

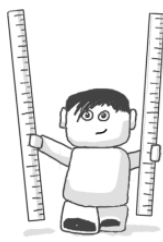


Tema 1. Estadística descriptiva.

Curso 2023-2024

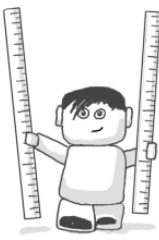
Grado de Física

Técnicas Experimentales II



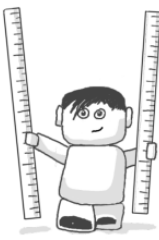
Objetivos tema:

- Introducción al tema, glosario y bibliografía.
- Distribuciones de frecuencias.
- Representaciones gráficas de la información.
- Medidas características: centralización, dispersión, asimetría y apuntamiento.
- Transformaciones de variable aleatoria: lineales y no lineales.
- Muestras multivariantes.



Bibliografía:

- [VAR10] **Tratamiento de Datos Fisicos.** L. M. Varela, F. Gomez, J. Carrete. Servizo de Publicacions e Intercambio Científico. Universidade de Santiago, 2010.
- [VIM12] **Vocabulario Internacional de Metrología** Conceptos fundamentales y generales, y términos asociados (VIM) (**JCGM 200:2012**) 3ª Ed. CEM
- [Pena08] **Fundamentos de Estadística,** D. Peña. Alianza Editorial, 2008.
- [Wonn87] **Introducción a la estadística,** T. H. Wonnacott, R. J. Wonnacott. Editorial Limusa, 1987.
- [Spie03] **Probabilidad y estadística.** Spiegel Murray. Editorial McGrawHill, 2003.

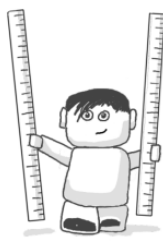


Vocabulario. Conceptos y definiciones.

Magnitud: propiedad de un fenómeno, cuerpo o sustancia, que puede expresarse cuantitativamente mediante un número y una referencia.

longitud, l	radio, r	radio del círculo A, r_A o $r(A)$
	longitud de onda, λ	longitud de onda de la radiación D del sodio, λ_D o $\lambda(D; Na)$
energía, E	energía cinética, T	energía cinética de la partícula i en un sistema dado, T_i
	calor, Q	calor de vaporización de la muestra i de agua, Q_i
Carga eléctrica, Q		Carga eléctrica del protón, e
Resistencia eléctrica, R		Valor de la resistencia eléctrica i en un circuito dado, R_i
Concentración de cantidad de sustancia del constituyente B, c_B		Concentración: cantidad de sustancia de etanol en la muestra i de vino, $c_i(C_2H_5OH)$
Concentración de número de partículas del constituyente B, C_B		Concentración: número de eritrocitos en la muestra i de sangre, $C(Erc; Sg_i)$
Dureza Rockwell C, HRC		Dureza Rockwell C de la muestra i de acero, HRC_i

La referencia puede ser una **unidad de medida**, un **procedimiento de medida**, un **material de referencia** o una combinación de ellos.



Vocabulario. Conceptos y definiciones.

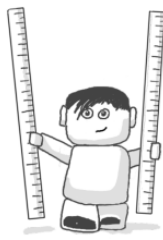
Toda **magnitud física** se expresa en función de las magnitudes básicas del Sistema Internacional de Magnitudes (ISQ)[§].

Magnitud básica	Dimensión
longitud	L
masa	M
tiempo	T
corriente eléctrica	I
temperatura termodinámica	Θ
cantidad de sustancia	N
intensidad luminosa	J

La dimensión de una magnitud Q se expresará como:

$$\dim(Q) = L^{\alpha} M^{\beta} T^{\gamma} I^{\delta} \Theta^{\varepsilon} N^{\zeta} J^{\eta}$$

Siendo los exponentes positivos, negativos o nulos.



Vocabulario. Conceptos y definiciones.

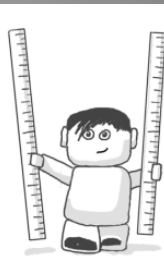
Unidad de medida: valor de una magnitud escalar real, definida y adoptada por convenio, con la que se puede comparar cualquier otra magnitud de la misma **naturaleza** para expresar la relación entre ambas mediante un número.



Magnitud básica	Unidad básica	
	Nombre	Símbolo
longitud	metro	m
masa	kilogramo	kg
tiempo	segundo	s
corriente eléctrica	ampère	A
temperatura termodinámica	kelvin	K
cantidad de sustancia	mol	mol
intensidad luminosa	candela	cd

Unidades de las magnitudes básicas del Sistema Internacional de Unidades (SI).

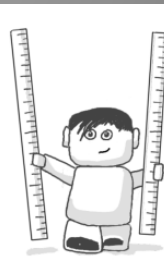
Las unidades se expresan mediante nombres y símbolos, asignados por convenio.



Vocabulario. Conceptos y definiciones.

Medición o medida: proceso que consiste en obtener experimentalmente uno o varios **valores** que pueden atribuirse razonablemente a una **magnitud**.

- Las mediciones no son de aplicación a las **propiedades cualitativas**.
- Una medición supone una comparación de magnitudes o el conteo de entidades (i.e. Geiger).
- Una medición supone una descripción de la magnitud compatible con el uso previsto de un **resultado de medida**, un **procedimiento de medida** y un **sistema de medida** calibrado conforme a un procedimiento de medida especificado, incluyendo las condiciones de medida.



Vocabulario. Conceptos y definiciones.

Mensurando:

magnitud que se desea medir. La especificación de un mensurando requiere el conocimiento de la **naturaleza de la magnitud** y la descripción del estado del fenómeno, cuerpo o sustancia de la cual esta magnitud es una propiedad. Será necesario incluir las componentes pertinentes involucradas.



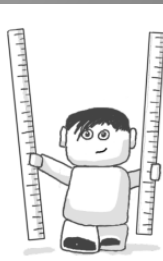
Definición del mensurando (¿son iguales todos los diámetros? ¿Cuánto se desvía la moneda de una geometría cilíndrica?)

Magnitudes de influencia (¿Temperatura?)

Trazabilidad (¿Calibración del calibre?)

Repetibilidad (¿Dependencia entre la posición relativa de la moneda y calibre?)

¿Holguras/imperfecciones del aparato de medida?)



Vocabulario. Conceptos y definiciones.

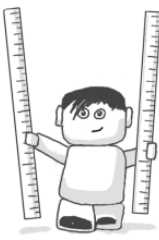
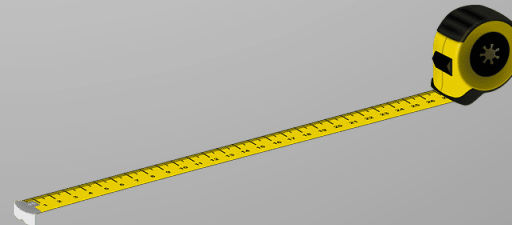
Magnitudes de influencia:

las magnitudes que afectan al **sistema de medida** o a la **magnitud de interés**, pero que no constituyen el objeto de la medida.



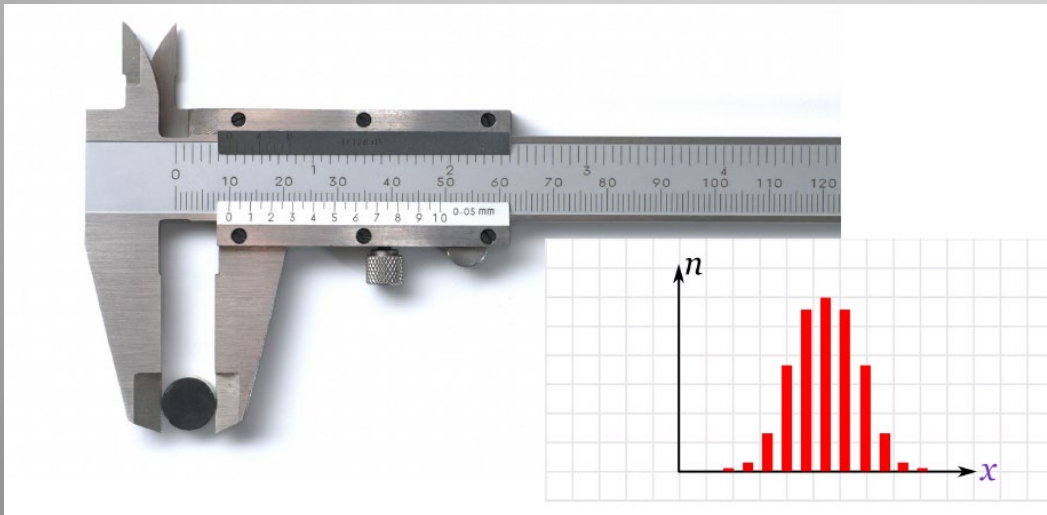
Forma de onda en la medida de valor eficaces de corriente alterna.

