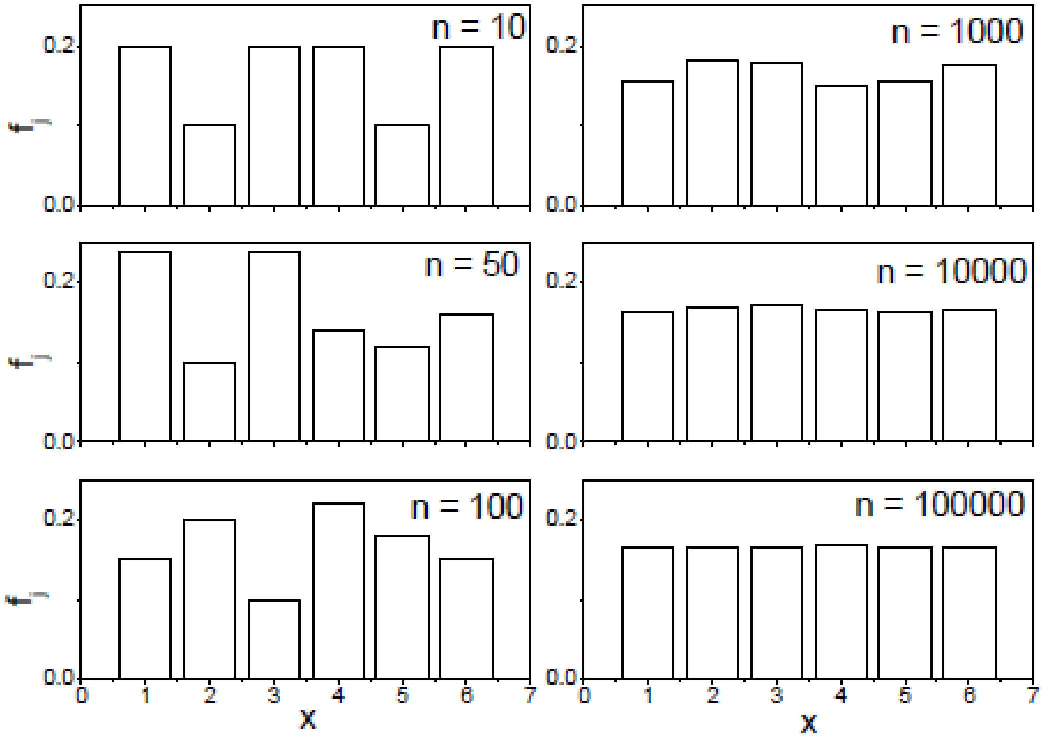
Tratamiento estadístico de datos

Supongamos que tiramos un dado N veces.

Los posibles resultados son: 1, 2, 3, 4, 5, 6

A priori, sabemos que cualquiera de los resultados es igualmente probable. ¿Qué es la probabilidad desde el punto de vista cuantitativo?

