CONOCIENDO LA DISTRIBUCIÓN NORMAL

Para conocer profundamente el concepto de distribución de densidad de probabilidad y más concretamente de distribución normal solo necesitas tocar los datos y ver qué está pasando.

Básicamente jugar un poco.

¡Vamos a ello!

La cuestión es que vas a crear tú mismo/a algunas distribuciones normales y las compararás entre sí.

Crea dos distribuciones normales (misma dispersión)

Crea una distribución normal “m1” con estas dos características:

- Valor central (media): 65 kg
- Dispersión (desviación estándar): 12 kg

Y otra distribución normal “m2”, con estos dos parámetros:

- Valor central (media): 42kg
- Dispersión (desviación estándar): 12 kg

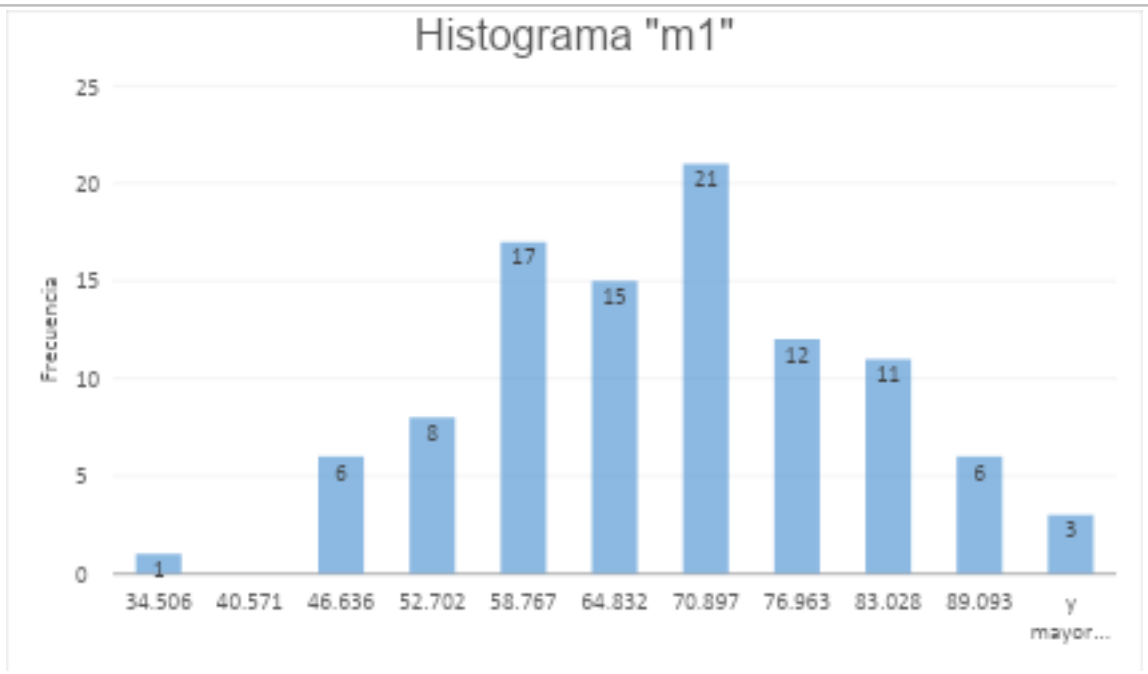
Para crear estas muestras puedes ver el vídeo Tech Zen para calcular muestras de distribuciones normales.

Dibuja los dos histogramas, calcula la media y la desviación estándar de las dos muestras de 100 observaciones y compara las dos distribuciones normales gráficamente y numéricamente.

Copias los dos histogramas y copia las medias y las desviaciones estándar que has calculado.

