

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

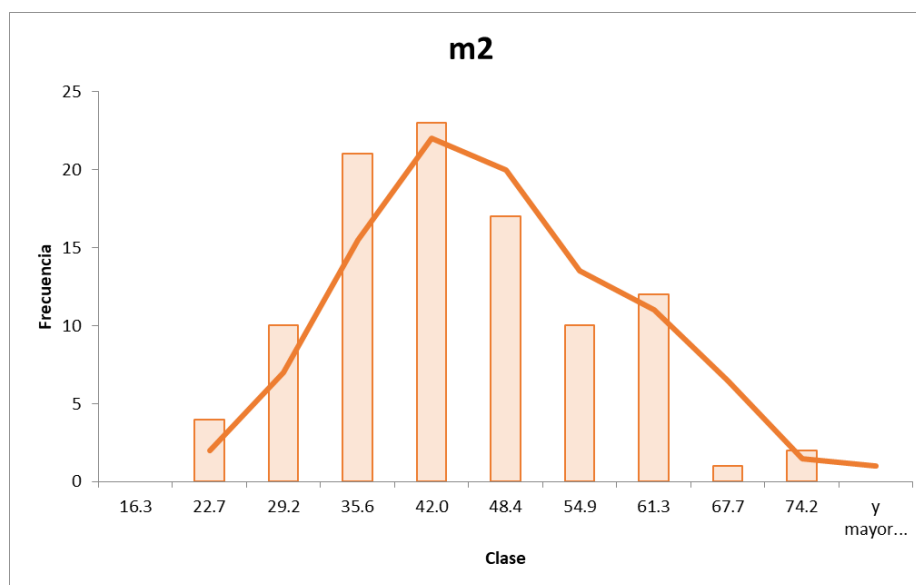
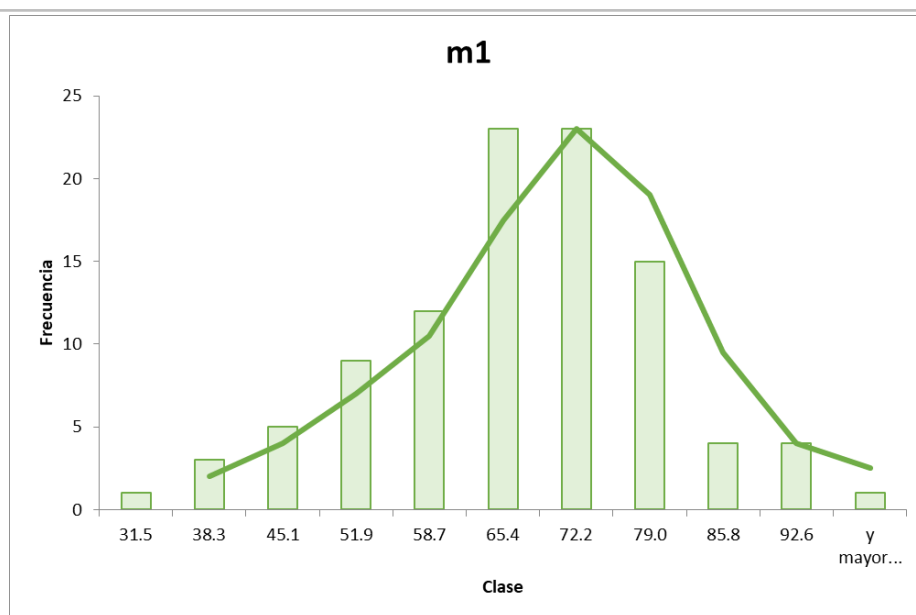
Para crear estas muestras puedes ver el vídeo Tech Zen para calcular muestras de distribuciones normales.

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
- Muestra m2

CONOCIENDO LA DISTRIBUCIÓN NORMAL



Análisis descriptivo:

	m1	m2
Medi		
a	63.73	41.24
SD	12.92	11.43

¿Qué observas?

