

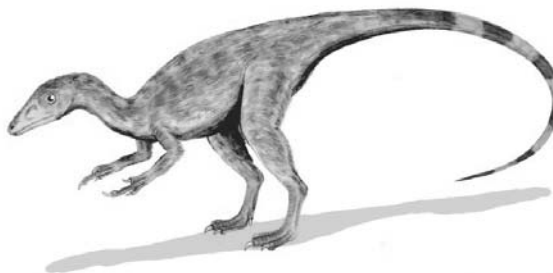
## Lección 9: Uso de la Curva para Modelar la Distribución de Datos

### Trabajo en clase

#### Ejemplo 1: Las Alturas de los Dinosaurios y la Curva

##### Normal

Un paleontólogo estudia la vida prehistórica ya veces trabaja con fósiles de dinosaurios. La siguiente tabla muestra la distribución de las alturas (redondeado a la pulgada más cercana) de 660 procompsognathus o "compis".

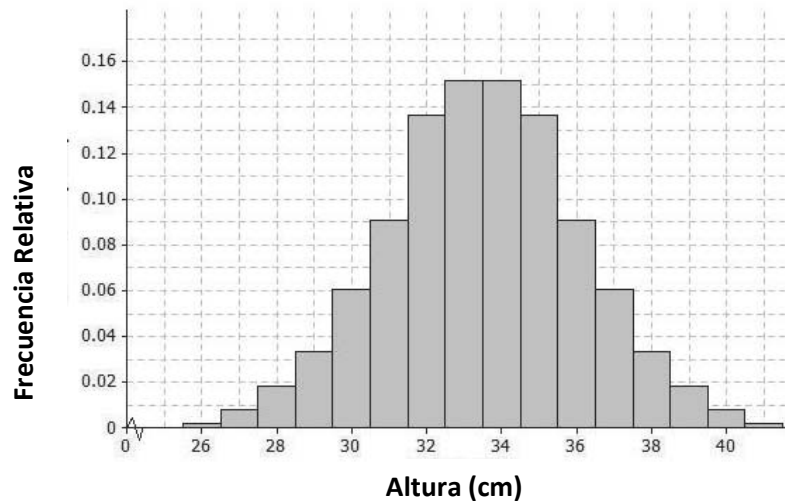


Las alturas se determinaron mediante el estudio de los restos fósiles de los compis.

Altura (cm)	Número de compis	Frecuencia Relativa
26	1	0.002
27	5	0.008
28	12	0.018
29	22	0.033
30	40	0.061
31	60	0.091
32	90	0.136
33	100	0.152
34	100	0.152
35	90	0.136
36	60	0.091
37	40	0.061
38	22	0.033
39	12	0.018
40	5	0.008
41	1	0.002
Total	660	1.00

## Ejercicios 1–8

El siguiente es un histograma de frecuencias relativas de las alturas de los compis.



1. ¿Qué significa la frecuencia relativa de la 0.136 nueva para la altura de 32 cm?
2. ¿Cuál es el ancho de cada barra? ¿Qué representa la altura de la barra?
3. ¿Cuál es el área de la barra que representa la frecuencia relativa de los compis con una altura de 32 cm?
4. La media de la distribución de alturas de los compis es de 33.5 cm, y la desviación estándar es de 2.56 cm. Interpreta la media y la desviación estándar en este contexto.

5. Marca la media en la gráfica y marca una desviación por encima y por debajo de la media.
- a. Aproximadamente, ¿qué porcentaje de valores de este conjunto de datos se encuentran dentro de una desviación estándar de la media? (Es decir, entre  $33.5 - 2.56 = 30.94$  cm y  $33.5 + 2.56 = 36.06$  cm.)
- b. Aproximadamente ¿qué porcentaje de valores de este conjunto de datos se encuentran dentro de dos desviaciones estándar de la media?
6. Dibuja una curva suave que pase bastante cerca a través de los puntos medios de las cimas de las barras del histograma. Describe la forma de la distribución.
7. Sombrea el área bajo la curva que representa la proporción de alturas que están dentro de la desviación estándar de la media.
8. Basándote en nuestro análisis, cómo responderías a la pregunta, "¿Qué tan alto era un compis?"