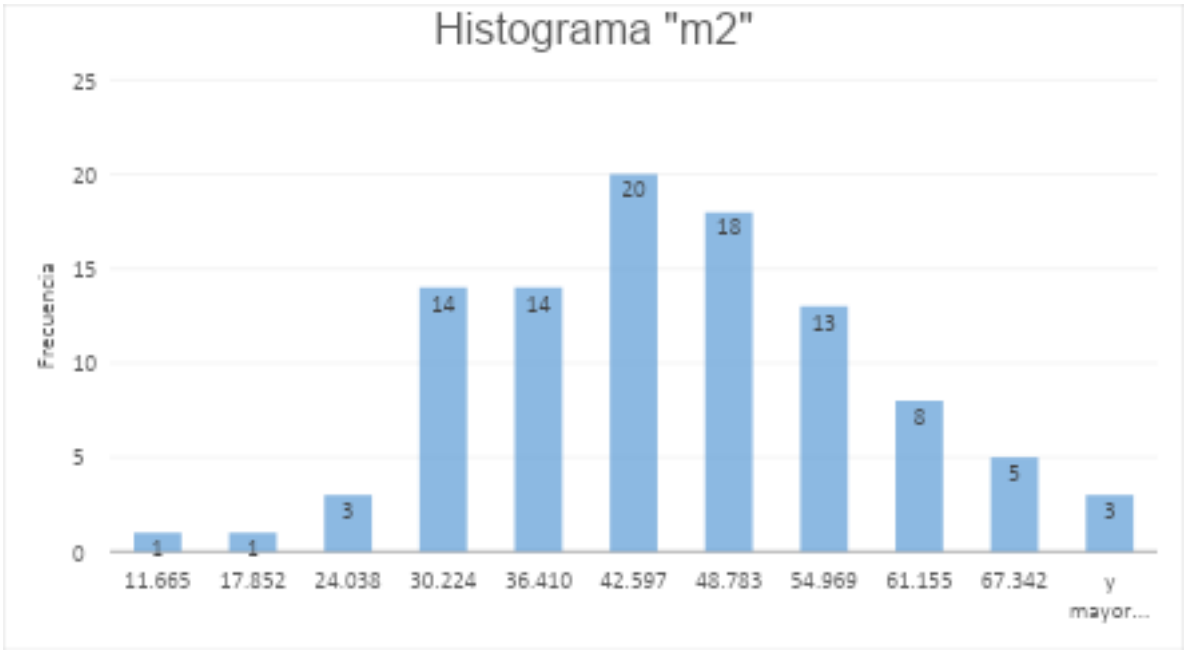
- Muestra m1:

MEDIA m1 = 65.597
DESVIACIÓN ESTÁNDAR m1 = 12.631



- Muestra m2:

MEDIA m2 = 42.276
DESVIACIÓN ESTÁNDAR m2 = 12.607



¿Qué observas?

Las variables “m1” y “m2” tienen histogramas simétricos y semejantes, ya que son distribuciones normales y tienen la misma dispersión (12). Sin embargo, la variable “m1” tiene una medida de centralidad mayor que la variable “m2”, por lo que su histograma está desplazado hacia la derecha respecto al de la variable “m2”.

Crea dos distribuciones normales (misma centralidad)

Crea otra distribución normal “m3”, con estos dos parámetros:

- Valor central (media): 65kg
- Dispersión (desviación estándar): 2 kg

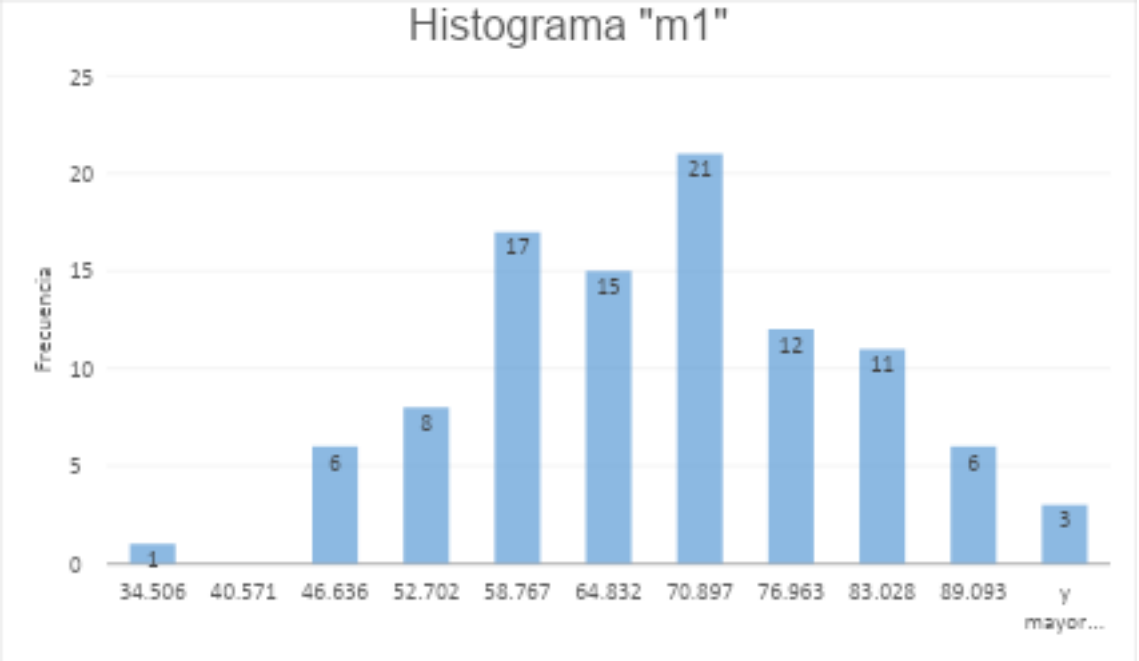
Para crear estas muestras puedes ver el vídeo Tech Zen para calcular muestras de distribuciones normales.