En los histogramas se aprecia que ambas variables siguen una distribución normal, es decir más o menos simétrica y siguiendo una campana de Gauss. En ambos casos se observa una dispersión de los datos parecida (que concuerda con la desviación estándar obtenida en el análisis descriptivo calculado), pero existe una distinta

centralidad, siendo mayor en la variable m_1 que en m_2 , como también se obtiene en la media de ambas variables calculadas en el análisis descriptivo.

Crea dos distribuciones normales (misma centralidad)

Crea otra distribución normal " m_3 ", con estos dos parámetros:

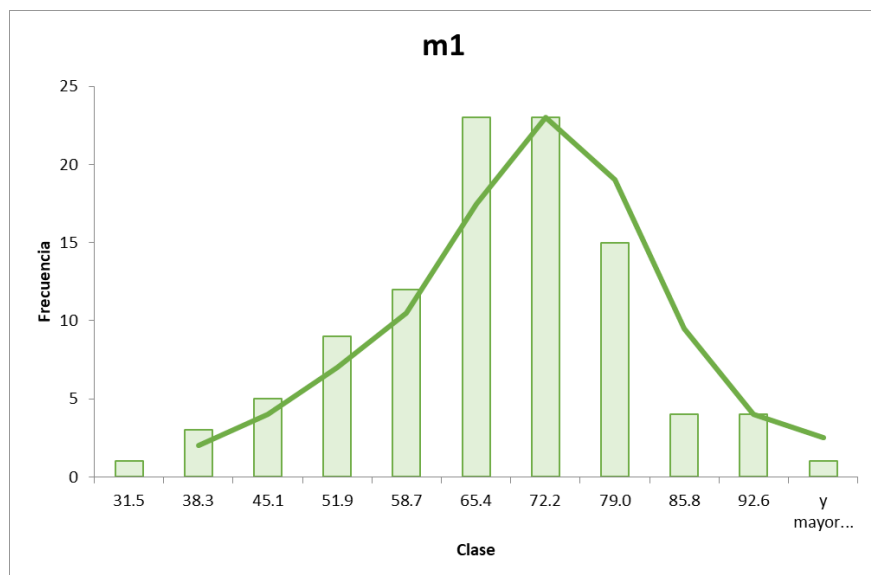
- Valor central (media): 65kg
- Dispersión (desviación estándar): 2 kg

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

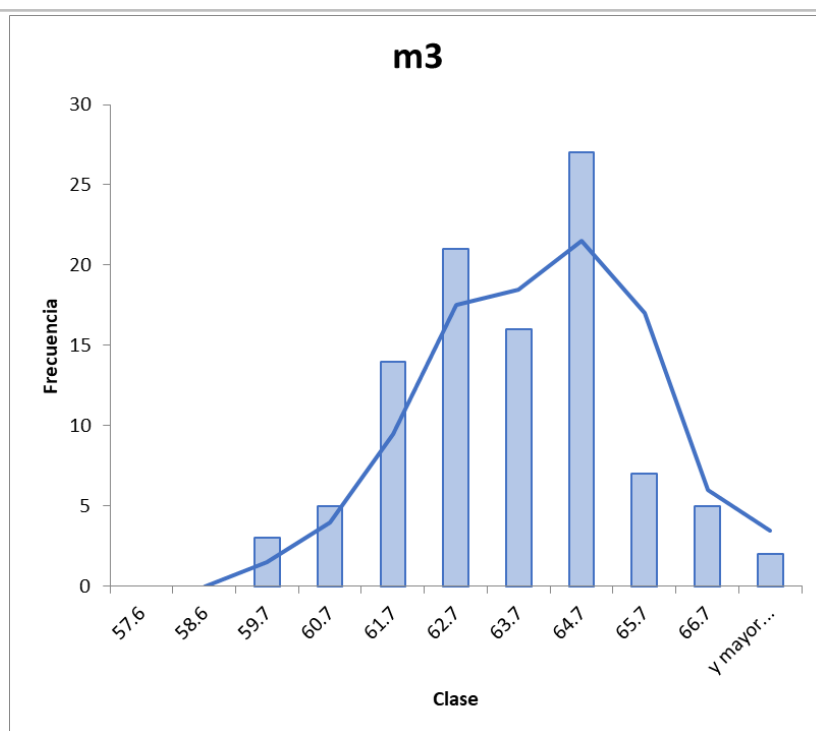
Dibuja los histogramas de la muestra " m_1 " y " m_3 ", calcula la media y la desviación estándar de las dos muestras de 100 observaciones y compara las dos distribuciones normales.

