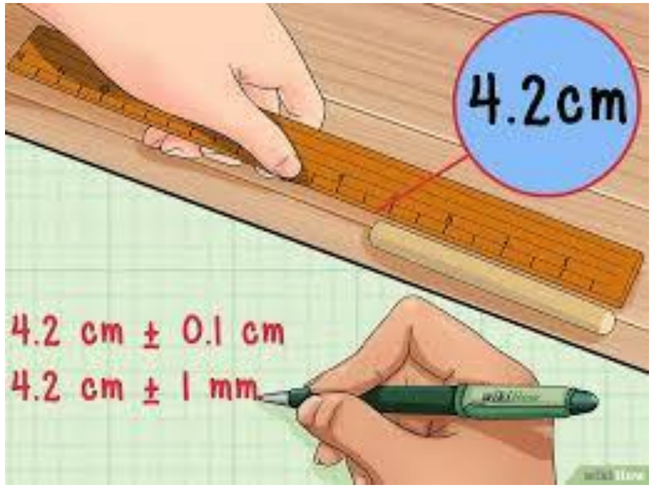
Retroalimentemos lo aprendido

Tratamiento estadístico de una magnitud.

Cuando medimos una magnitud una única vez (N = 1), el mejor valor es simplemente el valor medido, y su incertidumbre está dada por la incertidumbre nominal, σ_{nom} , que tiene en cuenta los errores del instrumento, del método y de las operaciones. En muchos casos prácticos estamos interesados en el estudio estadístico de la variación de las mediciones y realizamos mediciones de la magnitud de interés.



Supongamos que hemos hecho N medidas de una misma magnitud x con los resultados. Estas determinaciones pueden ser consideradas una muestra de todas las posibles mediciones que se podrían efectuar, es decir una muestra de la población. Bajo condiciones muy generales puede demostrarse que el mejor estimador de la magnitud x está dado por el promedio de los valores:

$\overline{x} = \frac{\sum_{j=1}^{n} x_j}{N}$ (promedio)

Es la mejor aproximación al valor desconocido en torno al cual fluctúan todos los resultados: el valor más probable del mesurando. El promedio de la muestra converge hacia el valor más probable cuando N tiende a infinito.

Llamaremos a:

$\Delta \boldsymbol{x} = \boldsymbol{x_j} - \overline{\boldsymbol{x}}; \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \dots \dots, \boldsymbol{N}$ **la desviación de cada medición respecto a** $\overline{\boldsymbol{x}}$

Desviación estándar

Para tener una idea cuantitativa de cuanto se aleja una serie de medidas $\{x_j\}$ del valor más representativo $\overline{x}$ debido a efectos aleatorios, podemos calcular $\sum \Delta x_j^2$, esto es, la suma de las 'desviaciones cuadráticas'.

Podríamos usar como estimación de la dispersión $\sum \Delta x_j^2$, pero esta no es una buena estimación, pues aún teniendo pequeñas desviaciones Δx_j una suma de muchas de ellas resultaría en un número grande. ***<u>Para independizarse del número de medidas se promedian esos valores, definiendo s_x , la desviación estándar de cada muestra de medidas</u>:***

$$s_x = \sqrt{\frac{\sum_{j=1}^{N}(x_j - \overline{x})^2}{N-1}} = \sqrt{\frac{\sum_{j=1}^{N} \Delta x_j^2}{N-1}}$$

Esta desviación es un estimador muestral de la desviación estándar poblacional , y nos da una idea global acerca de la dispersión de los puntos alrededor del promedio $\overline{x}$. Si S_x es grande, la probabilidad de hallar valores alejados del promedio es mayor, y si es pequeño los valores se concentran sobre el promedio. La desviación cuadrática tiene las mismas dimensiones físicas que $\overline{x}$, lo que hace posible una comparación directa a través del cociente $S_x/\overline{x}$. La calidad del proceso de medición será mayor cuanto menor sea $S_x/\overline{x}$. En la mayoría de los casos esta fracción es una constante del proceso de medición y no depende del número de medidas, una vez que es suficientemente grande.