**Ejemplo 2: Rendimiento de la Gasolina y Distribución Normal**

Una curva normal es una curva suave que es simétrica y con forma de campana. Las distribuciones de datos que se encuentran en forma de campana a menudo se modelan mediante una curva normal, y nos dicen que tal distribución es aproximadamente normal. Un ejemplo de una distribución que es aproximadamente normal es la distribución de las alturas de los Compis del Ejemplo 1. Las distribuciones que son aproximadamente normales ocurren en muchos entornos diferentes. Por ejemplo, un vendedor mantiene un registro del rendimiento de la gasolina de su automóvil en un lapso de 25 semanas.

El millaje (millas por galón redondeado al número entero más cercano) fue

23, 27, 27, 28, 25, 26, 25, 29, 26, 27, 24, 26, 26, 24, 27, 25, 28, 25, 26, 25, 29, 26, 27, 24, 26.

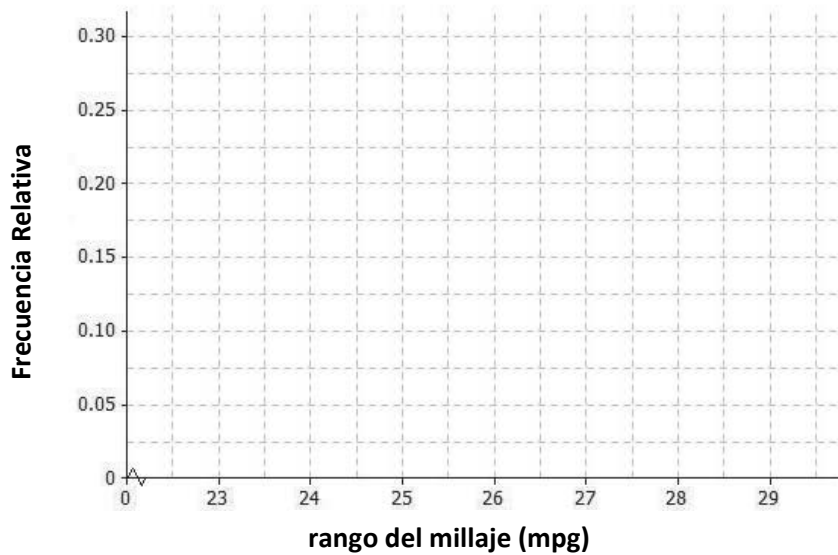
**Ejercicio**

9. Considera lo siguiente:

- Usando la tecnología para encontrar la media y la desviación estándar de los datos del millaje. ¿Cómo utilizaste la tecnología para ayudarte?
- Calcula la frecuencia relativa de cada uno de los valores del millaje. Por ejemplo, el millaje de 26 mpg tiene una frecuencia de 7. Para encontrar la frecuencia relativa, divide 7 en 25, el número total del millaje registrado. Completa la siguiente tabla.

Millaje	Frecuencia	Frecuencia Relativa
23		
24		
25		
26	7	
27		
28		
29		
Total	25	

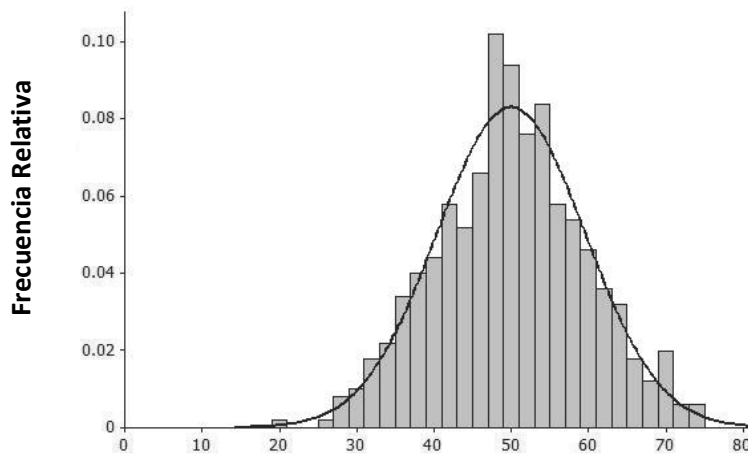
- c. Construye un histograma de frecuencias relativas utilizando la escala siguiente.



