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:

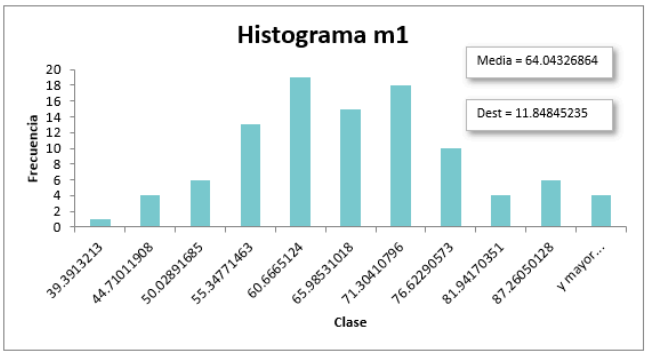
- Valor central (media): 65 kg
- Dispersión (desviación estándar): 12 kg

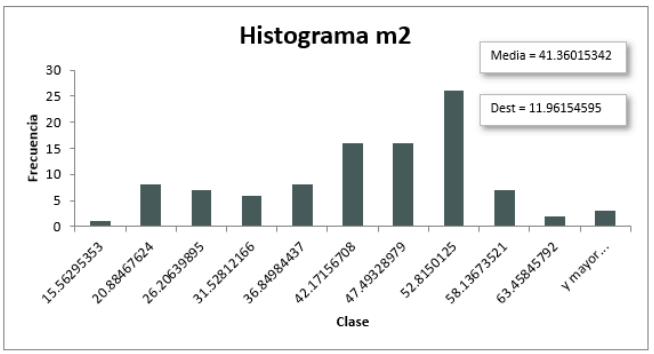
Y otra distribución normal “m2”, con estos dos parámetros:

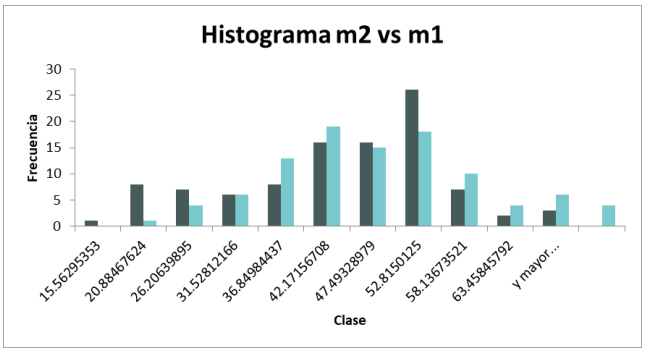
- Valor central (media): 42kg
- Dispersión (desviación estándar): 12 kg

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.







¿Qué observas?

La muestra 1 tiene forma de campana alta y la muestra 2 tiene forma de campana achatada. Al tener la misma desviación estándar pero diferente media las gráficas también varían en su altura. Asimismo, los datos se trasladan hacia la izquierda de la gráfica cuando la media es menor.

Crea dos distribuciones normales (misma centralidad)

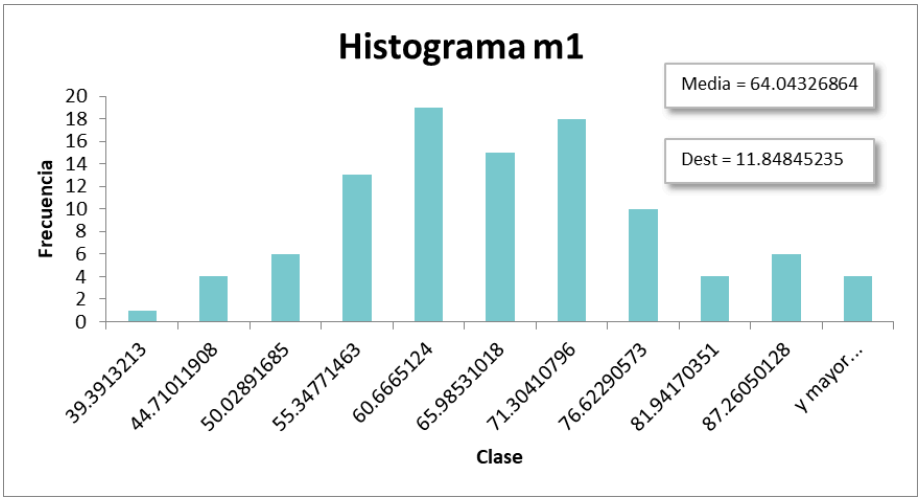
Crea otra distribución normal “m3”, con estos dos parámetros:

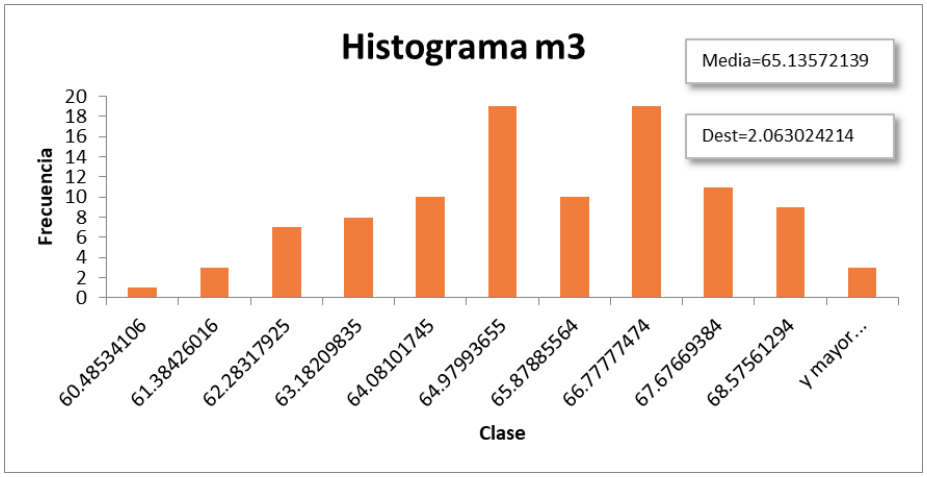
- Valor central (media): 65kg
- Dispersión (desviación estándar): 2 kg

