

## Apéndice. Teoría de errores y presentación de resultados

### 1 Introducción

Este apéndice trata de dos aspectos esenciales a la hora de elaborar una Memoria de Prácticas y de una presentación de cálculos y resultados en general. El primero de estos dos aspectos es la forma de expresar resultados numéricos, así como gráficas y tablas adjuntas. El segundo es cómo determinar la incertidumbre en una cantidad, a fin de indicar la fiabilidad de los resultados y los márgenes de error.

El apéndice se ha estructurado en dos tipos de secciones. Básicas, donde las fórmulas y métodos pertinentes son enunciados y razonados pero omitiendo los detalles matemáticos, y avanzadas, en las que se explica el origen de las fórmulas indicadas y se plantea cómo se generalizan a situaciones en las que puedan no ser aplicables las fórmulas básicas. Las secciones avanzadas han sido indicadas con un asterisco (\*).

### 2 Incertidumbre en los datos

En Ingeniería y en las Ciencias Experimentales, a diferencia de en Matemáticas, la mayoría de las cantidades no son conocidas exactamente (salvo que se trate de factores numéricos como 2, 10 o  $\pi$ ). Los datos y resultados de cálculos físicos proceden en última instancia de la experiencia y por tanto están sujetos a error.

Así, por ejemplo, la aceleración de la gravedad en la superficie terrestre no es exactamente  $9.8 \text{ m/s}^2$  sino una cantidad que varía de punto a punto y cuyo valor conocido está limitado por la precisión de los aparatos experimentales empleados para medirla.

Por esta razón, en Ingeniería o Física no existen las cantidades con infinitos decimales, ya que esto implicaría un conocimiento perfecto de un dato experimental. Esto quiere decir que si medimos, por ejemplo, un periodo de oscilación y obtenemos un valor de aproximadamente 3 s, y queremos hallar la frecuencia natural, su valor no puede ser

$$f = \frac{1}{T} = 0.\hat{3} \text{ Hz} = 0.333333 \dots \text{ Hz} \quad (1.1)$$

ya que esto significaría que conocemos *todos* los decimales de la inversa, aun cuando la cantidad original era solo aproximada. Un valor de

$$f = \frac{1}{T} = 0.3 \text{ S} \quad (1.2)$$

sería probablemente más correcto.

#### 2.1 Cifras significativas

El primer factor que nos da información sobre la certidumbre de un dato es el *número de cifras significativas*. Éstas son las que, como su nombre indica, nos dan información detallada sobre el valor de la cantidad.

Como regla básica (que luego matizaremos) podemos definirlo como el número de cifras del dato, sin contar los ceros iniciales o finales. Veamos algunos ejemplos:

2373 mm Tiene cuatro cifras significativas.

12.45 kg Tiene también cuatro cifras significativas. La posición del punto decimal es irrelevante. Por ello no hay que confundir el número de cifras significativas con el número de decimales.

0.00987 m Tiene tres cifras significativas. Los ceros iniciales nos informan del *orden de magnitud* de la cantidad, pero no de su precisión. Esto se ve más claramente en la *notación científica* de un número

$$0.00987 \text{ m} = 9.87 \times 10^{-3} \text{ m} \quad (1.3)$$

En esta notación, cada cantidad se expresa como el producto de dos: la *mantisa*, formada por todas las cifras significativas, siendo mayor que 1 y menor que 10, y el *exponente*, que nos informa del orden de magnitud de la cantidad, que podríamos escribir como 9.87 mm.

24.50 s Tiene cuatro cifras significativas. En este caso, el cero final es una cifra significativa, ya que si no sería superfluo. Por ello, no es lo mismo dar un resultado como 24.50 que darlo como 24.5 ya que la primera forma corresponde a una medida más precisa.

