

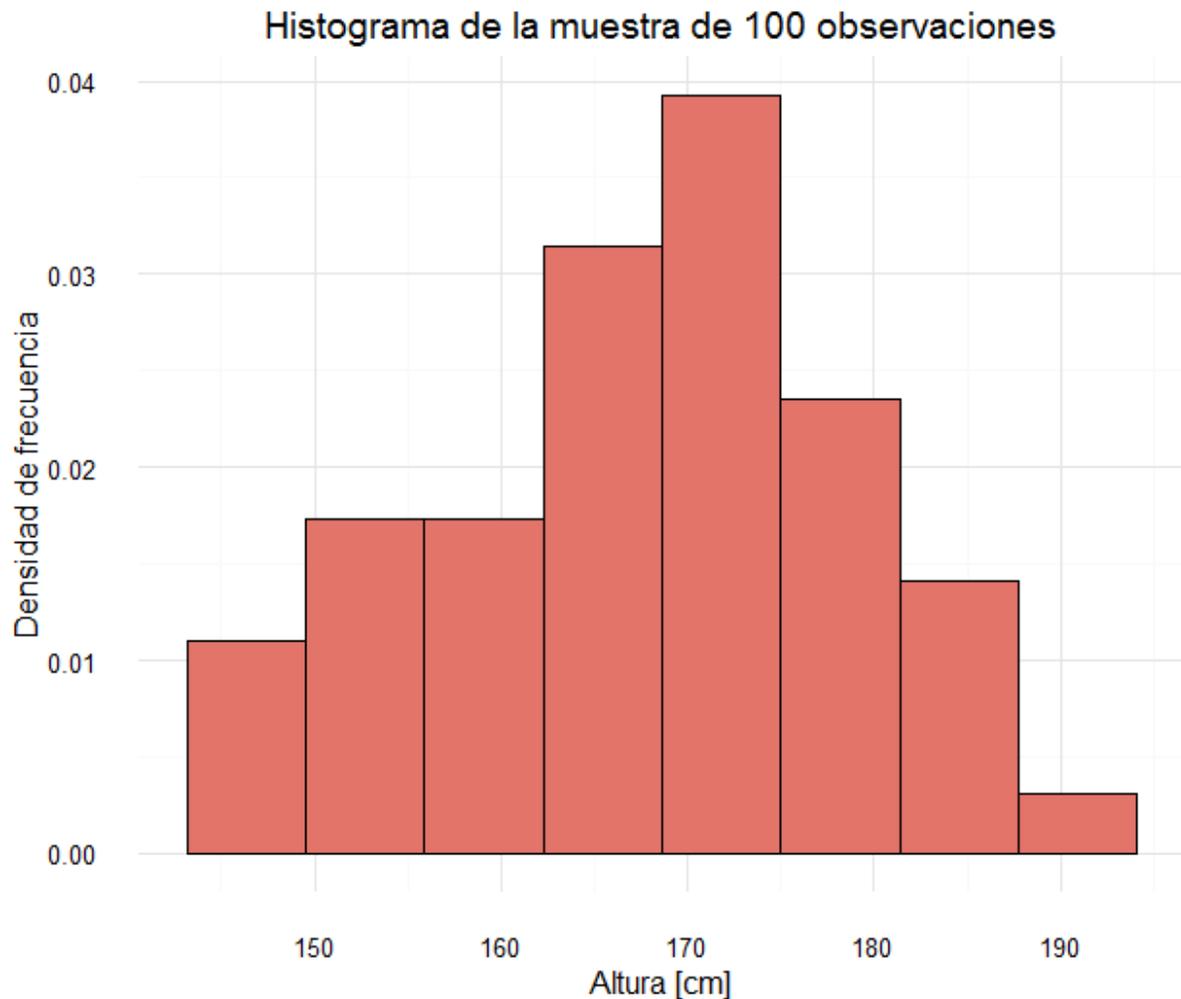
Función de Densidad de Probabilidad - Explicación intuitiva