Al cuadrado de S_x se le da el nombre de varianza

$$V = {S_x}^2$$

La varianza es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media

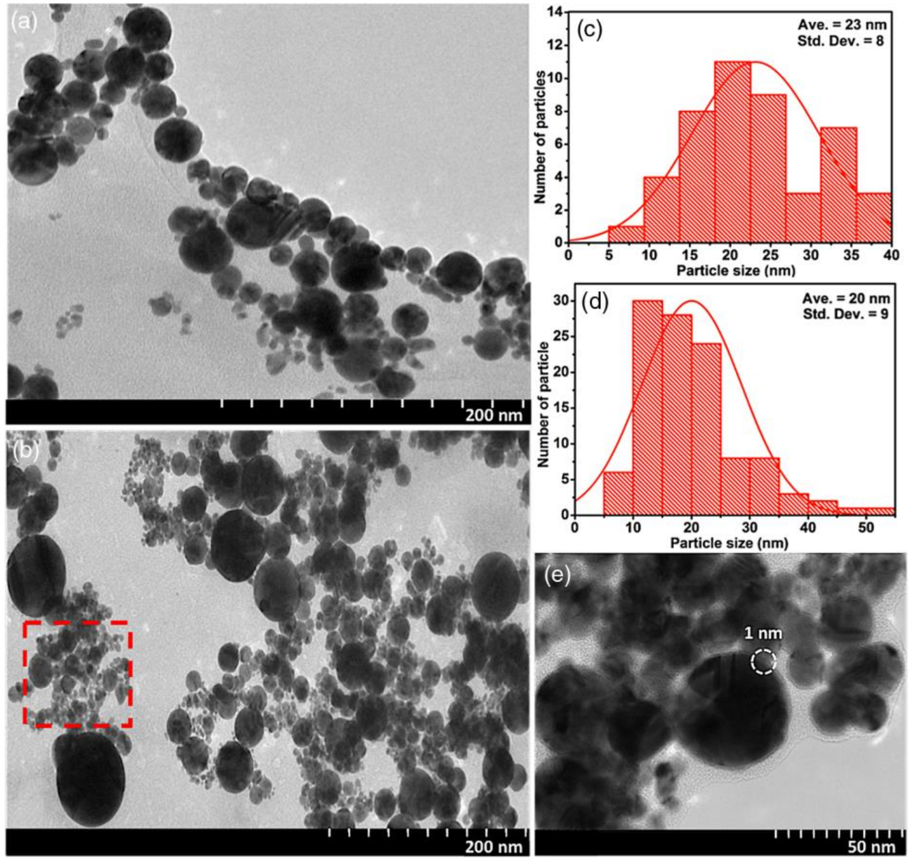
Desviación estándar del promedio

$$\sigma_{\overline{x}} = \frac{S_x}{\sqrt{N}}$$

y en un experimento es una medida de la incertidumbre estadística asociada a $\overline{x}$ en el proceso de medir la magnitud *N* veces.
esta definición tiene sentido si la desviación muestral es siempre la misma. Si el experimento se realiza en condiciones idénticas cada vez esto es siempre cierto (en efecto ya hemos supuesto previamente que las condiciones experimentales no cambian). La otra característica destacable de la expresión anterior es que su valor puede ser reducido arbitrariamente, siempre que se tome N suficientemente grande

Histogramas y distribución estadística

Veamos algunas elementos útiles para el tratamiento estadístico del resultado de una serie de medidas***. El problema central de la estadística es inferir propiedades de una población a partir de las propiedades de una muestra de algunos individuos de la misma***. Por población entendemos no sólo un grupo de personas, sino que podemos referirnos a objetos más generales como la población de todos los resultados posibles de una medida.



Concentrémonos en una propiedad x de una población. Si tomamos una muestra de tamaño *N* y observamos el valor de esta propiedad para cada individuo, encontramos *N* resultados:, x_1, x_2, x_3,…….., x_n comprendidos en un intervalo de valores $[x_{min}, x_{max}]$ entre el menor y mayor valor registrado.