45000 m En este caso tenemos una situación ambigua, ya que el número de cifras significativas podría ser 2, 3, 4 o 5, según que los ceros finales aporten información sobre la mantisa o sólo sobre el orden de magnitud. Esta ambigüedad, de nuevo, desaparece si expresamos el número en notación científica, ya que no es lo mismo  $4.5 \times 10^4 \text{ m}$  que  $4.5000 \times 10^4 \text{ m}$ .

El último ejemplo nos muestra que la expresión normal de un número no proporciona suficiente información sobre el número de cifras significativas. Para completar esta información debemos acotar la posible *incertidumbre* de una cantidad

A la incertidumbre de una magnitud se la denomina tradicionalmente como el “error de la magnitud”. Esta terminología es inadecuada, porque da a entender que los resultados de una medida con error son incorrectos, cuando lo que realmente ocurre es que son imprecisos, debido a las limitaciones naturales de los aparatos y procesos de medida. Por ello, aunque en lo que sigue se usará el vocabulario tradicional, donde pone “error” hay que entender siempre “incertidumbre”.

## 2.2 Errores sistemáticos

Como primera causa de incertidumbre, que nos limita nuestros resultados, tenemos los llamados *errores sistemáticos*, que se producen cuando, por algún factor externo, todas nuestras medidas resultan desviadas en el mismo sentido. En este caso, si somos capaces de identificar y medir este agente externo, podemos restar su efecto a nuestros datos.

## 2.3 Bandas de error

Más común que los errores sistemáticos, son los *errores aleatorios*, debidos a las imprecisiones intrínsecas del dispositivo experimental o del procedimiento de medida y que unas veces actúan en un sentido y otras en el contrario. Estos errores llevan asociada una *banda de error*, esto es, unos márgenes entre los cuales puede oscilar la magnitud que estamos midiendo. Si indicamos que un resultado vale

$$L = 2.7 \pm 0.3 \text{ cm} \quad (1.4)$$

estamos diciendo que la longitud tiene una elevada probabilidad (convencionalmente se toma el 95%) de encontrarse en el intervalo entre 2.4 cm y 3.0 cm.

En algunas ocasiones las bandas de error pueden ser asimétricas, esto es, el límite inferior puede estar a diferente distancia del valor de la magnitud que el límite superior, indicando que tenemos la magnitud quizás mejor acotada por abajo que por arriba. Sin embargo, no es esto lo habitual. En lo que sigue nos atenderemos a bandas de error simétricas en torno a su valor central.

### Error absoluto y error relativo

Las bandas de error determinan lo que se conoce como *error absoluto*,  $E_x$ , de una cantidad  $x$ , definido como la mitad de la anchura de la banda de error. Este error absoluto posee las mismas unidades que la cantidad que limita.

Una cantidad derivada de interés es el llamado *error relativo* de la misma cantidad, definido como

$$\epsilon_x = \frac{E_x}{x} \quad (1.5)$$

a diferencia del error absoluto, el error relativo no posee unidades, sino que se mide en tantos por uno o tantos por ciento.

Así, por ejemplo, se nos puede decir que una resistencia de  $2200 \Omega$  posee un error de  $\pm 110 \Omega$  (que sería su error absoluto) o que tiene una *tolerancia* (esto es, un error relativo) del 5%. Ambas expresiones son equivalentes. En ambos casos

$$R = 2200 \pm 110 \Omega \quad (1.6)$$

¿Cuál de los dos errores es más importante? ¿Qué debemos procurar que sea más pequeño, el error absoluto o el relativo? En lo que a información se refiere ambos proporcionan la misma ya que si conocemos uno conocemos el otro (si conocemos el valor de  $x$ , evidentemente).

En cuanto a cuál es más “importante”, depende de lo que se desee. Habrá aparatos o experimentos en los que nos interese una alta precisión (esto es, un error absoluto muy pequeño) o situaciones en las que convenga reducir el error relativo.

Como regla general, la pequeñez del error relativo es más informativa que la del error absoluto. Así, si decimos que el error al medir una masa es de 3 g, no nos dice mucho, si no sabemos cuánto vale la masa medida. No es lo mismo un error de 3 g en 1 kg que en 10 g.