Temperatura en una medida de longitud.

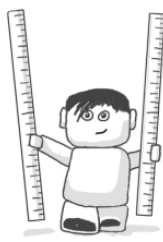


Vocabulario. Conceptos y definiciones.

La **reiteración de la medida** puede arrojar valores variables que se podrían atribuir razonadamente al mensurando. Esto implica la existencia de una **incertidumbre** en la determinación del valor del mensurando.

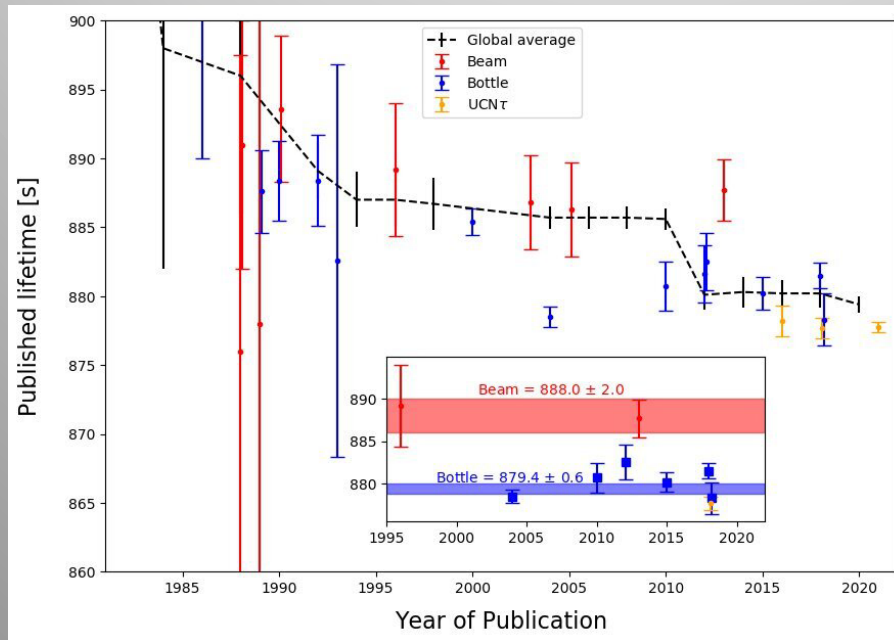


En general sólo podemos asignar un **intervalo, conjunto o distribución de valores** para describir el mensurando.



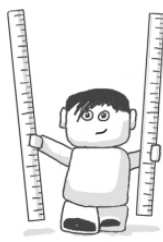
Vocabulario. Conceptos y definiciones.

Valor verdadero de una magnitud: valor compatible con la definición de la **magnitud**. En el enfoque en torno al concepto de error, el valor verdadero de la magnitud se considera único y, en la práctica, imposible de conocer en la descripción de la **medición**.

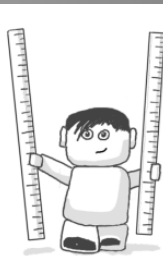
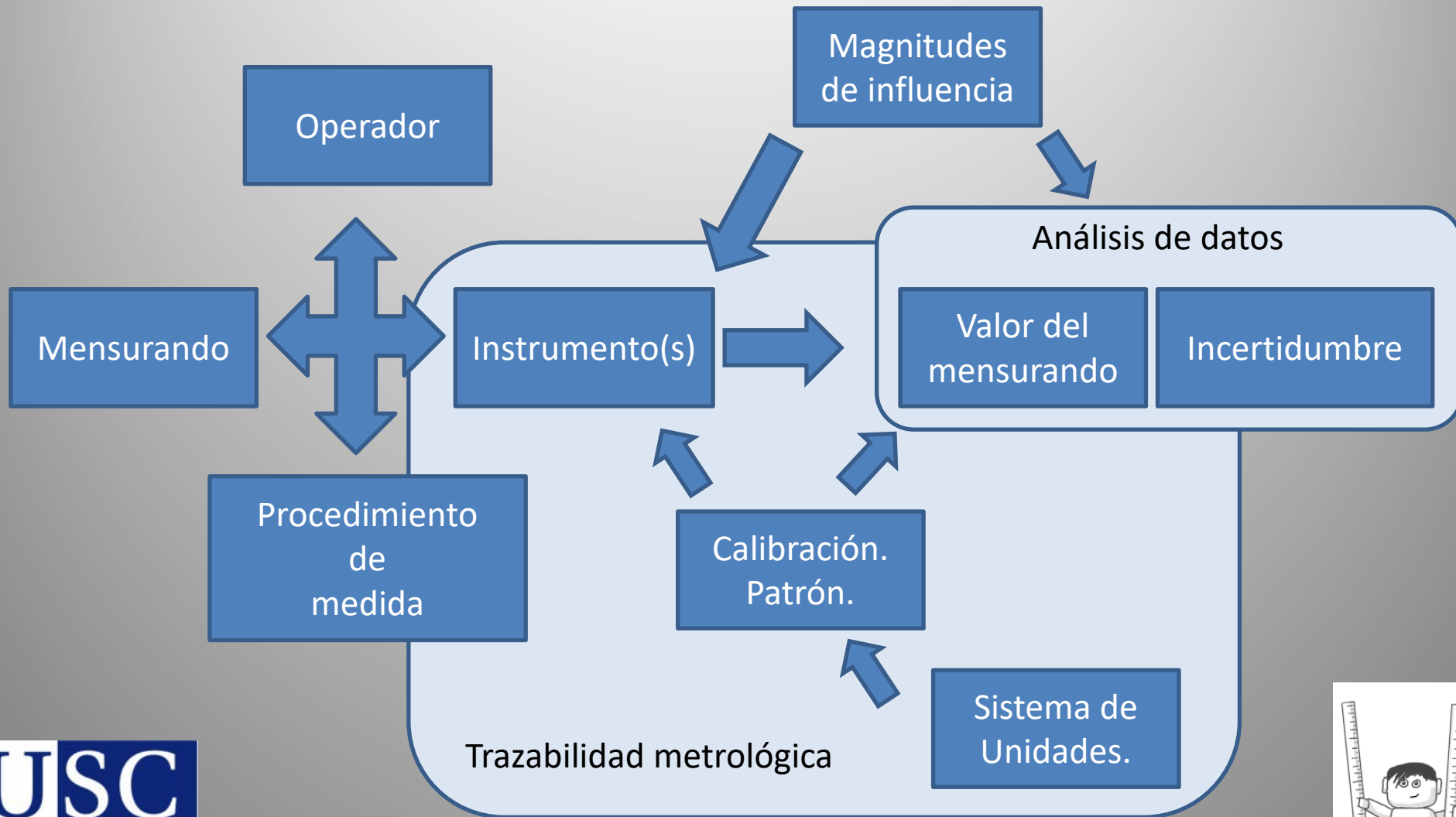


Vida media del neutrón.

Improved Neutron Lifetime Measurement with UCN τ
F. M. Gonzalez *et al.* (UCN τ Collaboration)
Phys. Rev. Lett. **127**, 162501



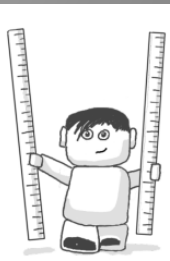
Vocabulario. Conceptos y definiciones.



Vocabulario. Conceptos y definiciones.

Valor convencional de una magnitud: valor asignado a una **magnitud**, mediante un acuerdo, para un determinado propósito.