Una manera útil de visualizar las características de este conjunto de datos consiste en dividir el intervalo $[x_{min}, x_{max}]$ en *m* sub-intervalos o clases delimitados por los puntos y_1, y_1, y_1,…….., y_m . Luego, contamos el número n_1 de individuos de la muestra cuyas edades estén comprendidas en el primer intervalo $[y_j, y_{j+1}]$, etc., el número de los individuos de la muestra que están en el *j*-ésimo intervalo , etc., hasta llegar al *m*-ésimo sub-intervalo. Estos valores son la ***frecuencia absoluta de cada clase***.

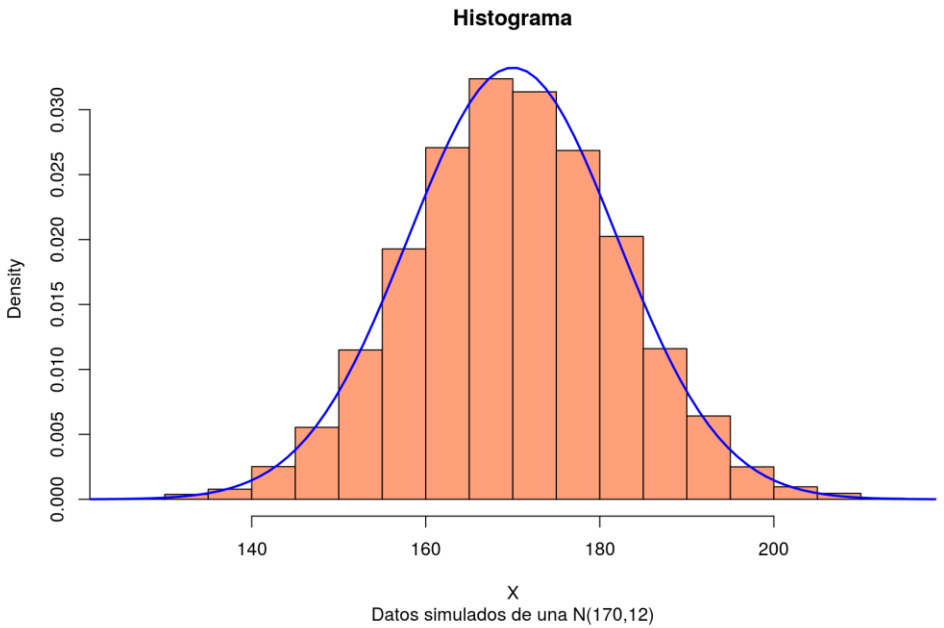
A partir de la frecuencia absoluta definimos la función f_j de distribución como:

$$f_j = \frac{n_j}{\sum_j n_j}$$

La función de distribución está normalizada, es decir:

$$\sum_{j=1}^{m} f_j = 1$$

Si construimos un gráfico de f_j en función de x_j para $j = 1,2,3, \dots, m$ donde $x_j = (y_{j-1} + y_j)/2$ son los centros de los intervalos, nos dará una idea clara de cómo se distribuyen las edades de los individuos de la muestra en estudio. Este tipo de gráfico se llama ***histograma***. Usualmente estos gráficos se representan con barras.

