	13,3771285

Distribución Exponencial (continua)

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

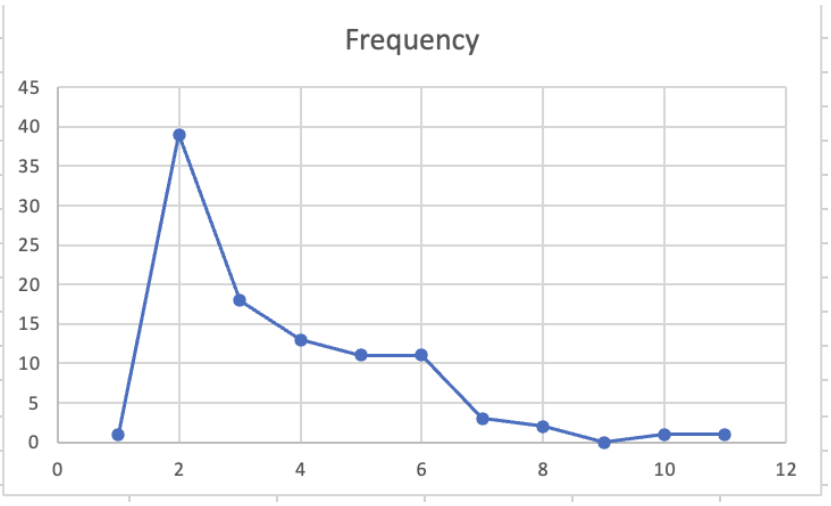
- Landa: 3

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

Bin	Frequency
0,00556264	1
1,1722119	39
2,33886116	18
3,50551042	13
4,67215968	11
5,83880894	11
7,0054582	3
8,17210746	2
9,33875672	0
10,505406	1
More	1



Mean	2,38764037
Median	1,58699502
Standard Deviation	2,32995421
Cuartil 1	0,48217283
Cuartil 3	3,66740455
IQR	3,18523172

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

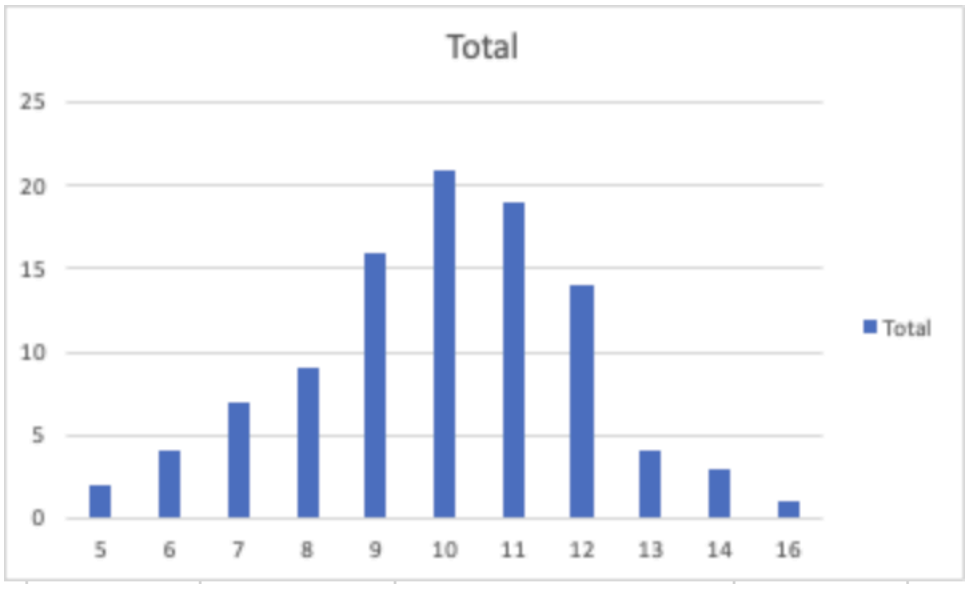
La distribución no tiene una forma normal, por tanto es aconsejable usar la mediana y el rango intercuartílico como medidas de centralidad y dispersión respectivamente.

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?



Muestra una distribución normal.

¡Bien hecho!

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

¡Enhorabuena!