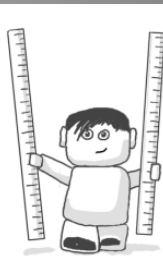
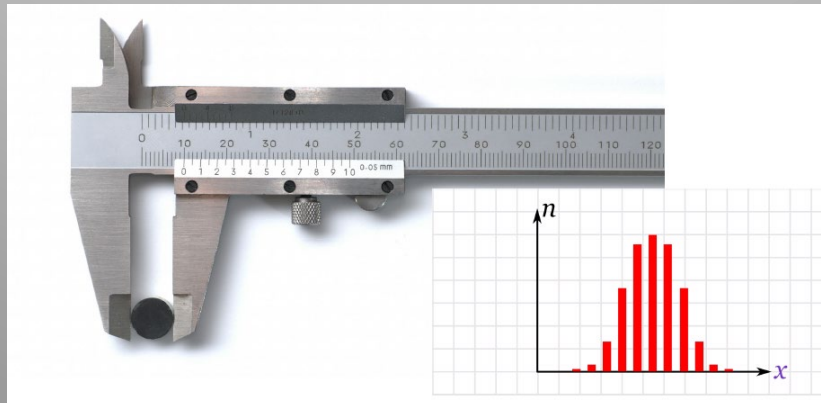
Valor convencional de la velocidad de la luz en el vacío: 299 792 458 m/s
Equivalentemente, un metro es la longitud recorrida por la luz en el vacío durante $1/299\,792\,458$ de segundo.



Vocabulario. Conceptos y definiciones.

El **objetivo de las mediciones en el enfoque “de la incertidumbre” no es determinar el mejor valor verdadero posible**. Se supone más bien que la información obtenida de la medición permite únicamente atribuir al mensurando un intervalo de valores razonables, suponiendo que la medición se ha efectuado correctamente.

Sin embargo, **ni la medición más refinada permite reducir el intervalo a un único valor**, a causa de la cantidad finita de detalles que intervienen en la definición del mensurando. La incertidumbre de la definición del mensurando (incertidumbre intrínseca) impone un límite inferior a toda incertidumbre de medida.

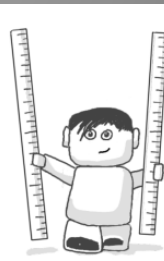
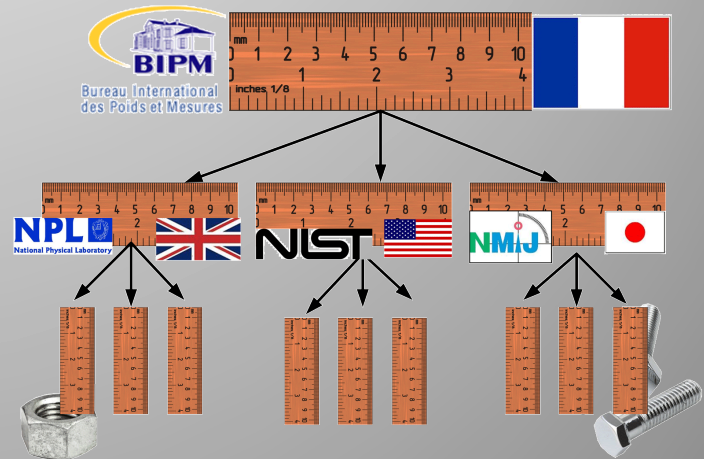


Vocabulario. Conceptos y definiciones.

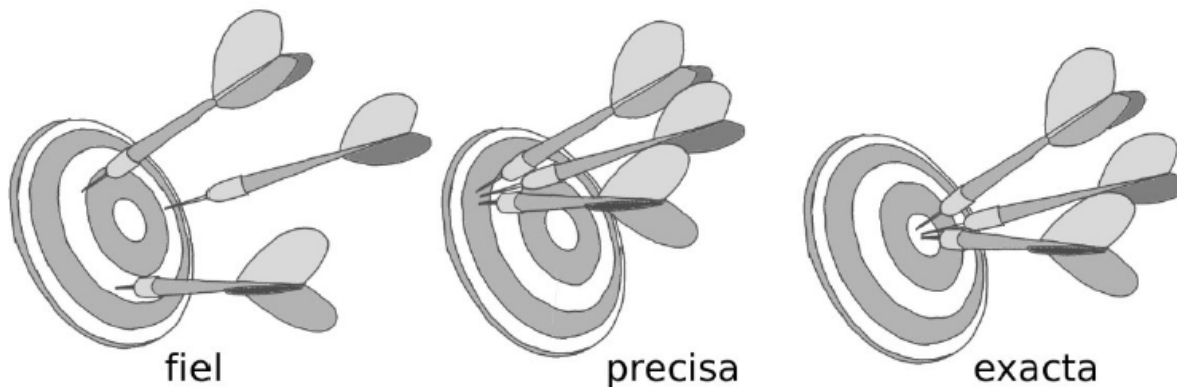
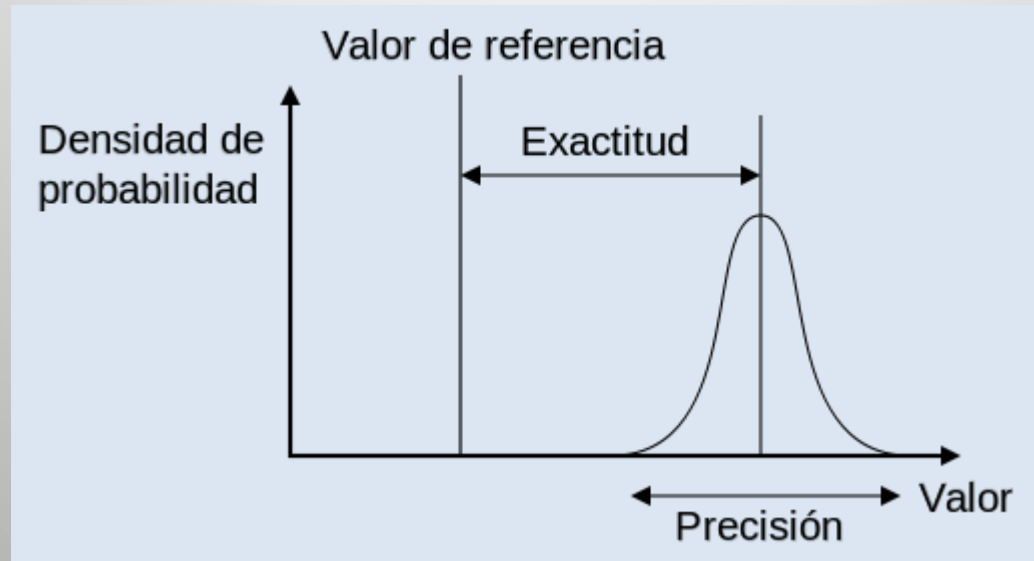
Trazabilidad metrológica:

propiedad de un **resultado de medida** por la cual el resultado puede relacionarse con una referencia mediante una cadena ininterrumpida y documentada de **calibraciones**, cada una de las cuales contribuye a la **incertidumbre de medida**.

- En esta definición, la referencia puede ser la definición de una **unidad de medida**, mediante una realización práctica, un **procedimiento de medida** que incluya la unidad de medida cuando se trate de una **magnitud no ordinal**, o un **patrón**.
- La trazabilidad metrológica requiere una **jerarquía de calibración** establecida.



Características de la medida



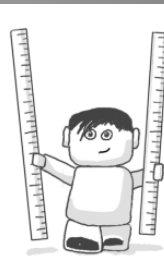
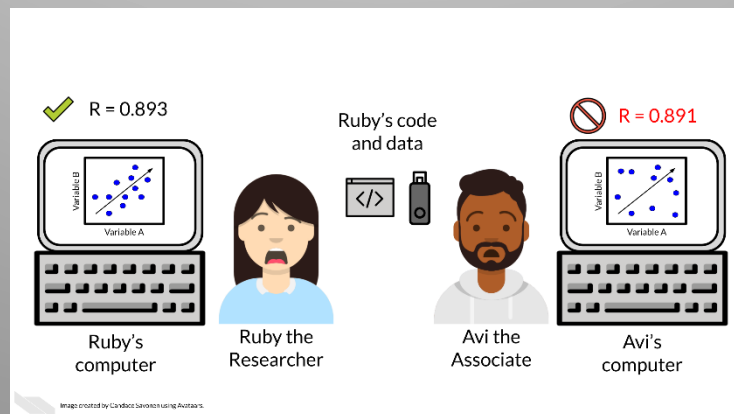
Características de la medida

Repetibilidad de una medida:

Precisión obtenida en una medición dentro de un conjunto de condiciones que incluye el mismo **procedimiento de medida**, los mismos operadores, el mismo **sistema de medida**, las mismas condiciones de operación y el mismo lugar, así como mediciones repetidas del mismo objeto o de un objeto similar en un periodo corto de tiempo.