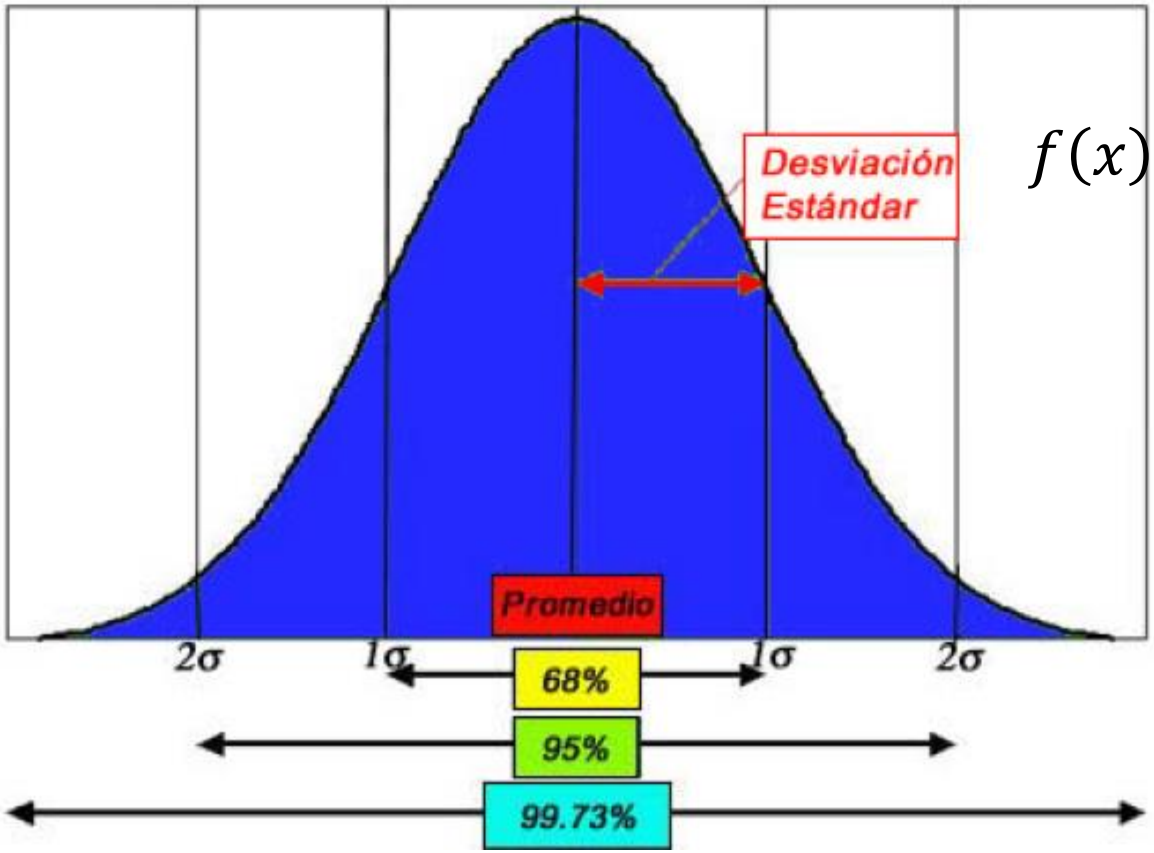


Es una herramienta usada para representar una distribución por medio de barras. La altura de la barra está en función de la frecuencia (eje y) y el rango (eje x) de una variable continua.

Nos ofrece un vistazo general del comportamiento de las variables, donde logramos analizar aspectos como distribución, dispersión, aleatoriedad y tendencia.

Distribución Normal o Gaussiana

La distribución de probabilidad más importante es la distribución gaussiana o normal, que tiene una forma de campana . Corresponde a la función



$$f(x) = N(x; x_o, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x - x_o)^2}{2\sigma^2}\right]$$

Donde $x_o y\ \sigma$ son constante arbitrarias y x es el punto de evaluación. El factor que multiplica la exponencial garantiza que el area total bajo la curva sea igual a uno.

La campana gaussiana es simétrica respecto a x_o y su ancho está determinado por el parámetro σ (‘casualmente’ este parámetro resulta ser la desviación estándar de la distribución).

Ejercicio Mostrar que la suma de las áreas de las barras de un histograma cualquiera en el que el ancho de cada intervalo del rango de clases *Δx* está fijo, es igual a *N* Δ*x*

Tiempo en minutos por usuario				
11,50	10,26	10,08	13,00	11,14
13,73	13,41	10,44	11,36	14,40
11,64	12,39	12,82	14,25	15,41
14,35	9,35	12,40	9,04	15,30
14,79	15,27	10,63	14,30	15,48
14,80	8,78	14,00	13,09	10,00
12,20	11,70	15,37	11,81	10,06
12,49	8,58	11,32	12,20	12,45
11,28	12,60	14,36	13,08	13,50
12,68	9,19	14,32	12,17	9,10