Existe no obstante una excepción importante, la de aquellas medidas en que el valor de la magnitud es aproximadamente nulo. Supongamos que estamos equilibrando una balanza de dos platillos, buscando que el fiel marque 0, en la posición de equilibrio. Medimos lo que marca empleando una escala de precisión 0.01 g y resulta una medida  $m = 1.00 \pm 0.01$  g. Esto nos da un error relativo del 1%. Supongamos que en una medida posterior de la posición del fiel obtenemos un resultado  $m = 0.00 \pm 0.01$  g, con lo que ahora el error relativo es infinito, y pierde todo significado. Esto se extiende a todos los experimentos en que se busque obtener un resultado nulo (que se basa en el equilibrio de dos cantidades, como las masas de los platos de la balanza).

De hecho, podemos decir que una medida es nula cuando el valor de la magnitud es mucho menor que su error absoluto, como mucho la mitad de este  $|x| < E_x/2$ , o lo que es equivalente, cuando su error relativo es superior al 200%.

## 2.4 Expresión de una cantidad con incertidumbre. Redondeo

La existencia de una banda de error nos limita el número de cifras significativas que podemos presentar. Por ejemplo, el siguiente resultado sería absurdo:

$$L = 8.4575308 \pm 0.2 \text{ cm} \quad (1.7)$$

ya que la banda de error nos dice que el resultado es incierto ya en su primera cifra decimal, por lo que dar siete no tiene ningún sentido, ya que todo lo que esté más allá de la primera es absolutamente incierto. No es que las siguientes cifras sean correctas o incorrectas, es que son desconocidas por completo.

Tampoco tiene mucho sentido excederse en la expresión del error. Por ejemplo, también sería incorrecta la expresión

$$L = 8.4575308 \pm 0.2189095 \text{ cm} \quad (1.8)$$

ya que la primera o quizás las dos primeras cifras del error nos dicen donde está la incertidumbre. Las cifras sucesivas suponen un refinamiento innecesario. Recordemos que el límite del 95% es convencional y no pasa nada si la banda de error delimita una banda del 94% o del 96%.

Por ello, es importante truncar las expresiones tanto del error como de la cantidad. Esto requiere una cierta técnica sistemática para redondear adecuadamente que consta de los siguientes pasos:

**1)** Se expresa la magnitud objeto de estudio y su error con todas las cifras conocidas.

**Ejemplo:** Supongamos que una medida de una longitud nos da el valor

$$L = 2.30408415 \text{ cm} \quad (1.9)$$

vamos a ver como se expresa el resultado para seis casos distintos de errores de esta cantidad, como pueden ser los siguientes:

(a).  $E_L = 0.002156 \text{ cm}$

(b).  $E_L = 0.03674 \text{ cm}$

(c).  $E_L = 0.2036 \text{ cm}$

(d).  $E_L = 2.87 \text{ cm}$

(e).  $E_L = 234 \text{ cm}$

(f).  $E_L = 0.00962 \text{ cm}$

(g).  $E_L = 0.257 \text{ cm}$

(h).  $E_L = 0.1 \text{ cm}$

**2)** Se examinan las dos primeras cifras significativas del error (esto es, descontando los ceros situados a la izquierda del número). Si, como consecuencia del proceso de medida solo disponemos de una cifra, nos quedamos solo con esta. No tenemos “permiso” para inventarnos cifras significativas.

**Ejemplo (cont.):** Para los seis errores anteriores las cifras a considerar serán:

(a).  $E_L = 0.002156 \text{ cm}$

(b).  $E_L = 0.03674 \text{ cm}$

(c).  $E_L = 0.2036 \text{ cm}$

(d).  $E_L = \underline{2.8}7$  cm

(e).  $E_L = \underline{23}4$  cm

(f).  $E_L = 0.009\underline{6}2$  cm

(g).  $E_L = 0.\underline{25}7$  cm

(h).  $E_L = 0.\underline{1}$  cm

**3)** Si el conjunto de las dos cifras seleccionadas es un número menor o igual a 25, se conservan ambas. Si es un número mayor que 25, se conserva sólo la primera. El resto de las cifras del error se elimina. Téngase cuidado de redondear adecuadamente la última cifra restante (hacia abajo si lo que viene detrás es menor que 5, hacia arriba si es mayor o igual).