Reproducibilidad de una medida:

Precisión obtenida en una medida dentro de un conjunto de condiciones que incluye **diferentes lugares, operadores, sistemas de medida** y mediciones repetidas de los mismos objetos u objetos similares durante períodos de tiempo arbitrarios. El conocimiento científico requiere la reproducibilidad.



Experimento aleatorio

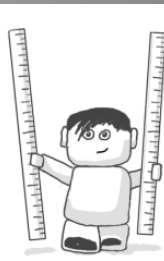
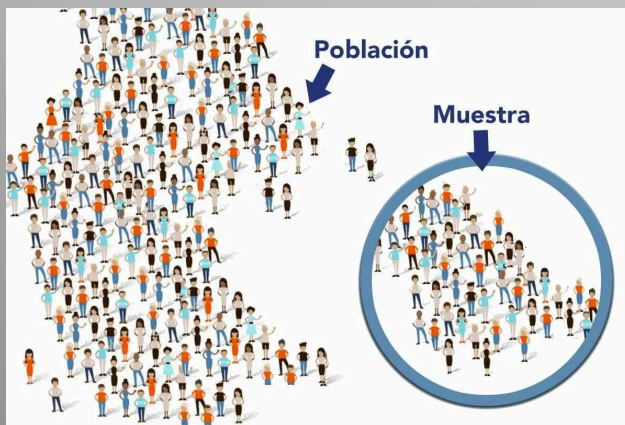
Población: conjunto de todos los elementos, fenómenos, individuos o unidades estadísticas sobre los que se puede manifestar una propiedad objeto de medida. La población ser finita o infinita.

Muestra: subconjunto de la población escogido para su análisis estadístico.

Experimento aleatorio: aquel en el que es imposible conocer de antemano el valor concreto de cada resultado (suceso aleatorio). Cada repetición se considera realizada en condiciones idénticas. El conjunto de posibles valores resultado del experimento es conocido de antemano.

Variables (cantidades o mensurandos) cuantitativas o estadísticas: corresponden a valores numéricos asignables a los sucesos aleatorios. Pueden ser discretas o continuas.

Variables (cantidades o mensurandos) cualitativas: son las que se corresponden a valores no numéricos a los sucesos aleatorios.



Variable aleatoria

Espacio muestral: el espacio muestral (Ω) de un determinado experimento aleatorio es el conjunto de todos los posibles valores asociados a los resultados del experimento aleatorio.

$$\Omega = \{1,2,3,4,5,6\} \quad \text{Dado}$$

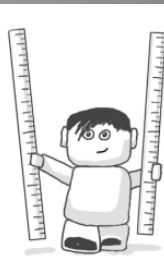
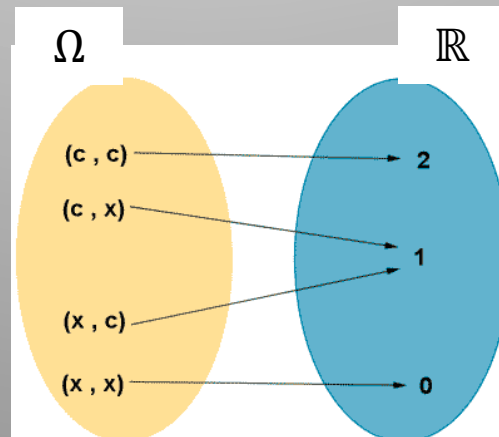
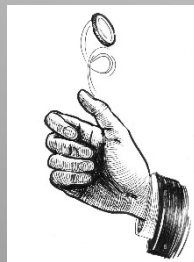
$$\Omega' = \{C, X\} \quad \text{Moneda}$$

En general para realizar evaluaciones cuantitativas asociadas a los experimentos aleatorios consideraremos que el resultado de cada suceso puede hacerse corresponder con un número real.

Variable aleatoria: sobre un espacio muestral (Ω) de un determinado experimento aleatorio en el que se denomina la función real X que asocia cada suceso posible (A) a un valor de la recta real, X se denomina **variable aleatoria unidimensional**.

$$X: \Omega \rightarrow \mathbb{R}$$

$$A \rightarrow X(A)$$



Distribución de frecuencias

Tras la realización de un experimento aleatorio N veces obtendremos un conjunto o muestra \mathcal{M} de N datos de la forma:

$$\mathcal{M} = \{y_1, y_2, \dots, y_N\}$$

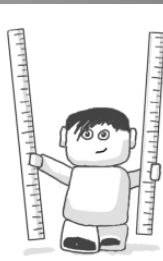
Consideremos que el recorrido de la variable aleatoria es numerable, de forma que podemos ordenar los k valores distintos que es posible obtener de la variable:

$$x_1 < x_2 < \dots < x_k$$

Denominaremos **frecuencia absoluta** n_i del valor i -ésimo de la variable aleatoria al número de veces que se observa dicho valor en el total de las observaciones realizadas en el experimento.

$$j \in \{j_1, j_2, \dots, j_{n_i}\}; \quad y_j = x_i$$

$$\sum_{i=1}^k n_i = N$$



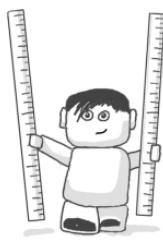
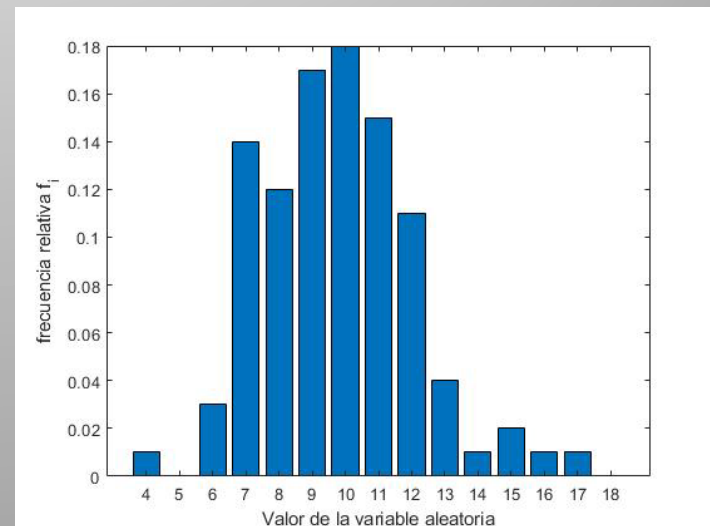
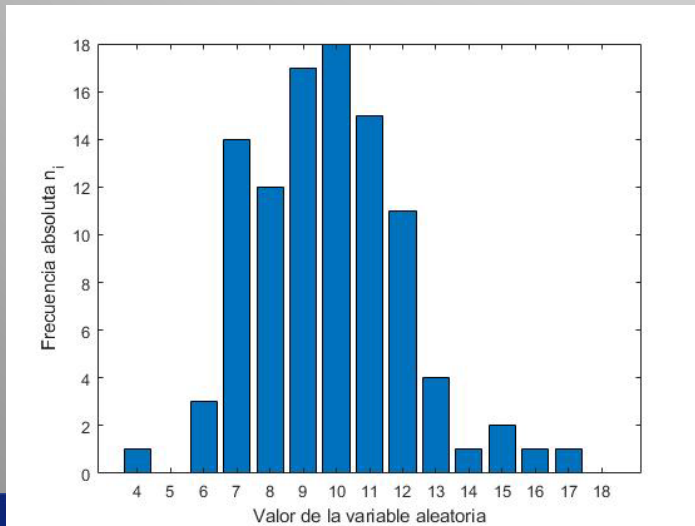
Distribución de frecuencias

Denominaremos **frecuencia relativa** f_i del valor i -ésimo de la variable aleatoria a la fracción de veces n_i/N que se observa dicho valor en el total de las observaciones realizadas en el experimento.