Ejemplo: Distribución de la población de una gran ciudad

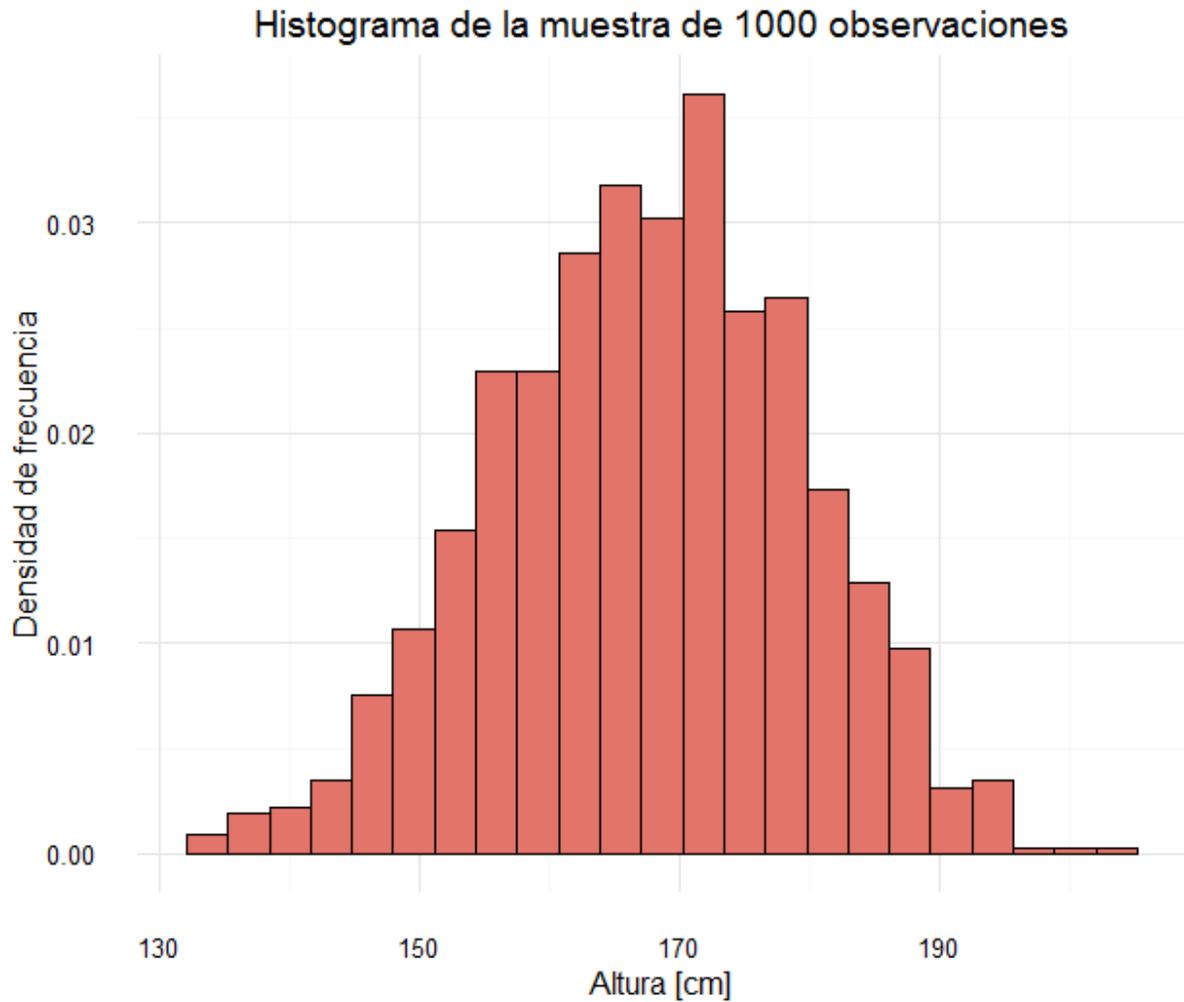
Imagínate que quieres estudiar la altura de las personas de una población de una gran ciudad. La variable numérica es la altura de las personas en cm.

En esta ciudad son 2M de personas. Quieres ver como se distribuye la altura en cm de todas las personas que habitan en la gran ciudad. Pero recoger todos los datos de la población es imposible. Como dice un colega mío «¡Tardarías 1000!»