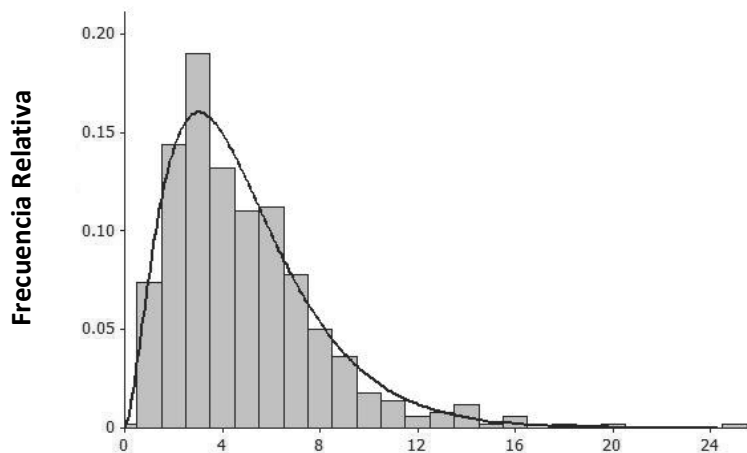
- d. Describe la forma de la distribución del millaje. Dibuja una curva suave que pase bastante cerca a través de los puntos medios de las cimas de las barras del histograma. ¿Es esto aproximadamente una curva normal?
- e. Marca la media en el histograma. Marca una desviación estándar a la izquierda y derecha de la media. Sombrea el área bajo la curva que representa la proporción de datos dentro de una desviación estándar de la media. Encuentra la proporción de datos dentro de una desviación estándar de la media.

## Resumen de la Lección

- Una curva normal es simétrica y con forma de campana. La media de una distribución normal se encuentra en el centro de la distribución. Las áreas bajo la curva normal se pueden usar para estimar la proporción de los valores de datos que caen dentro de un intervalo dado.

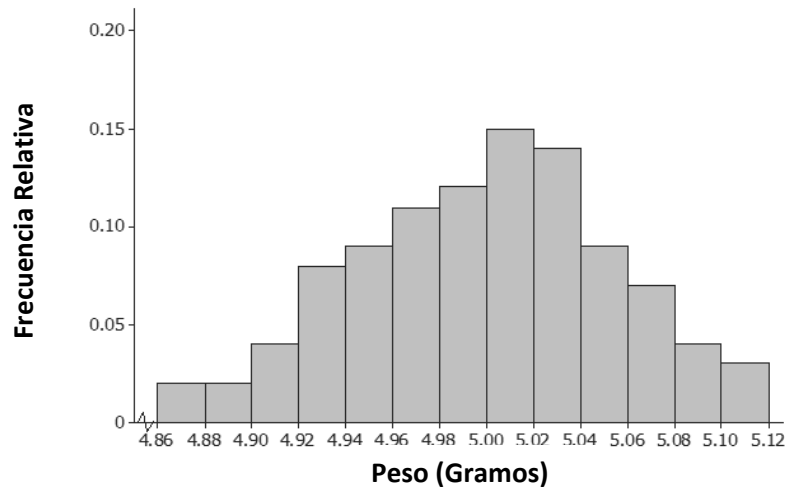


- Cuando una distribución es sesgada, no es apropiado modelar la distribución de los datos con una curva normal.