$$j \in \{j_1, j_2, \dots, j_{n_i}\}; \quad y_j = x_i$$

$$f_i = \frac{n_i}{N}$$

$$\sum_{i=1}^k f_i = 1$$



Distribución de frecuencias

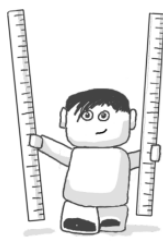
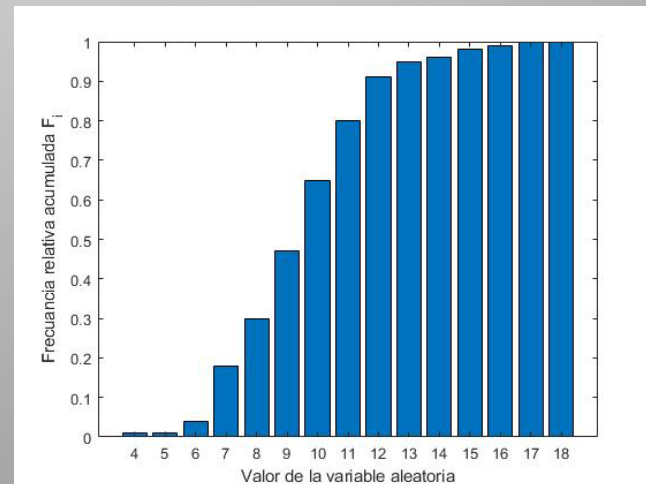
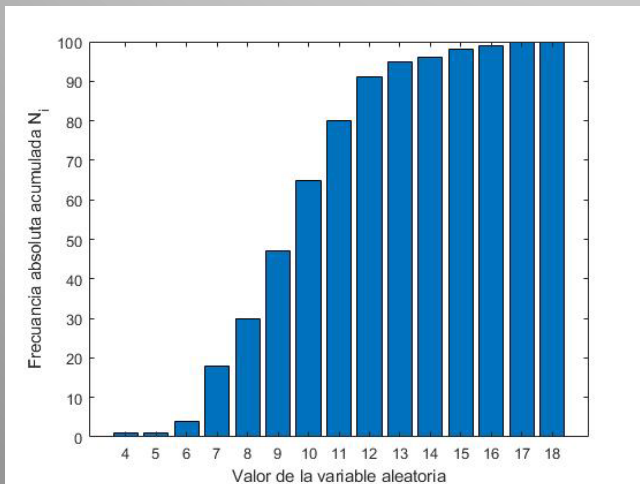
Denominaremos **frecuencia absoluta acumulada** N_i del valor i -ésimo de la variable aleatoria al número de veces que se observa un valor igual o inferior a x_i en el total de las observaciones realizadas en el experimento.

$$j \in \{j_1, j_2, \dots, j_{N_i}\}; \quad y_j \leq x_i$$

$$\sum_{j=1}^i n_j = N_i$$

La **frecuencia relativa acumulada** F_i se define mediante:

$$F_i = \frac{N_i}{N}$$



Distribución de frecuencias

Se cumplen las siguientes propiedades para las frecuencias absolutas y relativas:

i)

$$0 \leq n_i \leq N; 0 \leq f_i \leq 1.$$

ii)

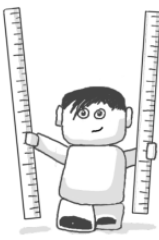
$$\sum_{i=1}^k n_i = N \Leftrightarrow \sum_{i=1}^k f_i = 1.$$

iii)

$$N_k = \sum_{i=1}^k n_i = N \Leftrightarrow F_k = 1.$$

iv)

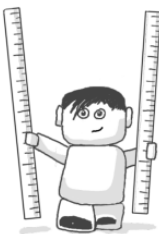
$$N_i = N_{i-1} + n_i; F_i = F_{i-1} + f_i.$$



Agrupación de datos.

Ejemplo: (1.8 en [VAR10]) Datos de una desintegración radiactiva medidas por un contador Geiger en el que se han efectuado un total de 300 medidas cada 30 segundos.

7	5	3	6	8	4	5	7	6	4	6	3	4	4
9	8	9	7	0	3	5	8	7	8	4	6	5	5
3	5	3	8	4	9	4	10	3	5	5	8	6	7
8	9	5	9	6	5	5	7	3	7	6	4	6	9
4	8	4	5	7	8	7	6	5	4	5	7	8	3
7	5	3	9	8	1	6	4	7	8	5	6	11	9
6	4	6	10	7	6	2	13	3	6	0	8	1	6
5	6	9	4	10	7	6	7	9	6	3	7	5	12
2	7	5	6	7	5	5	2	4	6	9	2	5	10
2	6	7	8	4	5	7	6	6	7	5	4	3	2
10	8	3	2	8	4	6	3	3	8	4	5	6	7
11	2	6	5	5	7	9	8	5	2	4	6	6	3
7	5	6	7	4	10	6	7	4	5	8	7	5	5
12	10	5	6	12	3	11	4	10	4	5	4	9	8
3	5	10	7	9	6	7	4	11	7	6	1	11	2



Agrupación de datos.

Ejemplo: (1.8 en [VAR10]) Datos de una desintegración radiactiva medidas por un contador Geiger en el que se han efectuado un total de 300 medidas cada 30 segundos.

x_i	n_i	f_i	N_i	F_i
0	2	0,0067	2	0,0067
1	7	0,0233	9	0,0300
2	15	0,0500	24	0,0800
3	25	0,0833	49	0,1633
4	38	0,1267	87	0,2900
5	52	0,1733	139	0,4633
6	52	0,1733	191	0,6367
7	40	0,1333	231	0,7700
8	30	0,1000	261	0,8700
9	19	0,0633	280	0,9333
10	10	0,0333	290	0,9667
11	6	0,0200	296	0,9867
12	3	0,0100	299	0,9967
13	1	0,0033	300	1,0000

x_i : resultados de la variable aleatoria x

n_i : frecuencia absoluta

f_i : frecuencia relativa

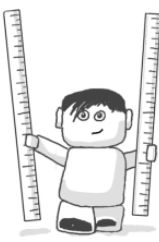
N_i : frecuencia acumulada absoluta

F_i : frecuencia acumulada relativa

$$\sum_{i=1}^{14} f_i = 1$$

$$\sum_{i=1}^{14} n_i = 300$$

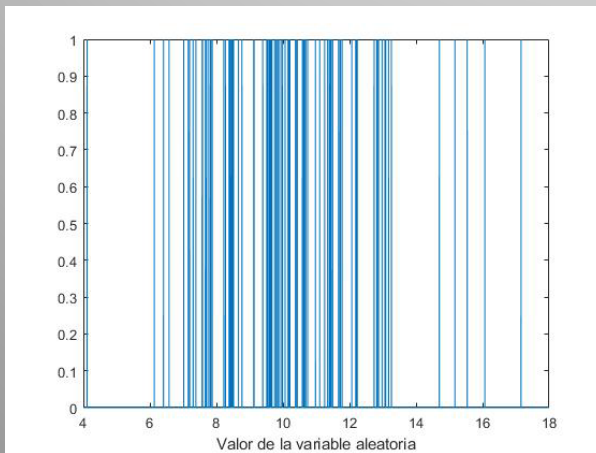
En este caso particular la variable aleatoria es discreta y resulta natural la asignación de la tabla de contingencia.



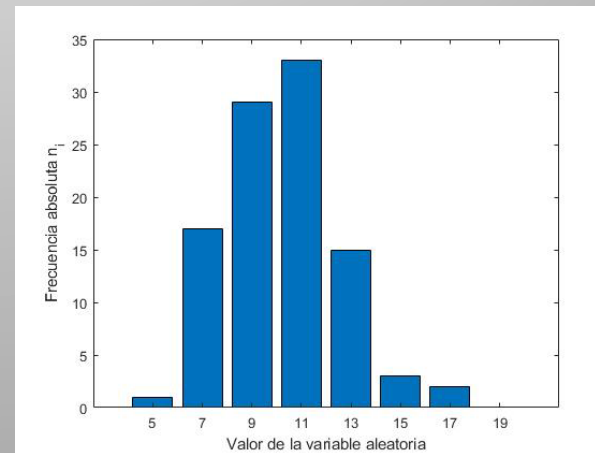
Agrupación de datos. Marcas de clase

En el caso de variable aleatoria continua (o de variable discreta con un gran número de posibles resultados), se **agruparán los resultados en una serie de intervalos disjuntos**, cuya unión cubra todo el posible recorrido de la variable. A los sucesos dentro de cada clase se les etiqueta por un suceso característico, el valor central del intervalo o, en general, una **marca de clase**.