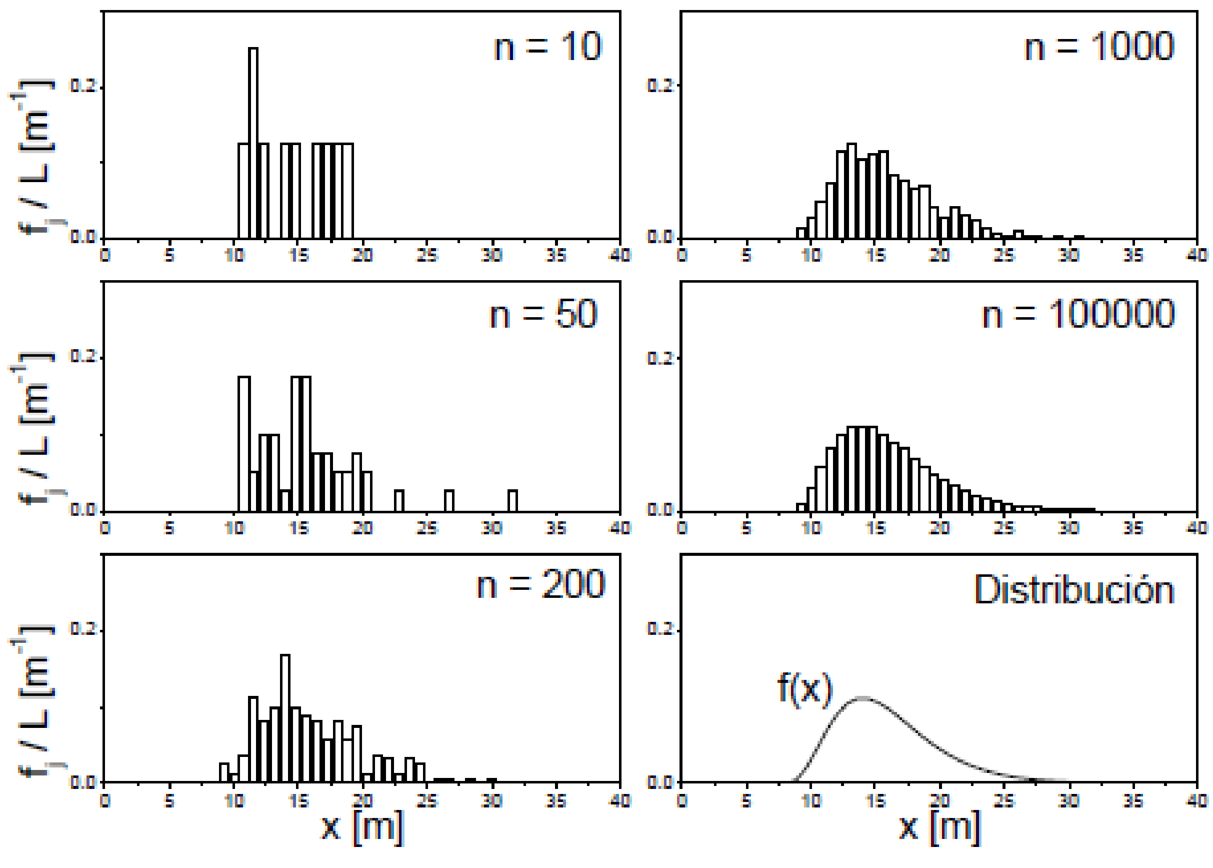
Supongamos que tiramos el dado N veces y hacemos un histograma de frecuencias relativas f_j , $j = 1, ..., 6$.



A medida que N aumenta todas las frecuencias tienden a un valor constante = 1/6, esa es la probabilidad de obtener cada resultado.

Las frecuencias relativas tienden a la probabilidad cuando N crece

Supongamos que podemos repetir una medición N veces, por ejemplo, la medida de la longitud de una mesa con una regla de 20 cm. Podemos medir: 10, 50, 200, 10000 veces y hacemos los histogramas correspondientes.



A medida que N aumenta, los valores de las frecuencias se estabilizan y la distribución tiende a una función continua f(x) (distribución)

¿Cuál es f (x) cuando se trata de los resultados de la medida de una magnitud?

¿Qué relación hay entre los valores del valor promedio y de la desviación estándar (calculados con los datos medidos) y la función de probabilidades (f (x))?