En los ejemplos anteriores, hay un pequeño problema con el penúltimo caso. En principio hay que retener dos cifras (25) pero al redondear el 7 siguiente nos resulta 26, que obligaría a retener una sola cifra (un 3, tras el redondeo del 6). Aunque en los casos límite tampoco es tan importante dejar una cifra o dos (ya que el 25 es de nuevo una convención), mantendremos el criterio de que realmente se trata de un número superior a 25 (sería 25.7) y por tanto se redondea a una sola cifra (3).

Si, como en el último caso, de entrada solo disponemos de una cifra significativa de error, nos quedamos solo con ella, aunque sea un 1 o un 2. La situación típica en que ocurre es en una medida directa con un aparato que tiene una incertidumbre fijada. Por ejemplo, medimos una longitud con una regla graduada en milímetros, de forma que la incertidumbre es  $\pm 1$  mm, aunque 10 sea un número menor que 25 no podemos añadir el 0, ya que no sabemos si es un 0, un 2 o cualquier otra cifra. Por tanto nos quedamos solo con una cifra de error también en este caso.

**Ejemplo (cont.):** Esto nos deja con los siguientes errores

(a).  $E_L = 0.0022$  cm

(b).  $E_L = 0.04$  cm

(c).  $E_L = 0.20$  cm

(d).  $E_L = 3$  cm

(e).  $E_L = 230$  cm

(f).  $E_L = 0.010$  cm

(g).  $E_L = 0.3$  cm

(h).  $E_L = 0.1$  cm

Obsérvese la diferencia entre colocar un cero de relleno para conservar el orden de magnitud correcto (caso (e)) y conservarlos al final cuando añaden información (casos (c) y (f))

**4)** Ya se ha redondeado el error. Ahora debe redondearse la magnitud. Para ello debe verse en qué posición se encuentra la última cifra del error. Hasta esa misma posición debe retenerse la cantidad original. Las cifras siguientes se desprecian, teniendo cuidado de redondear la última conservada.

**Ejemplo (cont.):** Para los siete casos anteriores nos quedan las siguientes expresiones para la longitud:

(a).  $L = 2.3041 \text{ cm}$

(b).  $L = 2.30 \text{ cm}$

(c).  $L = 2.30 \text{ cm}$

(d).  $L = 2 \text{ cm}$

(e).  $L = 0 \text{ cm}$

(f).  $L = 2.304 \text{ cm}$

(g).  $L = 2.3 \text{ cm}$

(h).  $L = 2.3 \text{ cm}$

Nótese cómo de nuevo se conservan los ceros cuando estos son significativos (pues expresan que esa cifra es efectivamente un cero). En el caso (e), en el cual el error es dos órdenes de magnitud más grande que la cantidad, debemos rellenar ésta con ceros a la izquierda (como  $002.3 \dots \text{ cm}$ ) y luego redondear las primeras cifras (que son nulas) dejando como resultado final un cero.

**5)** Por último, se expresan conjuntamente la cantidad, su error y sus unidades.

**Ejemplo (cont.):** Esto nos da:

(a).  $L = 2.3041 \pm 0.0022 \text{ cm}$

(b).  $L = 2.30 \pm 0.04 \text{ cm}$

(c).  $L = 2.30 \pm 0.20 \text{ cm}$

(d).  $L = 2 \pm 3 \text{ cm}$

(e).  $L = 0 \pm 230 \text{ cm}$

(f).  $L = 2.304 \pm 0.010 \text{ cm}$

(g).  $L = 2.3 \pm 0.3 \text{ cm}$

(h).  $L = 2.3 \pm 0.1 \text{ cm}$

6) Una presentación alternativa y más compacta consiste en expresar la medida y, al final, entre paréntesis el error que afecta a la última o últimas cifras. Esta notación es especialmente útil cuando se dispone de un espacio limitado, como una casilla de un formulario.