- Los resultados anteriores referidos a frecuencias o frecuencias acumuladas tanto relativas como absolutas, se generalizan sumando las frecuencias de todos los resultados que pertenezcan a una marca de clase determinada. Así, **la frecuencia absoluta para una marca de clase corresponderá al número de resultados que pertenezcan a su clase asociada**.
- No existen, sin embargo, normas referentes a como realizar la agrupación, aunque si ciertas reglas para optimizar la selección.



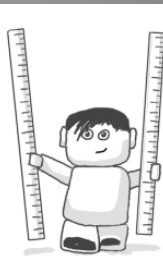
Sin agrupamiento.



Con agrupamiento.

[4,6);[6,8);[8,10);[10,12);[12,14);[14,16);
[16,18);[18,20)

$$\sum_{i=1}^8 n_i = 100$$



Agrupación de datos. Marcas de clase

- Para optimizar la selección de las marcas de clase en un conjunto de **N** medidas, **se suele elegir un número de clases del orden del entero más próximo a la raíz de N**.

En el ejemplo anterior, con **N** = 100, el número de divisiones óptimo estaría alrededor de 10 (y hemos elegido 8 en la forma de agrupamiento).

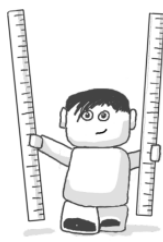
- **Las clases en las que se dividen los resultados se suelen tomar de la misma amplitud**, salvo en casos en los que la agrupación de valores sugiera otras soluciones.
- **No puede haber ambigüedad en las clases**: un resultado siempre debe ser atribuible a una clase y solo a una. Para ello deben definirse intervalos semiabiertos por un extremo y cerrados por el otro. Con esto tendremos una división del recorrido de la variable aleatoria en subintervalos o clases:

$$\{ [a_0, a_1), [a_1, a_2), [a_2, a_3), [a_3, a_4), \dots, [a_{k-1}, a_k) \}$$

A los que asociaremos marcas de clase: $\{ x_1, x_2, x_3, x_4, \dots, x_{k-1}, x_k \}$ en las que habitualmente se toma el centro de cada intervalo como marca asociada:

$$x_i = \frac{(a_{i-1} + a_i)}{2}; \quad i = 1, \dots, k$$

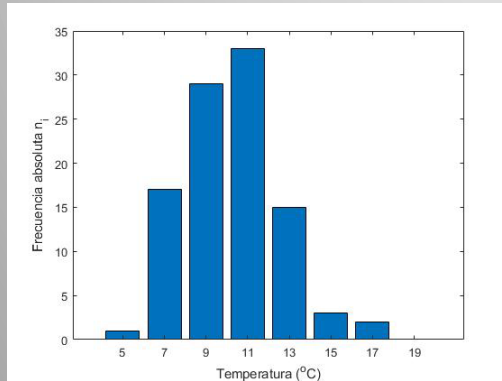
- Se establecen las frecuencias de los resultados pertenecientes a cada clase.



Agrupación de datos. Marcas de clase

Consideremos el ejemplo anterior de 100 medidas de la temperatura.

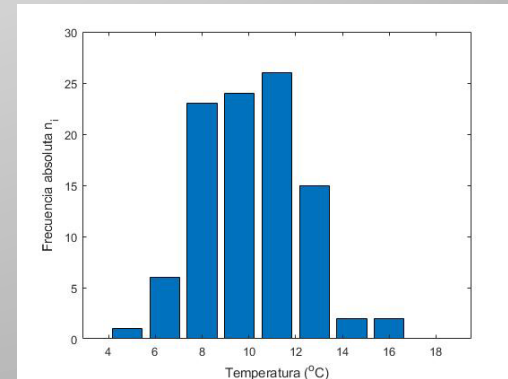
$T(^{\circ}\text{C}) = \{ 11.7243 \ 10.6375 \ 7.3846 \ 9.1328 \ 10.6852 \ 17.1568 \ 15.5389 \ 7.3002 \ 16.0698 \ 11.4508 \ 9.8739 \ 11.4295 \ 9.5901 \ 9.7517 \ 12.9794 \ 12.8181$
 $12.8344 \ 11.3430 \ 7.5850 \ 11.4345 \ 13.2605 \ 10.9778 \ 12.0694 \ 11.4538 \ 9.3931 \ 10.5877 \ 8.4254 \ 11.7768 \ 7.7059 \ 7.8623 \ 8.3810 \ 4.1114 \ 12.8768$
 $10.6504 \ 8.4901 \ 12.7406 \ 6.5770 \ 9.7955 \ 9.5171 \ 10.6384 \ 10.6257 \ 8.2702 \ 9.9399 \ 9.6702 \ 11.2554 \ 12.1865 \ 12.2185 \ 8.2727 \ 10.1547 \ 7.5718$
 $7.7730 \ 9.9863 \ 13.0653 \ 8.4607 \ 10.7428 \ 9.5488 \ 12.2347 \ 7.8219 \ 10.0651 \ 11.1051 \ 12.2012 \ 13.0884 \ 10.1719 \ 7.0168 \ 8.5154 \ 7.8768 \ 14.7009$
 $8.7688 \ 11.4962 \ 9.6152 \ 11.7772 \ 8.4703 \ 7.1955 \ 7.1552 \ 10.9764 \ 9.6452 \ 9.6079 \ 12.8386 \ 10.5832 \ 10.3956 \ 13.1754 \ 8.3911 \ 11.3932 \ 11.6702$
 $9.5126 \ 10.4313 \ 7.6683 \ 7.7041 \ 10.2097 \ 11.4445 \ 15.1710 \ 8.6662 \ 10.3747 \ 9.8350 \ 6.1340 \ 9.1221 \ 6.4106 \ 11.6808 \ 8.2239 \ 10.2002 \}$



$k=8$

Clases $[4,6); [6,8); [8,10); [10,12); [12,14); [14,16); [16,18); [18,20)$

Marcas de clase $\{5, 7, 9, 11, 13, 15, 17, 19\}$

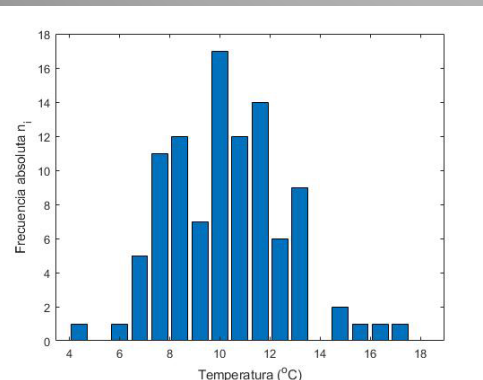


$k=10$

Clases $[4,5.6); [5.6,7.2); [7.2,8.8); [8.8,10.4); [10.4,12);$

$[12,13.6); [13.6,15.2); [15.2,16.8); [16.8,18.4); [18.4,20.0)$

Marcas de clase $\{4.8, 6.4, 8.0, 9.6, 11.2, 12.8, 14.4, 16.0, 17.6, 19.2\}$



$k=20$

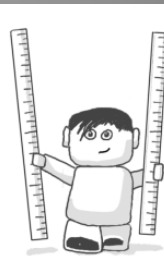
Clases $[4,4.8); [4.8,5.6); [5.6,6.4); [6.4,7.2); [7.2,8.0);$

$[8.0,8.8); [8.8,9.6); [9.6,10.4); [10.4,11.2); [11.2,12.0);$

$[12.0,12.8); [12.8,13.6); [13.6,14.4); [14.4,15.2); [15.2,16.0);$

$[16.0,16.8); [16.8,17.6); [17.6,18.4); [18.4,19.2); [19.2,20.0)$

Marcas de clase $\{4.4, 5.2, 6.0, 6.8, 7.6, 8.4, 9.2, 10.0, 10.8, 11.6,$
 $12.4, 13.2, 14.0, 14.8, 15.6, 16.4, 17.2, 18.0, 18.8, 19.6\}$

