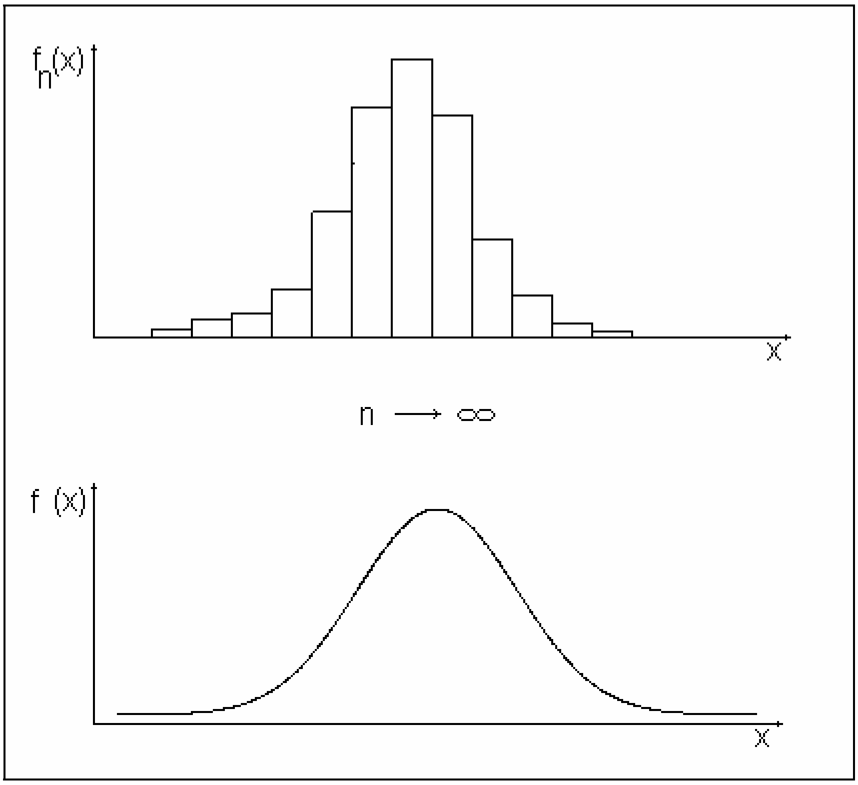
## Histogramas de Áreas - Curvas de densidad.

En un histograma de áreas, la altura de cada rectángulo de clase representa cuantos puntos fueron observados en el intervalo de clase por unidad, esto es la densidad de puntos por unidad y el área total es 1.

Si los valores que obtenemos provienen de una variable continua, y aumentamos el tamaño de la muestra, (siempre que los datos estén elegidos en forma adecuada si (muestra aleatoria)), podremos reducir la longitud de los intervalos de clase. De esta manera, a medida que aumenta la cantidad de datos las alturas de los intervalos va  formando una curva cada vez más suave. La curva límite es llamada curva de densidad de la variable correspondiente.



## Estimación de una curva de densidad.

Los histogramas se utilizan frecuentemente para describir datos para los cuales no se ha realizado ningún tipo de supuesto. Si los datos se modelan como una muestra aleatoria proveniente de alguna distribución continua. el histograma de áreas puede ser considerado como un estimador de la función de densidad de probabilidad. Desde este punto de vista tiene el inconveniente de no ser suave.

Un estimador de densidad de probabilidad puede ser construido de la siguiente manera. Consideremos una función no negativa, simétrica y centrada en cero w(x), función de peso, cuya integral sea igual a 1.
Por ejemplo w(x) puede ser la densidad normal estándar. La función

$$w_h(x) = \frac{1}{h} w(\frac{x}{h})$$

es una versión de w cambiada de escala. A medida que h tiende a cero $w_h$ se vuelve más picuda y concentrada alrededor de cero. A medida que h tiende a infinito $w_h$ se achata y dispersa. Cuando w es la densidad normal estándar, $w_h$ es la densidad normal con desvío h. Si ($x_1$, ..., $x_n$ ) son valores correspondientes a una muestra aleatoria de tamaño n de una densidad f, una estimación de f es:

$$f_h(x) = \frac{1}{h} \sum_{i=1}^{n} w_h(x - x_i)$$

Esta estimación es llamada, estimación por núcleos de la función de densidad (kernel estimation).