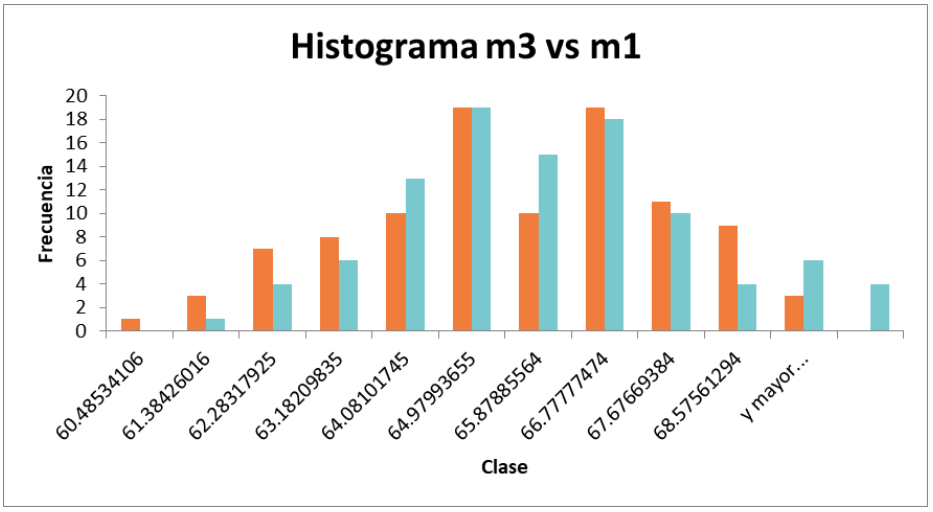
Para crear estas muestras puedes ver el vídeo Tech Zen para calcular muestras de distribuciones normales.

CONOCIENDO LA DISTRIBUCIÓN NORMAL

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.







¿Qué observas?

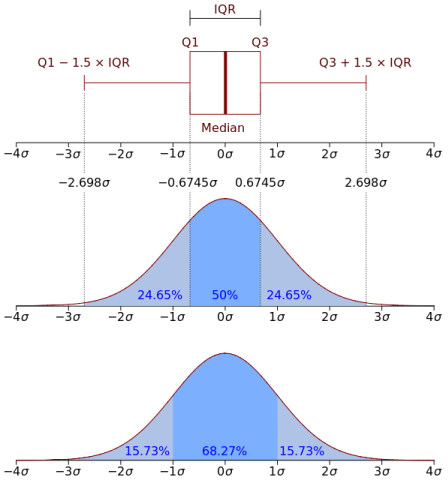
En m3 las medias son iguales y la desviación de la muestra es menor. Existe un leve desplazamiento hacia la izquierda, pero los valores que se encuentra en las colas del histograma tienen mayor magnitud que las de m1. Los valores centrales de m1 parecen tener mayor magnitud.

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	64.0433	41.3602	65.1357
	Mediana (cuartil 2)	63.1391	43.6613	65.1553
DISPERSIÓN	Desviación Estándar	11.8485	11.9615	2.0630
POSICIÓN del 50% de los datos más centrados	=Media-0.67*Desv. Estándar	56.1048	33.3459	63.7535
	Cuartil 3	71.2592	49.7942	66.7073
	=Media+0.67*Desv. Estándar	71.9817	49.3744	66.5179
	Cuartil 1	56.4252	33.5781	63.6521
POSICIÓN del 95% de los datos más centrados	=Media+1.96*Desv. Estándar	87.2662	64.8048	69.1792
	Percentil 97.5	89.3238	64.7253	68.7192
	=Media-1.96*Desv. Estándar	40.8203	17.9155	61.0922
	Percentil 2.5	41.9469	17.8461	60.8086
FORMA	Curtosis	-0.3562	-0.4771	-0.6643
	Asimetría	0.2352	-0.3632	-0.2437

¿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 ☺

En todas las muestras las medidas de centralidad son similares en m3 como la dispersión es menor los resultados son prácticamente iguales.
En el rango Inter cuartil o la posición del 50% de los datos y 95% de los datos los intervalos de confianza de igual forma son similares.
M2 con media más baja tiene datos más ajustados.
En cuanto a la forma la muestra con media más alta (m1) tiene un índice de asimetría mayor con signo positivo se puede asociar con tener una mayor concentración de datos al centro de la gráfica.

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.

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	87.2662	64.8048	69.1792
	Percentil 97.5	89.3238	64.7253	68.7192
	=Media-1.96*Desv. Estándar	40.8203	17.9155	61.0922
	Percentil 2.5	41.9469	17.8461	60.8086
Regla del 2. POSICIÓN del 95% de los datos	=Media+2*Desv. Estándar	87.74017335	65.28324533	69.26176982
	=Media-2*Desv. Estándar	40.34636394	17.43706152	61.00967296

¡Buen trabajo!

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