FLUCTUACIONES ESTADÍSTICAS

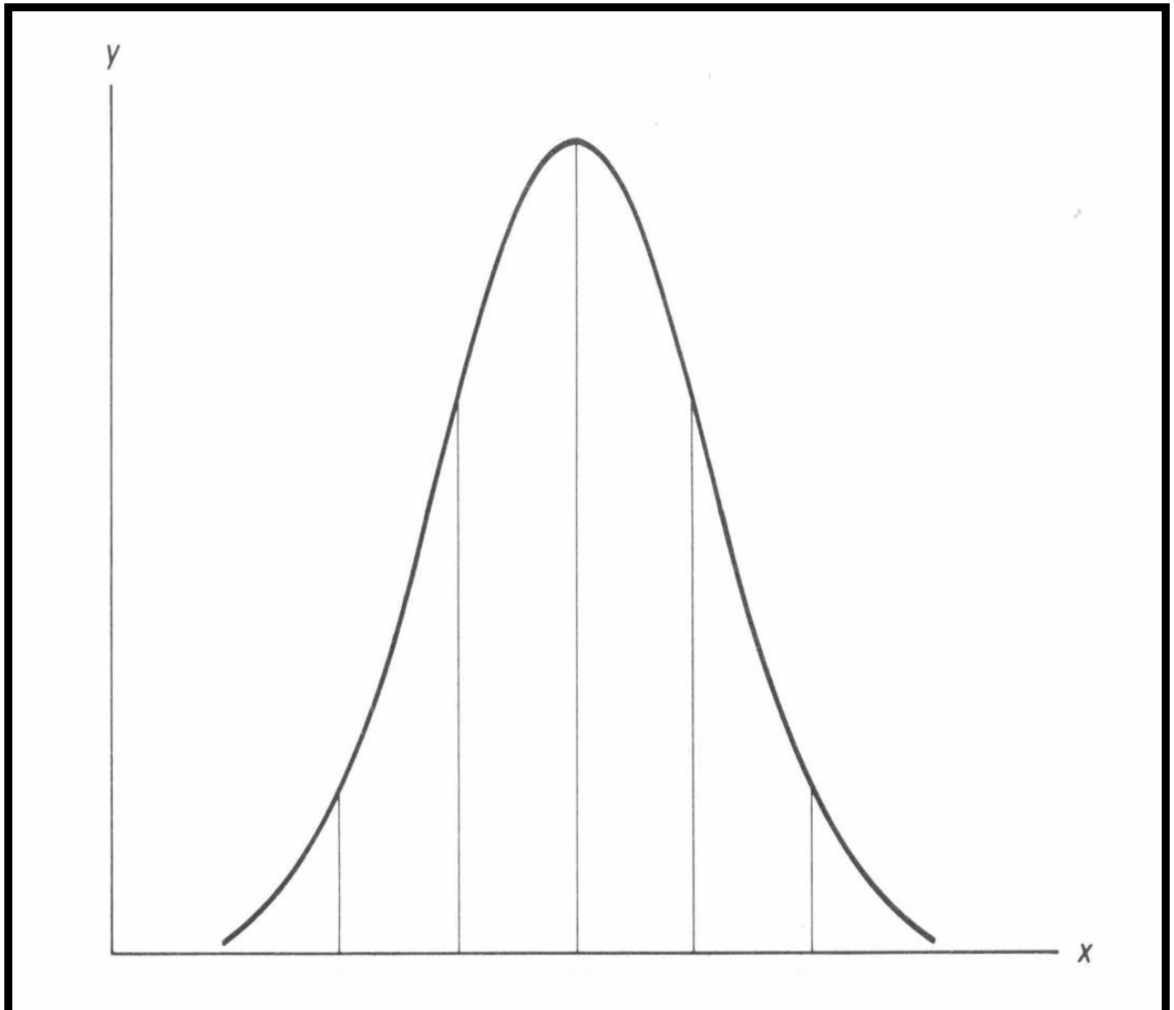
Los postulados fundamentales de la teoría estadística de “errores” establecen que, dado un conjunto de medidas, todas efectuadas en idénticas condiciones, suficientemente grande :

I. El valor más probable de la serie de mediciones es el valor medio (entones el máximo de f (x) coincide con el valor medio).

II. Es igualmente probable cometer “errores” de igual valor y distinto signo (Esto quiere decir que es igualmente probable obtener valores mayores que el valor medio, errores por exceso, o menores que el valor medio, por defecto, implica que la función f (x) debe ser simétrica).