**Ejemplo (cont.):** De esta forma quedaría como

(a).  $L = 2.3041(22)$  cm

(b).  $L = 2.30(4)$  cm

(c).  $L = 2.30(20)$  cm

(d).  $L = 2(3)$  cm

(e).  $L = 0(230)$  cm

(f).  $L = 2.304(10)$  cm

(g).  $L = 2.3(3)$  cm

(h).  $L = 2.3(1)$  cm

7) Cuando se expresa un resultado en la notación científica, es preferible agrupar el valor de la magnitud y su error en una sola mantisa, en lugar de como dos cantidades separadas, es decir que en vez de

$$t = 3.57 \times 10^{-4}\text{s} \pm 3 \times 10^{-6}\text{s}$$

debería escribirse

$$t = (3.57 \pm 0.01) \times 10^{-4}\text{s}$$

o, de forma aun más compacta,

$$t = 3.57(1) \times 10^{-4}\text{s}$$

De esta forma se consigue ver de una forma más clara la incertidumbre de la cantidad y a qué cifra significativa afecta.

Las técnicas de redondeo deben practicarse hasta que se conviertan en un procedimiento automático. No es necesario indicar explícitamente los pasos intermedios.

De todas formas, aunque en los resultados finales las cantidades deben aparecer redondeadas, es conveniente conservar en los borradores las cantidades obtenidas, con todas sus cifras, si pretendemos realizar algún cálculo posterior con ellas.

Como regla básica de conducta aplicaremos que los cálculos intermedios se hacen con todas las cifras y los resultados finales se presentan redondeados.

### 3 Cálculo de errores

En la sección anterior hemos comentado la necesidad de indicar la banda de error de una cantidad y la forma de expresarla, pero no de dónde sale este error ni cómo podemos determinar su valor. De esto nos encargaremos en esta sección.

### 3.1 Caso de una sola medida

La situación más sencilla es aquella en la que para determinar una cantidad hacemos una sola medida directa con un aparato.

En este caso, el error depende de la *precisión* del aparato, definida como la mínima separación que se puede conseguir entre medidas. Esto es, si un amperímetro puede medir 2.4 A, 2.5 A, 2.6 A, . . . , la precisión es de 0.1 A.

No siempre la precisión coincide con el número de decimales que aparecen en el resultado. Por ejemplo, si un termómetro digital proporciona los siguientes valores sucesivos de temperatura: 25.0 °C, 25.5 °C, 26.0 °C, su precisión es de 0.5 °C y no de una décima de grado. Por ello es necesario estar atento a los intervalos entre valores sucesivos.

Una vez determinada la precisión de un aparato, el *error* de una cantidad medida una sola vez lo tomaremos, para todos los aparatos como igual a esta precisión. Por ejemplo, para una regla graduada en milímetros, si hacemos una medida con resultado de 87 mm, le asignamos el valor

$$L = (87 \pm 1) \text{ mm}$$

En ocasiones, esta estimación puede parecer demasiado conservadora. Si tenemos una aguja sobre una escala graduada en centímetros, es seguro que el ojo humano puede apreciar una mayor precisión y afirmar si la aguja está más cerca de una línea o de la siguiente o a cuantos milímetros se halla. No obstante, mantendremos este criterio conservador por diversas razones:

- Por mantener un criterio uniforme, que emplee exclusivamente los valores medidos y no la apreciación subjetiva del observador.
- Porque siempre puede haber otras fuentes de incertidumbre, por la propia naturaleza del experimento o el funcionamiento interno del aparato de medida (que a menudo desconocemos).