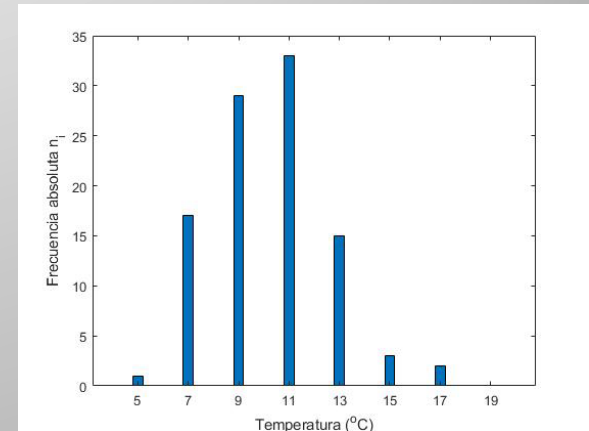


Representación gráfica

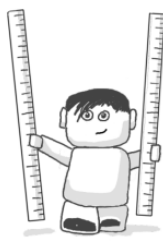
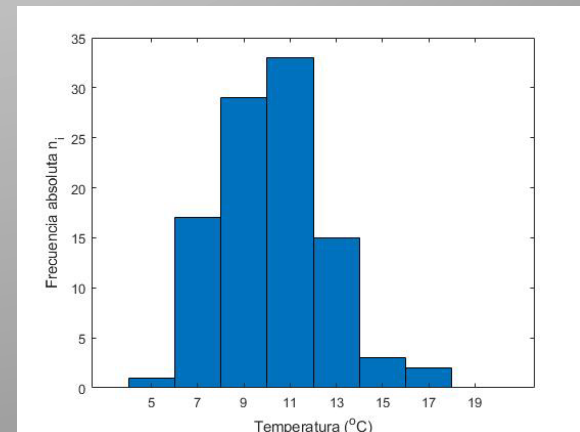
En general el número de clases debería de situarse en torno a \sqrt{N} pero hay casos en los que un mayor o menor número de divisiones mejoran la comprensión de la distribución.

La información contenida en los resultados de un experimento aleatorio puede interpretarse con mayor facilidad utilizando representaciones graficas de las distribuciones de frecuencia. Las más importantes representaciones graficas son:

- **Diagramas de barras:** asocia a cada marca de clase o valor de la variable una barra de altura igual a la frecuencia con la que la marca o el valor aparecen. Puede corresponder a frecuencias absolutas o relativas, ordinarias o acumuladas.



- **Histogramas:** asocia a cada clase de la variable un rectángulo cuya base sea la amplitud del intervalo y cuya altura es proporcional a la frecuencia con que aparecen los resultados en la clase.



Agrupación de datos.

Ejemplo: (1.8 en [VAR10]) Datos de una desintegración radiactiva medidas por un contador Geiger en el que se han efectuado un total de 300 medidas cada 30 segundos.

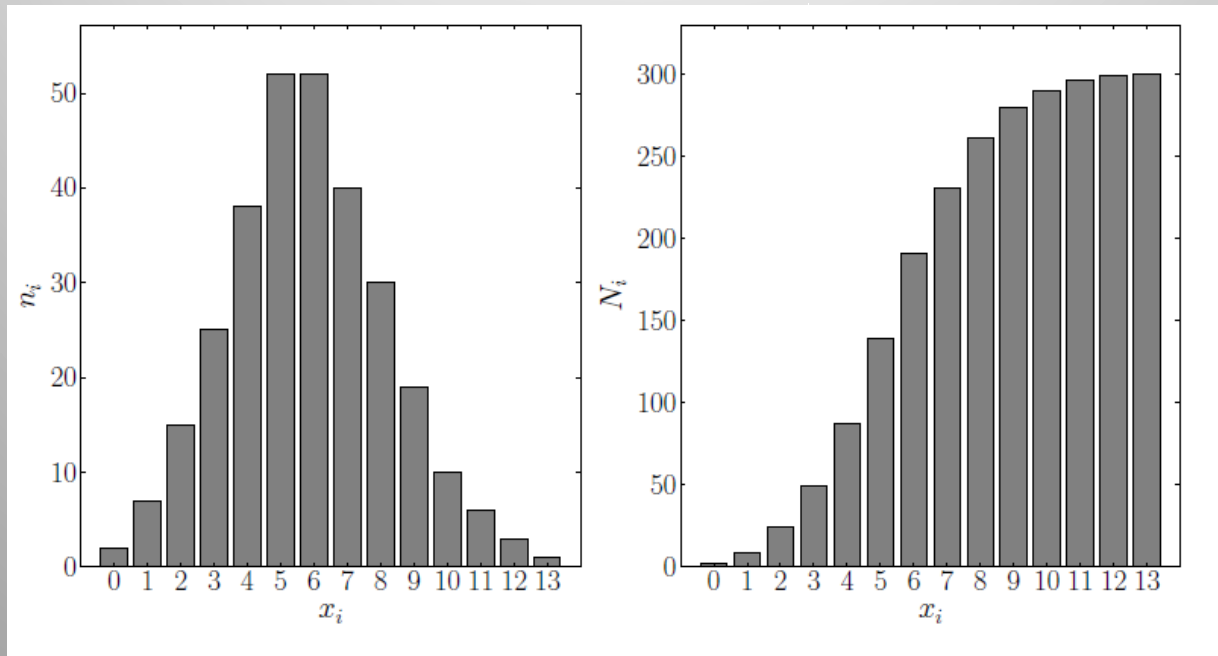
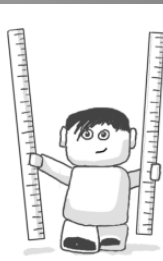
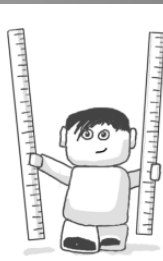
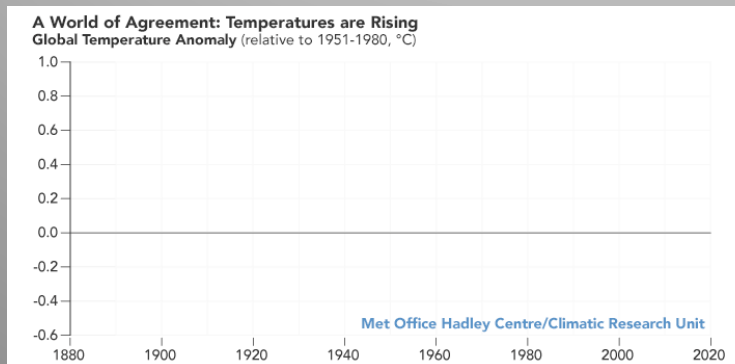
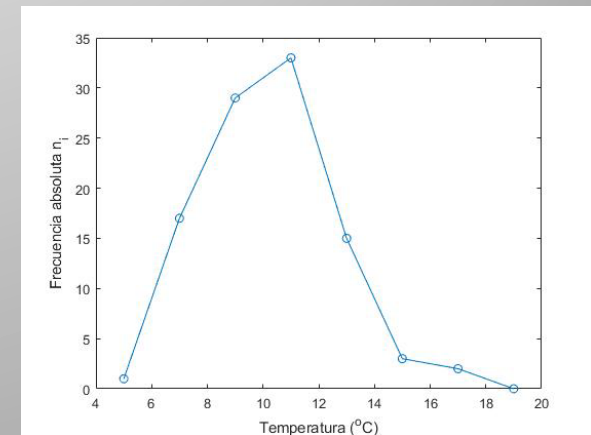
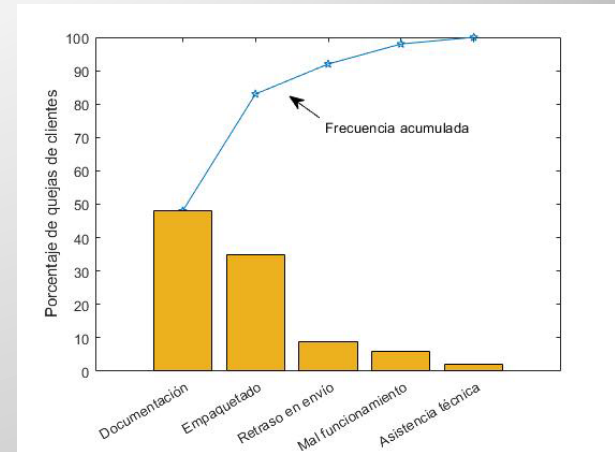


Diagrama de barras con la representación de la frecuencia absoluta (izq) y frecuencia absoluta acumulada (der).