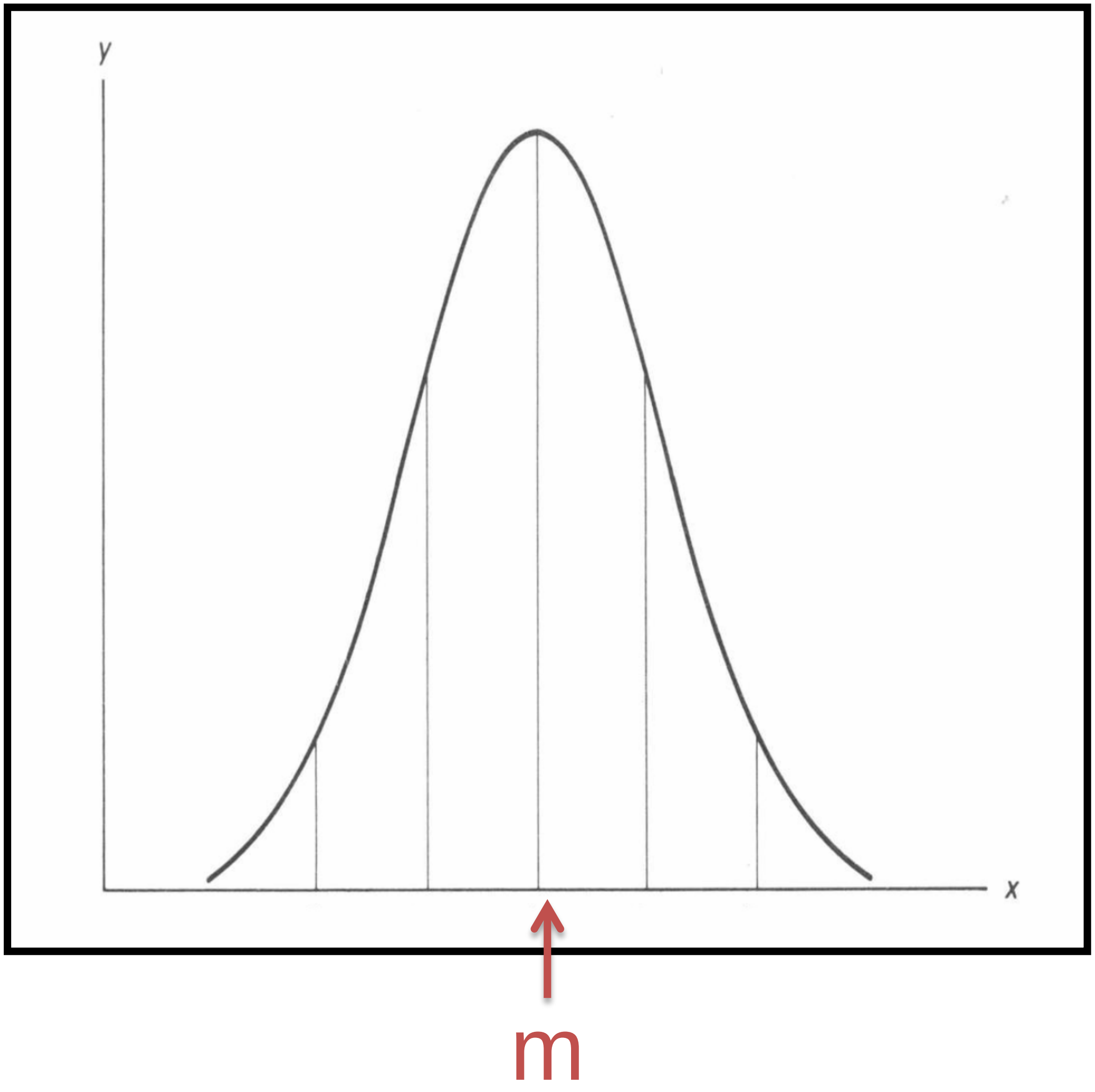
III. En una serie de mediciones es tanto más probable “cometer un error” cuanto menor sea su valor absoluto respecto del valor medio. (Esto significa que es más probable que los valores sean próximos al valor medio que alejados de éste. Esto implica que la función f (x) tiene a cero cuando me alejo del máximo).

En general, puede observarse que las fluctuaciones al azar de una medición se distribuyen de acuerdo a la distribución conocida como distribución Normal o de Gauss (cumple con los tres postulados).