### 3.2 Varias medidas de la misma magnitud

En muchos casos, una sola medida de una cantidad no nos da suficiente información sobre el error de una magnitud. Puede ocurrir que hagamos una medida única, con un error de medida muy pequeño, pero, por la propia naturaleza del experimento o del dispositivo de medida, sepamos que ese resultado puede haberse desviado del valor “real” de la cantidad, esto es, que haya una incertidumbre intrínseca, no achacable a la imprecisión del aparato de medida. En este caso, obtendremos una mejor aproximación efectuando una serie de medidas sucesivas y elaborando un estudio estadístico, que nos dirá como se distribuyen los resultados. También puede ocurrir que hayamos obtenido diferentes valores de la misma magnitud por procedimientos diferentes y queramos combinar los resultados en uno solo.

Si tenemos  $n$  valores,  $x_i$ , de la misma magnitud  $x$ , que se supone que representan al mismo valor real y que han sido medidos *independientemente* (esto es, que el hecho de realizar una de las medidas no afecta a los resultados de las demás medidas), tomaremos como estimación de la cantidad  $x$  la media aritmética de las distintas medidas

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{S_x}{S} \quad (1.10)$$

donde hemos definido las cantidades  $S$  y  $S_x$  (que luego usaremos más ampliamente) como

$$S = \sum_i 1 = n \quad S_x = \sum_i x_i \quad (1.11)$$



(en los sumatorios sobre  $i$  se entenderá siempre que se extiende desde 1 hasta  $n$ , el número de datos).

**Ejemplo:** Supongamos que las sucesivas medidas de una distancia hechas con una regla de precisión 1 mm, nos dan

$$L = 27, 28, 27, 25, 28, 26 \text{ (mm)}$$

teniendo cada valor un error individual de  $E_{L_i} = \pm 1$  mm en este caso asignaremos como valor real de la longitud

$$S = 6 \quad S_x = 161 \text{ mm} \quad \bar{L} = 161/6 \text{ mm} = 26.8 \text{ mm}$$

El cálculo exacto daría  $\bar{L} = 26.8333\dots$  mm. A la hora de expresar el resultado final no se puede dar con más decimales que los que permite la precisión. Sin embargo, para los cálculos intermedios, como los del error que veremos ahora, es conveniente anotar este valor con más decimales (dos o tres más, como mínimo).

Asignaremos como *error* de la cantidad  $x$  el doble de la desviación cuadrática media de la media de los datos, estimado como

$$E_x = 2\sigma_{\bar{x}} = 2\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}} \quad (1.12)$$

La justificación de esta fórmula se indica en la sección 4.1.

Esta cantidad no viene implementada directamente en las calculadoras u hojas de cálculo (como *Excel*), pero es inmediata a partir de la llamada desviación estándar de la muestra,  $\sigma_{n-1}$  (en las calculadoras  $x\sigma_{n-1}$ ) o de la llamada desviación estándar de la población,  $\sigma_n$  (en las calculadoras  $x\sigma_n$ ), que sí aparecen. Estas dos funciones están definidas como

$$\sigma_n = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}} \quad \sigma_{n-1} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} \quad (1.13)$$

de forma que el error puede calcularse a partir de cualquiera de ellas como

$$E_x = 2\frac{\sigma_n}{\sqrt{n-1}} = 2\frac{\sigma_{n-1}}{\sqrt{n}} \quad (1.14)$$

Caso de que no estén disponibles, el error, puede hallarse a partir de los sumatorios, ya que al desarrollar tenemos

$$E_x = 2\sqrt{\frac{(SS_{xx} - S_x^2)}{S^2(S-1)}} \quad S_{xx} = \sum_{i=1}^n x_i^2 \quad (1.15)$$

**Ejemplo:** Para los mismas medidas de la longitud

$$L = 27, 28, 27, 25, 28, 26 \text{ (mm)}$$

el cálculo del error seguiría los siguientes pasos (que la calculadora hace directamente)