El parámetro h (ancho de la ventana) controla su suavidad y corresponde a la longitud del intervalo de clase en un histograma. Si h es demasiado pequeña la estimación resulta demasiado rugosa. Si h es demasiado grande la estimación resulta demasiado desparramada.

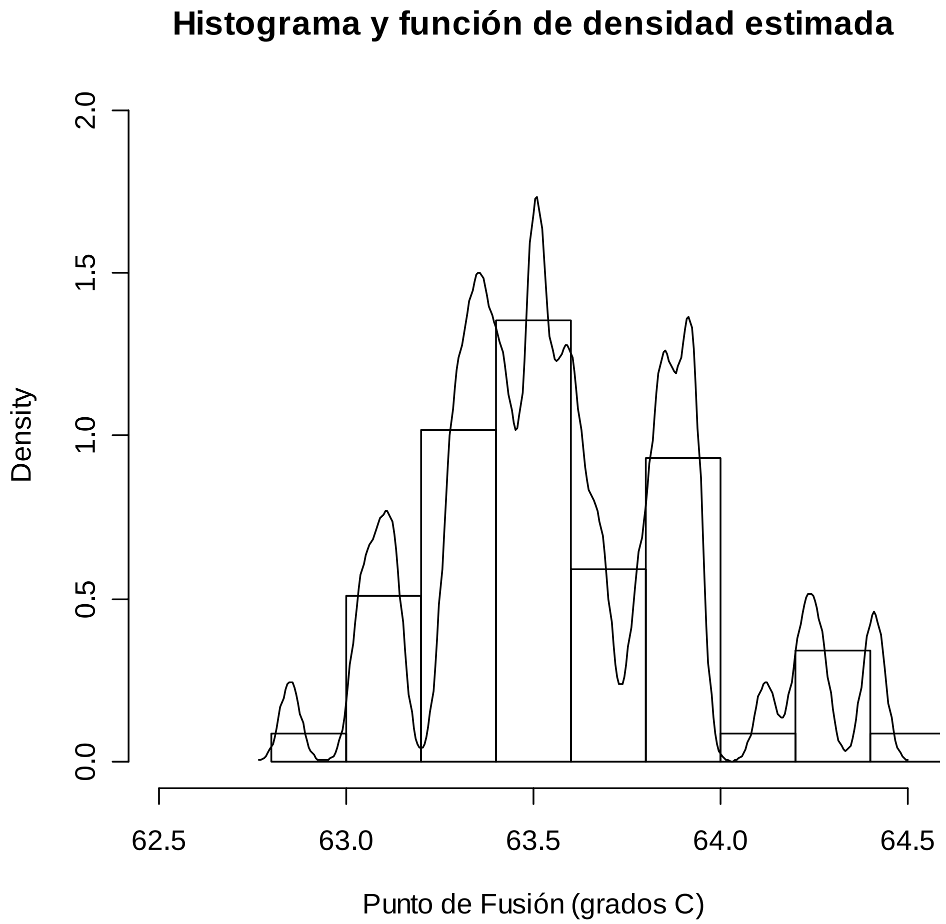



**Figura 3.** Histograma junto con una función de densidad estimada

La figura 3 muestra el histograma junto con una función de densidad estimada que, aunque es más suave que el histograma, es demasiado "rugosa". Fue obtenida mediante las siguientes instrucciones:

```
hist(Cera $CERA,probability=T, ylim=c(0,2),
  xlim=c(62.5,64.5),main="",
  xlab="Punto de Fusión (grados C)")
  lines(density(Cera $CERA,n=300,width=0.11))
   title ( "Histograma y función de densidad estimada")
```

Los gráficos de la figura 4 fueron obtenidos mediante las siguientes instrucciones:

```
 par(mfrow=c(3,1))
 plot(density(Cera
$CERA,n=500,width=0.11,from=62,to=66))
 plot(density(Cera$CERA,n=500,width=0.30,from=62,
to=66))
 plot(density(Cera$CERA,n=500,width=0.60,from=62,
to=66))
```