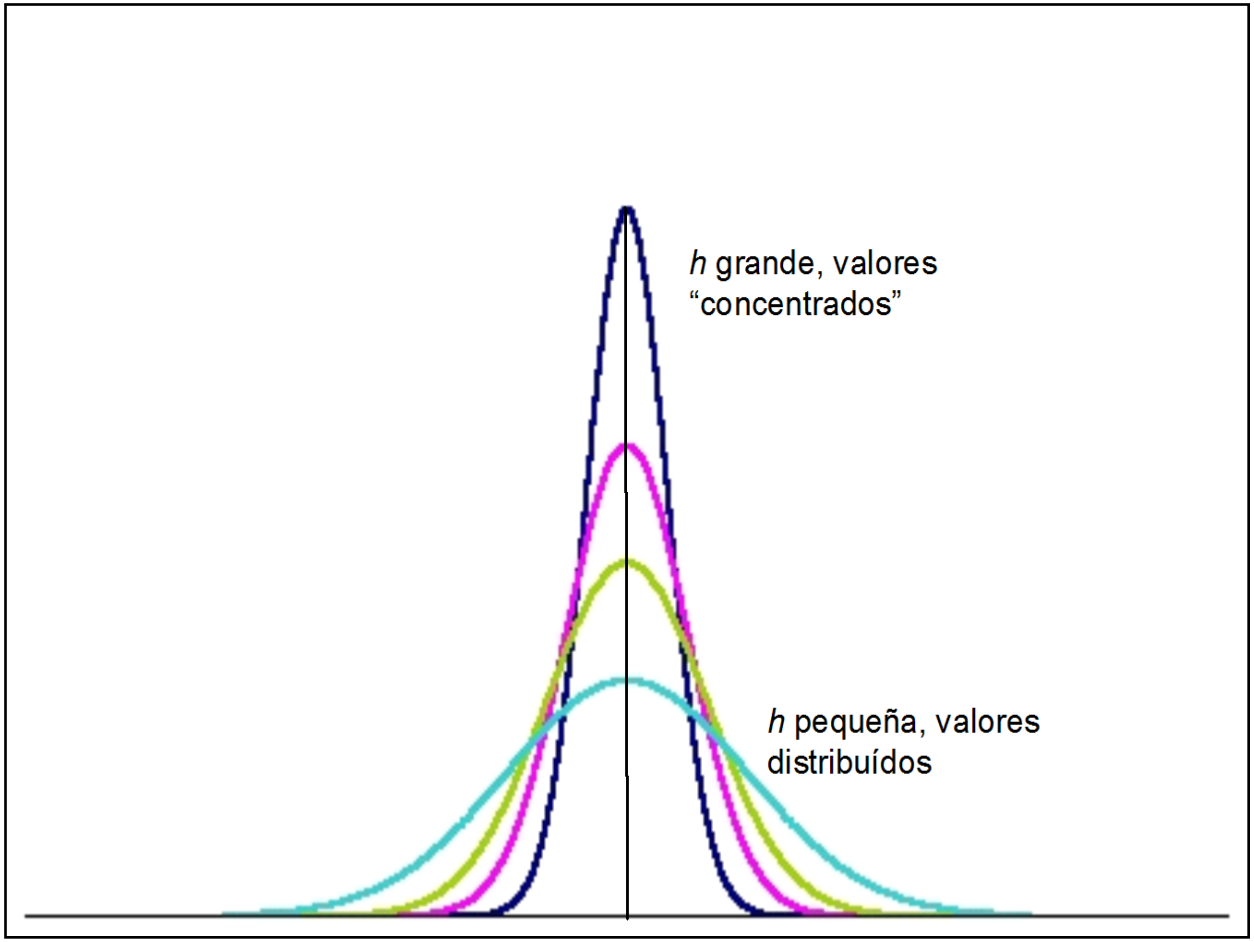


Representación de la función normal o "campana de Gauss" (responde a estos postulados)

$y = K\,exp[\,-h^2(x-m)^2]$



h: relacionado con el ancho de la distribución



Curva de Gauss para distintos valores de *h*
(inversamente proporcional al "ancho" de la distribución)