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

- Muestra m_1
- Muestra m_3



CONOCIENDO LA DISTRIBUCIÓN NORMAL



Análisis descriptivo:

	m1	m3
Media	63.73	63.32
SD	12.92	2.01

¿Qué observas?

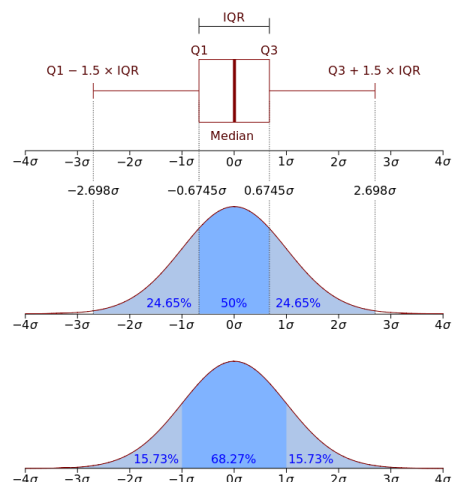
En los histogramas se aprecia que ambas variables siguen una distribución normal, más o menos simétrica y siguiendo una campana de Gauss. En ambos casos se observa una centralidad parecida (que concuerda con la media obtenida en el análisis descriptivo calculado), pero parece observarse una dispersión diferente entre ellas, siendo mayor en la variable m1 que en m3, ya que las barras del histograma parecen estar más abiertas a lo largo del eje X en la primera variable, lo que indica más dispersión en los datos que en m3. El dato numérico que lo acompaña es la comparación de las desviaciones estándar calculadas en el análisis descriptivo.

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	63.73	42.30	62.85
	Mediana (cuartil 2)	65.10	43.14	62.74
DISPERSIÓN	Desviación Estándar	12.92	11.98	1.80
POSICIÓN del 50% de los datos más centrados	=Media-0.67*Desv. Estándar	55.07	34.27	61.65
	Cuartil 3	71.49	51.09	64.04
	=Media+0.67*Desv. Estándar	72.39	50.33	64.06
	Cuartil 1	57.00	34.06	61.63
POSICIÓN del 95% de los datos más centrados	=Media+1.96*Desv. Estándar	89.06	65.78	66.37
	Percentil 97.5	91.71	64.87	66.48
	=Media-1.96*Desv. Estándar	38.40	18.82	59.34
	Percentil 2.5	36.20	15.90	59.60
FORMA	Curtosis	0.29	-0.22	-0.06
	Asimetría	-0.08	-0.21	0.17

¿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 las 3 variables se observa el mismo patrón. Una similitud entre los valores de la media y la mediana; entre el límite superior donde se encuentran el 50% de los datos y el cuartil 3; y entre el límite inferior donde se encuentran el 50% de los datos y el cuartil 1. Lo mismo ocurre con los percentiles 97.5 y el límite superior donde se encuentran el 95% de los datos y el percentil 2.5 y el límite inferior de esta área. Esto ocurre, debido a que la muestra sigue una distribución normal, y tanto el boxplot como el histograma en

CONOCIENDO LA DISTRIBUCIÓN NORMAL

estas distribuciones son simétricos, provocando así que los valores que limitan los porcentajes de la muestra sean muy parecidos.

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	89.06	64.54	66.68
	Percentil 97.5	91.71	63.83	66.83
	=Media-1.96*Desv. Estándar	38.40	18.86	59.50
	Percentil 2.5	36.20	14.98	59.58
Regla del 2. POSICIÓN del 95% de los datos	=Media+2*Desv. Estándar	89.57	65.01	66.76
	=Media-2*Desv. Estándar	37.88	18.39	59.42

¡Buen trabajo!

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