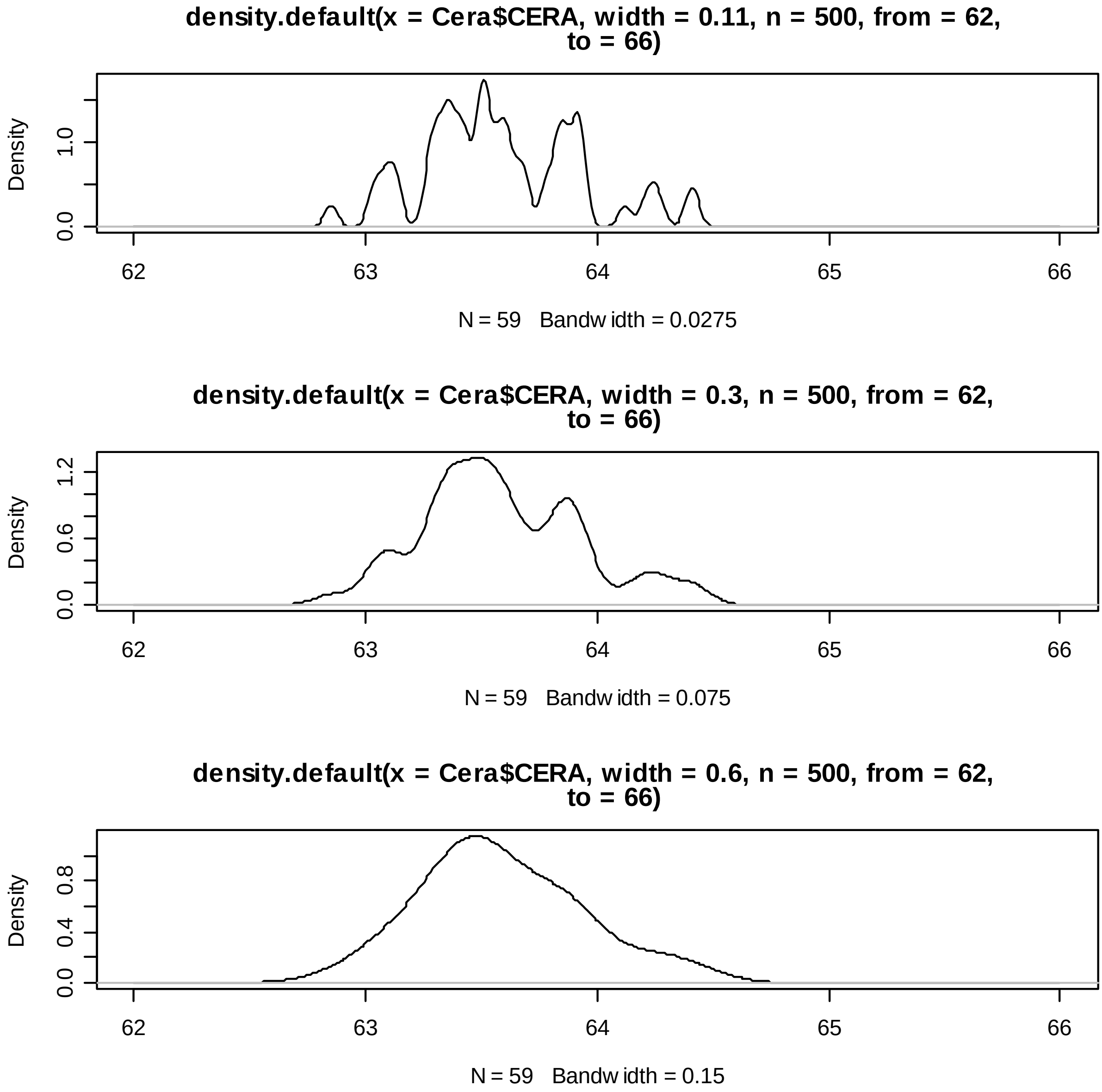



**Figura 4**. Estimaciones de densidad de probabilidad para los datos del punto de fusión de ceras naturales, para distintos tamaños del ancho de la ventana.