Paso 1: Determinar el rango: El valor mínimo y el valor máximo de los datos

Valor Max=15,48   valor min= 8,58

R=15,48-8,58=6,9

Paso 2. Se calcula el número de intervalos de clase(k).

$$K = \sqrt{N} = \sqrt{50} \approx 7{,}07 \quad \Rightarrow \quad K=7$$

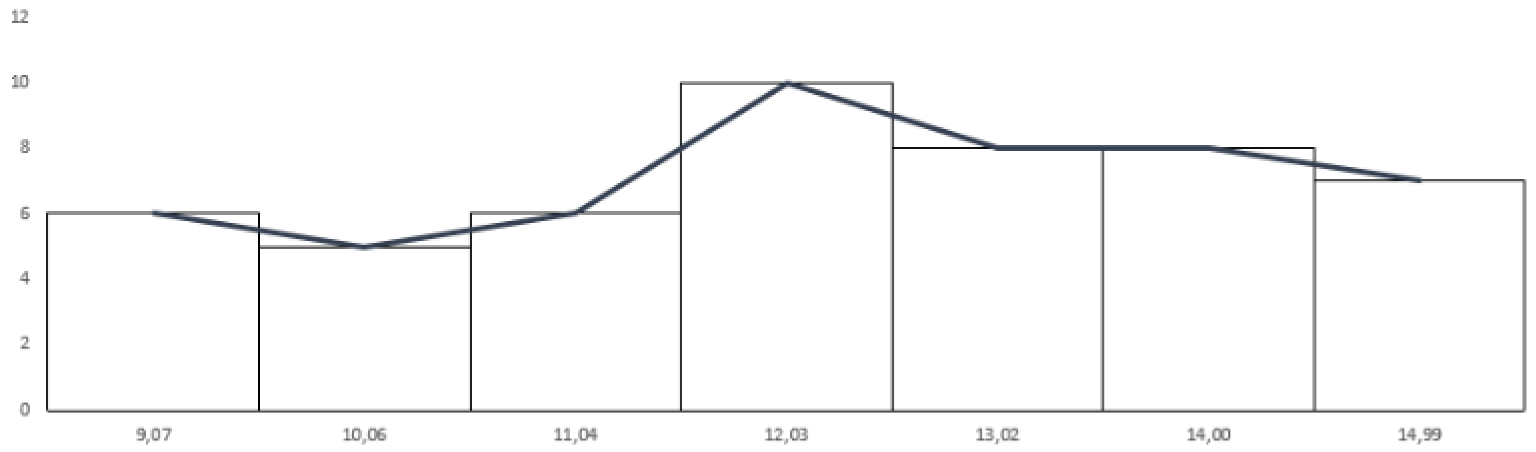
Paso 3: Se calcula la amplitud o ancho del intervalo

$$A = \frac{R}{K} = \frac{6{,}9}{7} = 0{,}99$$

Paso 4 : Se define las clases sumándole al valor más pequeño , el ancho del intervalo hasta que obtenga K intervalos

Paso 5: Se agrupa cada valor dentro del intervalo de clase, o dicho de otra forma, determinamos la frecuencia

Intervalo de clase		Frecuencia
Desde	Hasta	
8,58	9,57	6
9,57	10,55	5
10,55	11,54	6
11,54	12,52	10
12,52	13,51	8
13,51	14,49	8
14,49	15,48	7



Ejercicio: Considere la siguiente tabla, donde se detalla la lita de N=30 medida de la masa m de una muestra de cierto material

1,09	**1,01**	**1,10**	**1,14**	**1,16**
1,11	1,04	1,16	1,13	1,17
1,14	1,03	1,17	1,09	1,09
1,15	1,06	1,12	1,08	1,20
1,08	1,07	1,14	1,11	1,05
1,06	1,12	1,00	1,10	1,07

A). Realice el histograma correspondiente a estos valores.
B).Calcule la desviación estándar de las medidas