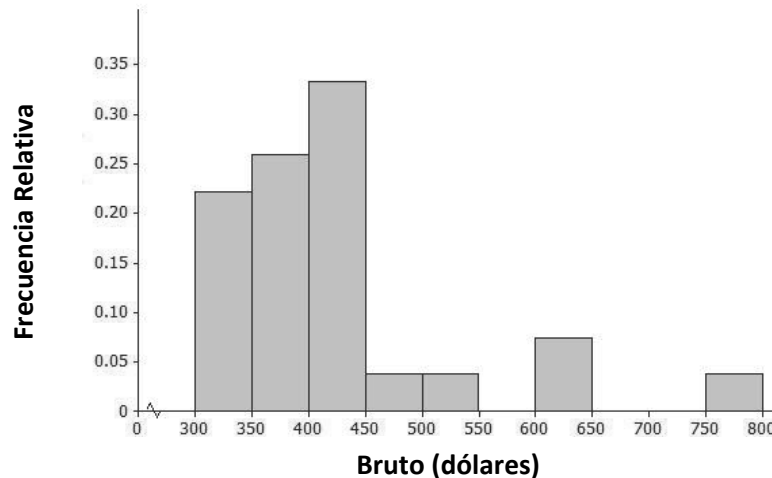
## Conjunto de Problemas

1. Periódicamente los EE.UU. Verifica el peso de las monedas recién acuñadas. A continuación se muestra un histograma de los pesos (en gramos) de una muestra aleatoria de 100 nuevas monedas de cinco centavos.



- La media y la desviación estándar de la distribución de los pesos de níquel son 5.00 gramos y 0.06 gramos, respectivamente. Marca la media en el histograma. Marca una desviación estándar por encima de la media y una desviación estándar por debajo de la media.
- Describe la forma de la distribución. Dibuja una curva suave que pase bastante cerca a través de los puntos medios de las cimas de las barras del histograma. ¿Es esto aproximadamente una curva normal?
- Sombrea el área bajo la curva que representa la proporción de datos dentro de una desviación estándar por encima y por debajo de la media. Encuentra la proporción de datos dentro de una desviación estándar por encima y por debajo de la media.

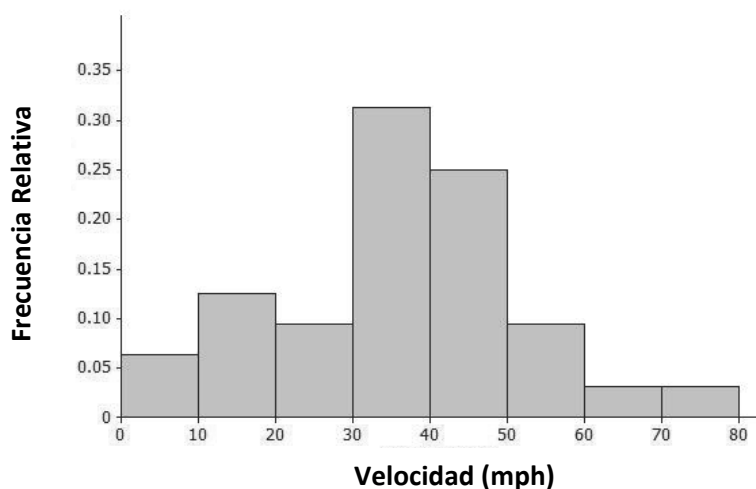
2. A continuación se muestra un histograma de frecuencia relativa de los ingresos brutos (en millones de dólares) para el punto más alto de las películas taquilleras Americanas (a partir de finales de 2012). Bruto es la cantidad total de dinero hecha antes de restar los gastos, como los gastos de publicidad y salarios de los actores.



- Describe la forma de la distribución del punto mas alto de las películas taquilleras de todos los tiempos . ¿Sería una curva normal la mejor para modelar esta distribución? Explica tu respuesta.
- ¿Cuál de las siguientes es una estimación razonable para la media de la distribución? Explica tu elección.
  - 325 millones
  - 375 millones
  - 425 millones
- ¿Cuál de las siguientes es una estimación razonable de la desviación estándar de la muestra? Explica tu elección.
  - 50 millones
  - 100 millones
  - 200 millones



3. A continuación se muestra un histograma de la velocidad máxima de diferentes tipos de animales.



- Describe la forma de la distribución de velocidad máxima.
- Estima la media y la desviación estándar de esta distribución. Describe cómo realizaste tu estimación.
- Dibuja una curva suave que sea aproximadamente una curva normal. La media y la desviación estándar real de este conjunto de datos son 34.1 y 15.3. Sombrea el área bajo la curva que representa la proporción de los valores de datos que están dentro de una desviación estándar de la media.