Dibuja los histogramas de la muestra “m1” y “m3”, calcula la media y la desviación estándar de las dos muestras de 100 observaciones y compara las dos distribuciones normales.

Copias los dos histogramas y copia las medias y las desviaciones estándar que has calculado.

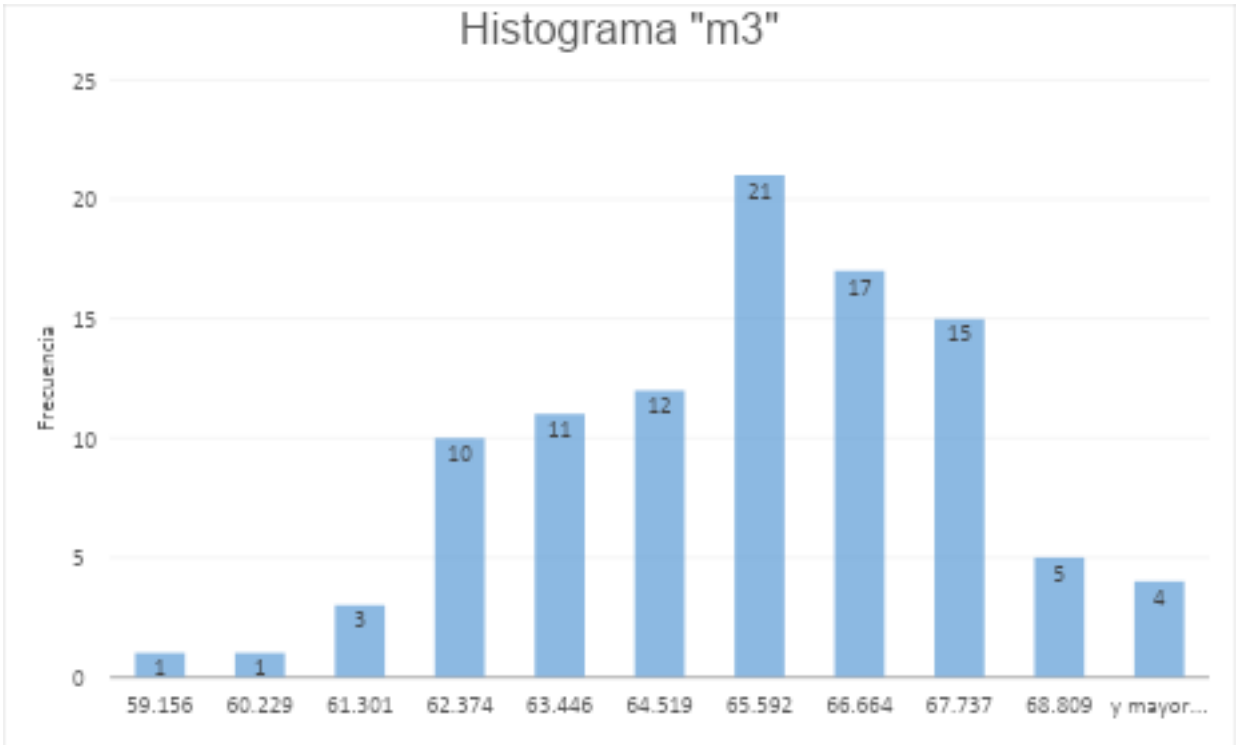
- Muestra m1:

MEDIA m1 = 65.597
DESVIACIÓN ESTÁNDAR m1 = 12.631



- Muestra m3:

MEDIA m3 = 65.003 Kg
DESVIACIÓN ESTÁNDAR m3 = 2.280 Kg



¿Qué observas?