$i$	$x_i$ (mm)	$x_i - \bar{x}$ (mm)	$(x_i - \bar{x})^2$ (mm <sup>2</sup> )
1	27	0.167	0.0278
2	28	1.167	1.3611
3	27	0.167	0.0278
4	25	-1.833	3.3611
5	28	1.167	1.3611
6	26	-0.833	0.6944
$n$	$\sum_i x_i$ (mm)	$\sum_i (x_i - \bar{x})$ (mm)	$\sum_i (x_i - \bar{x})^2$ (mm <sup>2</sup> )
6	161	0	6.8333
	$\bar{x}$ (mm)		$\sum_i (x_i - \bar{x})^2 / (n - 1)$ (mm <sup>2</sup> )
	26.833		1.3667

Por lo que la desviación de la muestra sería

$$\sigma_{n-1} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}} = 1.169 \text{ mm}$$

y la estimación del error de la media

$$E_{\bar{x}} = 2 \frac{\sigma_{n-1}}{\sqrt{n}} = 0.954 \text{ mm}$$

que redondeado nos da, para el resultado final

$$L = 26.8 \pm 1.0 \text{ mm}$$

Esta estimación para el error no puede resultar inferior a la precisión del aparato de medida. Supongamos que medimos la longitud de una mesa con una regla graduada en centímetros. Lo hacemos 10 veces y en todos los casos nos resulta  $L = 123 \pm 1$  cm. No podemos concluir que el error es nulo (que es lo que daría la expresión anterior), sino que carecemos de la precisión suficiente para determinarlo.

Por ello, adoptaremos el siguiente criterio, si hacemos  $n$  medidas, cada una de ellas con el mismo error  $E_{x_i}$  (según vimos en el apartado anterior) y el error estadístico, definido en la ecuación (1.12), nos da una cantidad  $E_{\text{est}}$ , tomaremos como error de la cantidad  $x$  a la mayor de estas dos cantidades, esto es

$$E_x = \text{máx}(E_{x_i}, E_{\text{est}}) \quad (1.16)$$

Con este error, la cantidad  $x$  se indicará, convenientemente redondeada, como

$$x = \bar{x} \pm E_x \quad (1.17)$$

### Expresión relativa a la media y error de redondeo

Una forma más compacta de escribir la expresión del error es introduciendo las variables *relativas a la media*, que definimos como

$$X_i = x_i - \bar{x} \quad (1.18)$$

de forma que el error queda como

$$E_x = 2\sqrt{\frac{\sum_i X_i^2}{n(n-1)}} = 2\sqrt{\frac{S_{XX}}{S(S-1)}} \quad (1.19)$$

Si es necesario calcular explícitamente esta cantidad (porque no se dispone de la función  $x\sigma_{n-1}$ ) es preferible emplear la expresión con las variables relativas a la media, ya que presentan un menor *error de redondeo*.

Este error numérico se da cuando debemos restar dos cantidades muy próximas entre sí. Por ejemplo, supongamos que tenemos la siguiente lista de valores de una temperatura

$$T = 298.1, 298.2, 298.2, 298.3 \text{ K}$$

y tratamos de hallar la media y el error de esta temperatura.

La media de estos valores nos da

$$\bar{T} = \frac{1}{4}(298.1 + 298.2 + 298.2 + 298.3) \text{ K} = 298.2 \text{ K}$$

y para el error tenemos, en primer lugar, que

$$S = 4 \quad S_x = 1198.2 \text{ K} \quad S_{xx} = 355693 \text{ K}^2$$

$$SS_{xx} = 1422771.92 \text{ K}^2 \quad S_x^2 = 1422771.84 \text{ K}^2$$

$$SS_{xx} - S_x^2 = 0.08 \text{ K}^2$$

$$E_{\bar{T}} = 2\sqrt{\frac{0.08}{4^2 \cdot 3}} \text{ K} = 0.08 \text{ K}$$

Vemos que a la hora de calcular el error podemos tener problemas al hacer la resta de dos números grandes como  $SS_{xx}$  y  $S_x^2$  ya que si la calculadora retiene solo 7 cifras perdemos los dos últimos y nos resulta (erróneamente) un error nulo.