El área bajo la curva es:

$$\int K\, exp[\, -h^2(x-m)^2]\, dx = K\, \sqrt{\pi} \,/\, h\ ,$$

si el área bajo la curva vale la unidad, K queda determinada, y la curva quedará determinada por los parámetros m y h

m = es el valor más probable

h = relacionado con el ancho de la distribución

VALOR MEDIO Y DESVIACION ESTANDAR DE UNA SERIE DE MEDIDAS

Si medimos N veces la magnitud de interés obtendremos los N datos experimentales:

$$x_1 , x_2 , x_3 , \ldots\ldots\ldots, x_N$$

La media aritmética de las N medidas (valor medio o promedio) será:

$$< x > = (x_1 + x_2 + x_3 + \ldots\ldots + x_N)/N$$

$< x >$ es el mejor valor de la medida (el que minimiza la dispersión en torno a él)

¿Cómo cuantificamos la dispersión de valores?

Si definimos $d_i = x_i - \langle x \rangle$,

d_i es la desviación *i-ésima* respecto del valor medio $\langle x \rangle$.

$$\Sigma d_i = 0$$

Por ese motivo se introduce otra magnitud: la desviación cuadrática media, o **varianza S^2:**

$S^2 = (d_1^2 + d_2^2 + d_3^2 + \ldots + d_N^2)/N$ *(varianza, promedio de las desv.cuadráticas)*

y

$S = [(d_1^2 + d_2^2 + d_3^2 + \ldots + d_N^2) / N]^{½} = [(\Sigma (x_i - \langle x \rangle)^2) / N]^{½}$

S, **desviación estándar**, es la raíz cuadrada de la desviación cuadrática media (o varianza).

Se puede demostrar que el promedio de valores medidos (valor medio) minimiza la desviación estándar y lo podemos elegir como el valor más probable.

$$m = \langle x \rangle$$

Trabajamos con S y no con la varianza S^2 ya que S posee las mismas dimensiones que las observaciones.

En realidad la expresión correcta de la desviación estándar es:

$$S = [\ \Sigma\ (x_i - \langle x \rangle)^2 / (N-1)]^{1/2}$$

(para un solo evento N=1, S da infinito)

La desviación estándar S se relaciona con h en la Gaussiana...

$$h^2 = 1/2S^2$$

Si el número de medidas fuera infinito (para N tendiendo a infinito), la función de distribución tiende a:

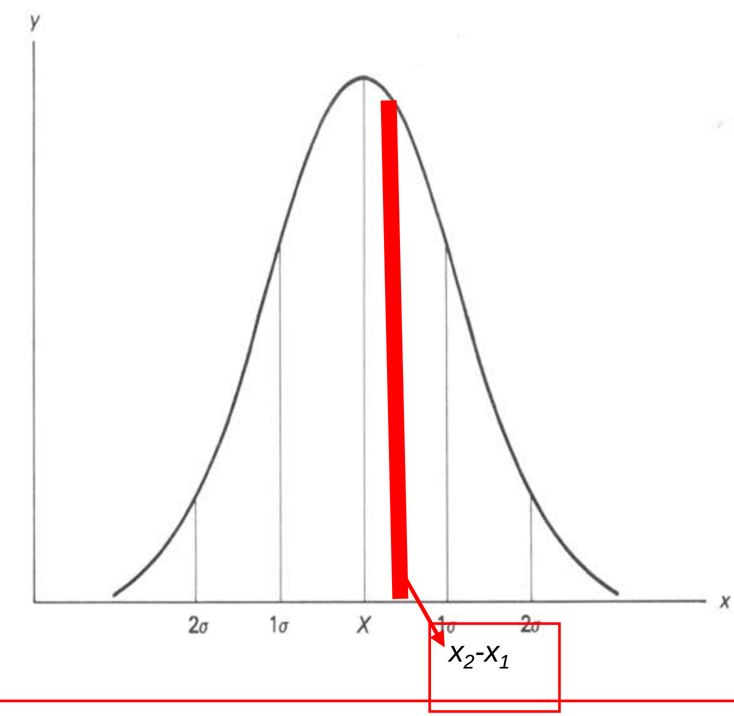
:

$$f(x) = 1 / [S(2\pi)^{1/2}] \exp [-(x-\langle x \rangle)^2/2S^2]$$

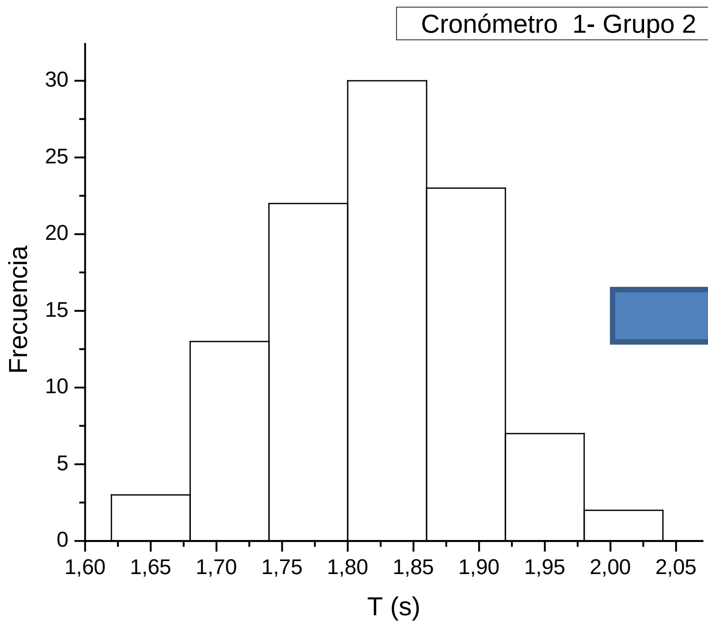
$$f(x) = 1 / [S(2\pi)^{1/2}] \exp [-(x-\langle x \rangle)^2/2S^2]$$

(para N tendiendo a infinito)

:



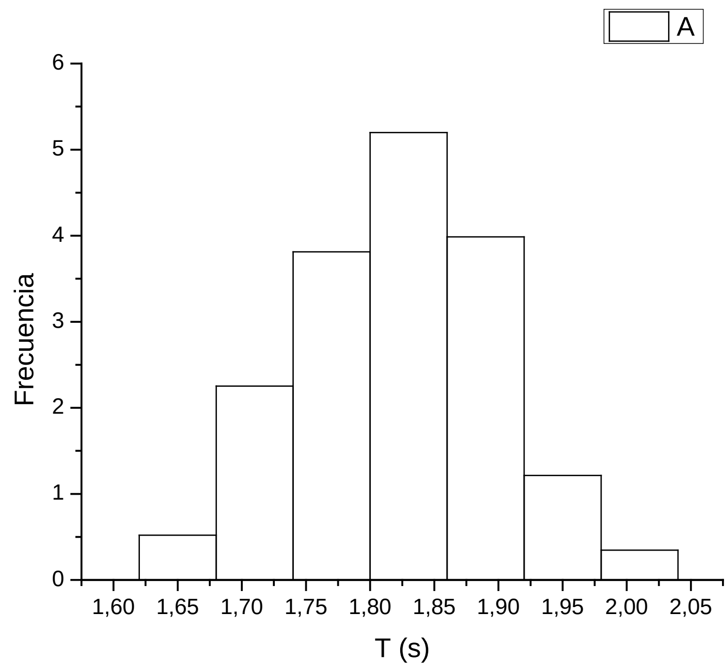
Si realizo una nueva medida habrá un 68% de probabilidad que ésta esté comprendida en $\langle x \rangle \pm S$, un 95% en $\langle x \rangle \pm 2S$ y un 97 % en $\langle x \rangle \pm 3S$.



Se normaliza dividiendo por el área (A):

$$A = \sum a.f_i = a.\, N$$

a: ancho de la clase
N: número de medidas



Para verificar si la distribución es de tipo normal o gaussiana, vamos a simular una función gaussiana con los datos obtenidos de nuestras medidas (valor medio y distribución estandar).

$$f(x) = 1 / [S(2\pi)^{1/2}] \exp[-(x-\langle x\rangle)^2/2S^2]$$

$\langle x\rangle$ = 1.8155 s

S= 0,0778 s