Así que decides encuestar a 100 personas al azar (es decir escoges una muestra aleatoria) y miras el histograma de densidad. (Recuerda que el de densidad significa que la suma de las áreas del histograma suma 1)



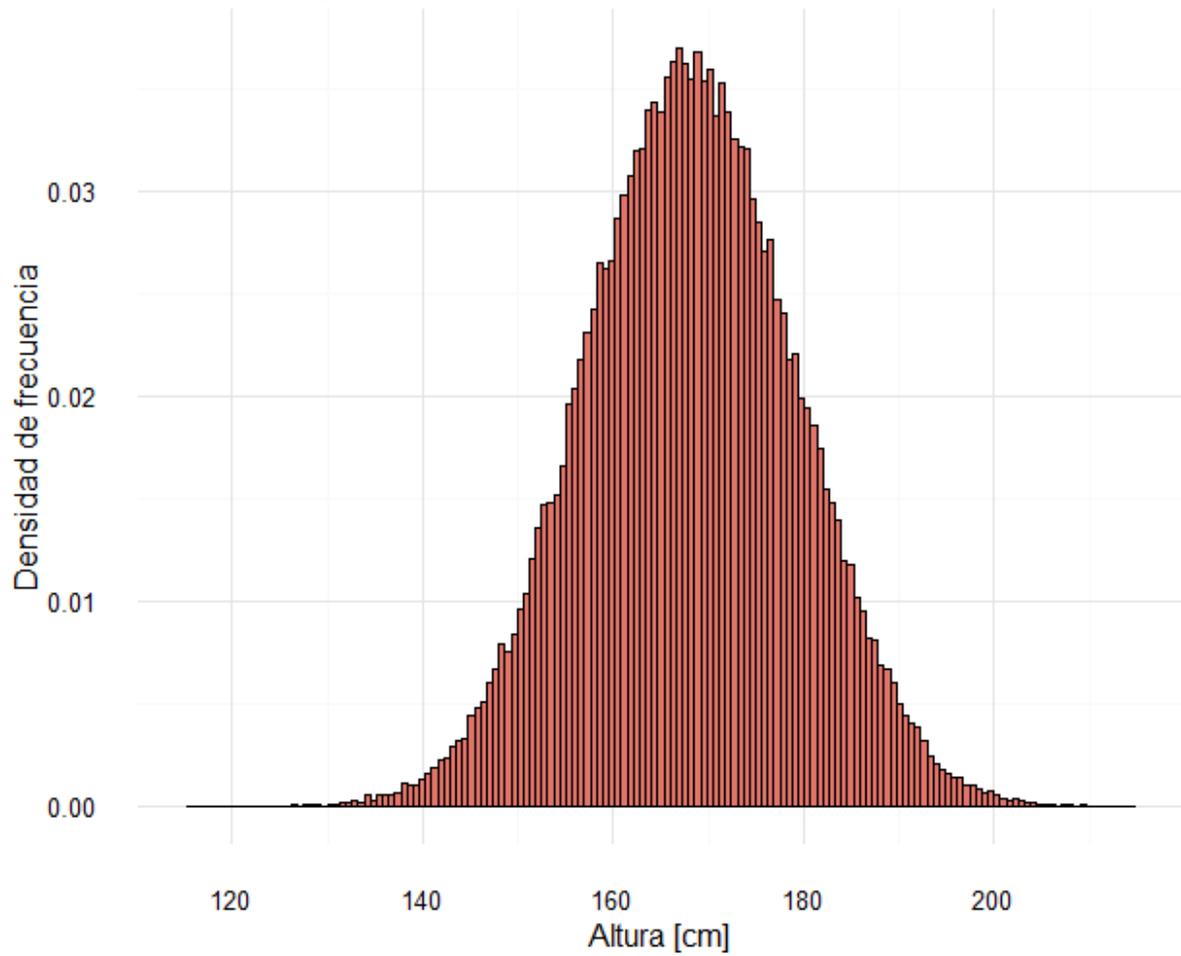
Pero ves que aún no te queda claro como se distribuyen las alturas de las personas. Crees que son pocas las personas que has encuestado. Te sientes con energía y encuestas a 1000 personas. Y pintas el histograma de densidad.