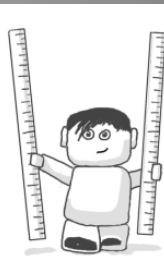
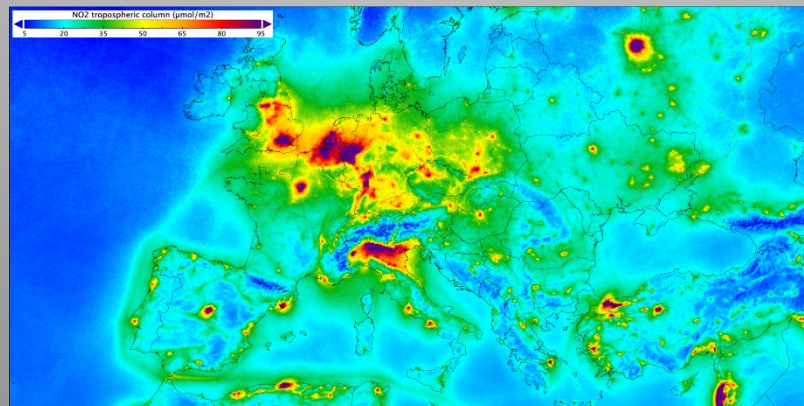
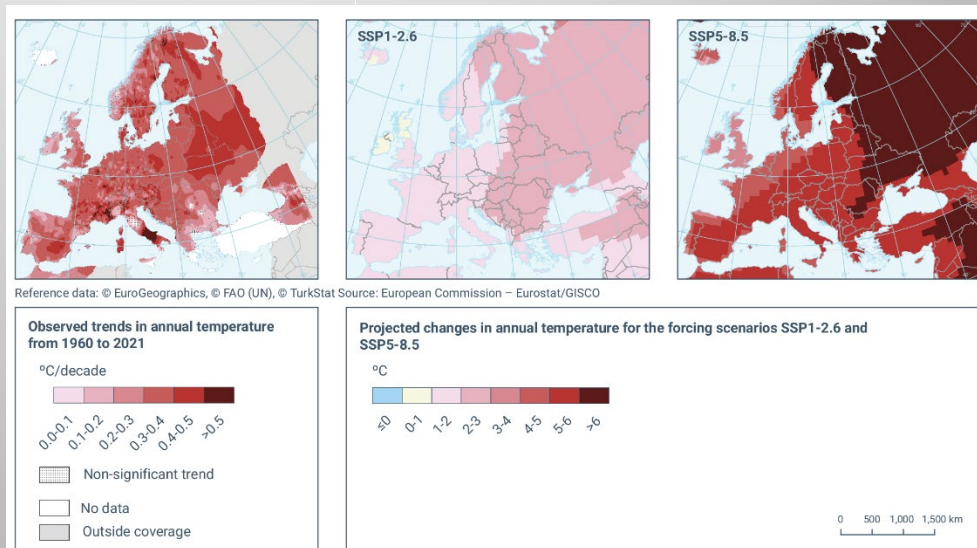
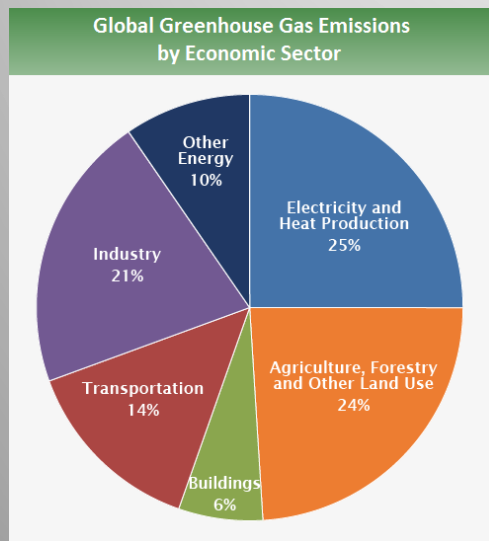
Representación gráfica

- **Diagramas de Pareto:** asocia rectángulos de altura igual a la frecuencia relativa a las clases o eventos ordenados según su frecuencia, en orden descendiente.
- **Poligonos o curvas de frecuencia:** se trazan líneas rectas o curvas que unan los extremos superiores de las barras del diagrama de barras de frecuencias o del histograma correspondiente.
- **Series temporales o de evolución:** corresponde a la representación de una variable aleatoria con respecto a una variable que evolucione, normalmente el tiempo de la medida.



Representación gráfica

Diagramas de sectores o circular, pictogramas, cartogramas, ... : otras representaciones estadísticas variadas que permiten representar los datos mediante la equivalencia de áreas, identificación ideográfica con los posibles resultados, localización geográfica de los datos, ...



Representación gráfica



Alluvial Diagram

The flow chart type is useful for showing changes over time across categories. The geometry is similar to a stacked area chart but with spacing and order changes.



Barcode Chart

A compact way to show a distribution, line segments represent individual points along a single axis.



Bubble Chart

Points are placed like in a scatter plot, but they are sized based on the square root of the data value.



Area Chart

The space between the data and the baseline is filled with a color or a pattern, usually emphasizing counts or percentages for a single variable.



Baseline Chart

It's like a line chart, but it specifically uses a reference point, which becomes the baseline, to compare all the other categories to.



Bump Chart

It looks like a line chart. However, the focus is specifically on ranks, usually over time.



Bar Chart

The old standby. The classic. The bar height or length represents data. The baseline starts at zero.



Beeswarm

It emphasizes individual points in a distribution instead of binning them like a histogram.



Calendar

A familiar, everyday layout to show data over time. It is especially useful to show a distribution over time.



Bar Chart Race

The standard bar chart is animated to show rankings and change over time. The alternative is a multi-line chart.



Box Plot

The classic chart of quartiles, median, minimum, and maximum shows a basic view of distributions.



Cartogram

Regions scale based on data instead of preserving geographic area.



Chord Diagram

Show connections between nodes.



Difference Chart

Fill the space between lines to highlight the greater-than and less-than differences over time.



Dot Plot

A generalized form of the scatter plot, the dots can be placed in various coordinate systems.



Choropleth Map

Regions are colored by a variable. Be careful with your color choices, as they can quickly change what the map shows.



Donut Chart

It is like a pie chart. But with a hole. It looks like a donut. That's why it's called a donut chart.



Frequency Trails

Area or line chart value axes overlap to fit more in a space and add another visual aspect to the magnitude of peaks. Also known as: ridgeline chart, joyplot.



Connected Scatter Plot

A cross between a scatterplot and a time series. Show two variables over time.



Dot Density Map

Dots are placed randomly within regions to show the density of populations. The dots and spacing allow for multiple groups to be shown at once.



Gantt Chart

It's a type of bar chart that shows the end and start times. The bars are offset by the former.



Density Plot

Look at distributions under the curve.



Dot Map

Each dot represents a data point. Collectively, geographic patterns emerge.



Glyph Chart

Symbols are used and various dimensions represent different variables in a dataset.



Grid Map

Instead of using geographic boundaries, same-size cells are used to represent areas to provide equal visual attention to all.



Line Chart

Typically used to show trends over time, the slope of the line between two points shows patterns of change.



Network Graph

Nodes and edges show connections, typically positioned to show strength of relationships.



Heatmap

Cells or bins are colored based on data. As with all visualization types that use color as the main visual encoding, choose shades carefully.



Line Map

Connect the lines in geographic space and see where they go.



Histogram

It looks like a bar chart, but it reads differently. The baseline is continuous instead of discrete categories, which allows one to see distributions.



Mosaic Plot

Also known as a Marimekko diagram, this chart uses the width and height of rectangles to represent separate variables. It can be useful to represent multidimensional data.



Horizon Graph

Compact the area chart by slicing it horizontally, and then shifting the slices to baseline zero. It's like a combo area chart and heatmap.



Moving Bubbles

Making use of a force-directed graph, bubbles move between different clusters to show grouping over time.



Packed Bubble Chart

Circle size represents data like with a bubble chart, but there is typically no x-y axis. Instead position often represents grouping or is used to maximize space.



Parallel Coordinates

Differing scales run alongside each other, sometimes showing relationships between multiple variables through connections. Can be tricky though.

Descripciones detalladas de representaciones gráficas de datos:
<https://flowingdata.com/chart-types/>