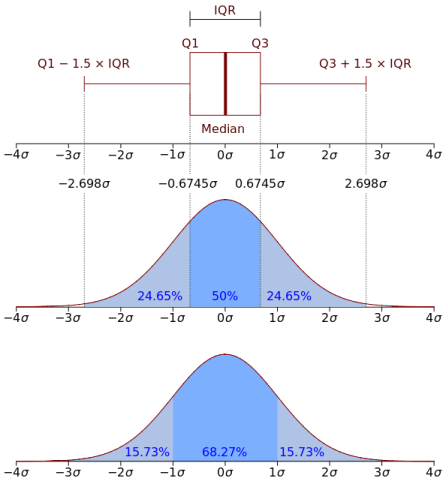
Las variables “m1” y “m3” tienen histogramas simétricos con la misma medida de centralidad (65), pero tienen diferente dispersión, ya que la variable “m1” tiene valor de desviación estándar mucho mayor que la variable “m3”.

Aplicando las reglas de la distribución normal

De las tres muestras que has creado calcula las características de posición con la ayuda de la tabla.

Recuerda que esta distribución es especial. A la derecha tienes un esquema para recordarte algunas características

Calcula los resultados siguientes:



Característica		Muestra "m1"	Muestra "m2"	Muestra "m3"
CENTRALIDAD	Media	65.597	42.276	65.003
	Mediana (cuartil 2)	66.679	41.922	65.051
DISPERSIÓN	Desviación Estándar	12.631	12.607	2.280
POSICIÓN del 50% de los datos más centrados	=Media-0.67*Desv. Estándar	57.134	33.829	63.475
	Cuartil 3	75.264	50.455	66.619
	=Media+0.67*Desv. Estándar	74.060	50.723	66.530
	Cuartil 1	56.159	32.632	63.276
POSICIÓN del 95% de los datos más centrados	=Media+1.96*Desv. Estándar	90.354	66.986	69.471
	Percentil 97.5	90.620	68.980	69.393
	=Media-1.96*Desv. Estándar	40.840	17.565	60.535
	Percentil 2.5	42.658	16.562	60.150
FORMA	Curtosis	-0.492	-0.229	-0.368
	Asimetría	0.019	0.144	-0.219

¿Qué patrón se repite en esta tabla?

Pista: media y mediana deberían ser parecidas, casi iguales

Para crear esta tabla utiliza las fórmulas que te enseñe en hojas de trabajo anteriores ☺

La media y la mediana son prácticamente iguales en las tres variables, ya que todas son distribuciones normales.
Además, el 50% de los datos de las tres variables están aproximadamente entre el cuartil 1 y el cuartil 3 (como son distribuciones normales, se puede aplicar la fórmula: Media ± 0.67*Desviación Estándar).
Por otro lado, el 95% de los datos de las tres variables están aproximadamente entre el percentil 97.5 y el percentil 2.5 (como son distribuciones normales, se puede aplicar la fórmula: Media ± 1.96*Desviación Estándar).
Los valores de asimetría y curtosis son muy cercanos a cero en todas las variables, lo que nos indica que todas tienen distribuciones bastante simétricas y mesocúrticas.

Aplicando la regla del 2

En la práctica normalmente se utiliza la regla del 2. (en lugar del 1.96 que has visto en la tabla anterior.

CONOCIENDO LA DISTRIBUCIÓN NORMAL

Si una variable cuantitativa presenta un histograma parecido a una distribución normal (con forma de montaña y simétrico) se aplica:

El 95% de los datos más centrados están entre la media +/- 2 veces la desviación estándar.

Es una regla efectiva para entender los límites de la variable rápidamente:

Característica		Muestra "m1"	Muestra "m2"	Muestra "m3"
POSICIÓN del 95% de los datos más centrados. Fórmula Exacta	=Media+1.96*Desv. Estándar	90.354	66.986	69.471
	Percentil 97.5	90.620	68.980	69.393
	=Media-1.96*Desv. Estándar	40.840	17.565	60.535
	Percentil 2.5	42.658	16.562	60.150
Regla del 2. POSICIÓN del 95% de los datos	=Media+2*Desv. Estándar	90.859	67.490	69.562
	=Media-2*Desv. Estándar	40.335	17.061	60.443

¡Buen trabajo!

¡Hoy has entendido con ejemplos la distribución normal y el concepto de distribución!