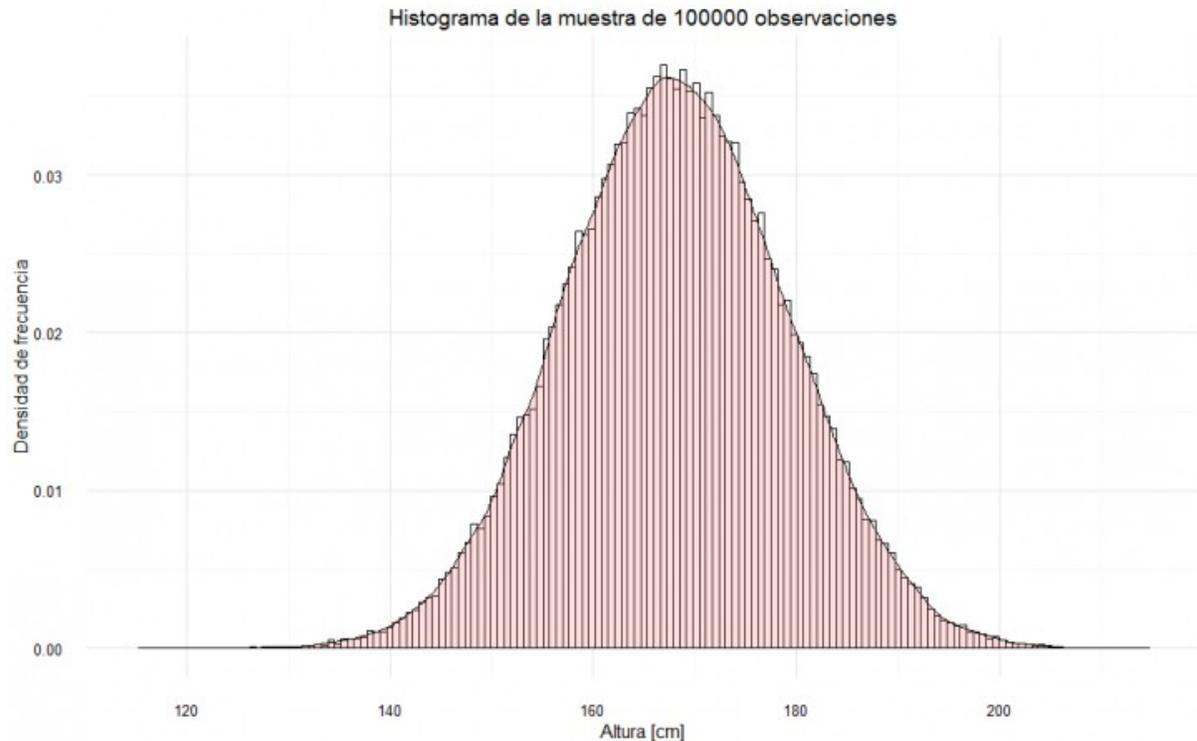
Fíjate como cada vez la distribución de tu variable es cada vez más suave y puedes intuir mejor como es la forma de la distribución. De hecho, de la densidad de la distribución.

Imagínate que eres capaz de obtener los valores de la altura de 100.000 personas. Y pintas el histograma de densidad.

Histograma de la muestra de 100000 observaciones



Ahora puedes ver como eres capaz de un contorno mucho más fino. Podrías dibujar el contorno fácilmente. ¡Mira esta imagen!



La función de densidad es precisamente este contorno. Es una línea continua que representa la distribución de densidad de TODA LA POBLACIÓN.

Como te he comentado el área de toda las barras del histograma suma 1. ¿Puedes adivinar qué área tiene la función de densidad? Fácil, 1. Es la probabilidad total de la distribución.

Sea X =altura de una persona elegida al azar. Suponemos que X sigue la función de densidad de la imagen. ¿Cómo calculamos $P(160 < X < 170)$? Fácil, integramos la función de densidad entre 160 y 170. ¿Cómo calculamos la esperanza de X ? Fácil, en vez de pensar en sumar todos los valores del rango multiplicados por su probabilidad puntual como haríamos en el caso de las variables discretas (por ejemplo, si pensábamos)