Por ello, desde el punto de vista numérico es preferible emplear la cantidades relativas a la media, con las que el cálculo quedaría como

$$T - \bar{T} = -0.1, 0.0, 0.0, 0.1 \text{ K}$$

$$S_{XX} = 0.02 \text{ K}^2$$

$$E_{\bar{T}} = 2\sqrt{\frac{0.02}{2 \cdot 3}} \text{ K} = 0.08 \text{ K}$$

En este caso no tenemos que preocuparnos de  $S_X$  (que es siempre nulo) ni de los errores de redondeo (aunque estos pueden aparecer al hallar  $T - \bar{T}$ , esta situación se da en muy pocos casos).

### 3.3 Caso de una magnitud función de otra(s)

En los dos apartados anteriores hemos visto cómo obtener y expresar una cantidad sujeta a error a partir de las medidas directas de dicha magnitud.

En muchos casos, no obstante, debemos trabajar con cantidades indirectas, obtenidas a partir de las medidas. Estas pueden depender de una sola variable experimental  $z = f(x)$  o de varias magnitudes diferentes  $z = f(x, y, \dots)$ .

### Función de una sola variable

A modo de ejemplo, supongamos que, dado el diámetro  $D_0 = 12.5 \pm 0.3$  cm de un cable, debemos determinar el área de su sección, de la que sabemos que es circular. El cálculo de la sección es inmediato

$$A = \frac{\pi D_0^2}{4} = 122.718 \text{ cm}^2 \quad (1.20)$$

pero si el diámetro es una cantidad con una cierta banda de error, ¿cuál es la banda de error del área?

Podemos hacer una primera estimación. Supongamos que  $E_D$  es el error del diámetro, de forma que hay un 95% de probabilidad de que  $D$  se encuentre en el intervalo

$$D \in (D_0 - E_D, D_0 + E_D) = (12.2 \text{ cm}, 12.8 \text{ cm}) \quad (1.21)$$

de forma que el área posee un 95% de probabilidad de encontrarse en el intervalo

$$A \in \left( \frac{\pi(D_0 - E_D)^2}{4}, \frac{\pi(D_0 + E_D)^2}{4} \right) = (116.899 \text{ cm}^2, 128.680 \text{ cm}^2) \quad (1.22)$$

Esta estimación, aun siendo correcta, tiene el inconveniente de que proporciona una banda de error asimétrica, ya que el valor de  $A$  indicado antes no se encuentra en el punto medio de este intervalo (que se halla en  $122.789 \text{ cm}^2$ ). O, equivalentemente, que el punto medio del intervalo no es la mejor aproximación para el área.

Podemos simplificar las fórmulas en el caso de que el error  $E_x$  sea mucho menor que la propia cantidad  $x$  (lo que ocurre a menudo, pero no siempre). En este caso, tal como se ve en la sección 4.2, resulta, en primer orden

$$E_f = \left| \frac{df}{dx} \right| E_x \quad (1.23)$$

de forma que la cantidad derivada se expresará como

$$f = f(x) \pm \left| \frac{df}{dx} \right| E_x \quad (1.24)$$

El valor absoluto de la derivada se debe a que debemos obtener siempre un error positivo de la magnitud.

Así, para el caso del área, tendremos

$$\frac{dA}{dD} = \frac{\pi D}{2} \quad A = \frac{\pi D^2}{4} \pm \frac{\pi D E_D}{2} = 122.718 \pm 5.890 \text{ cm}^2 = 123 \pm 6 \text{ cm}^2 \quad (1.25)$$

Puede comprobarse que esto es lo que resulta en el intervalo que hallamos antes si suponemos  $E_D \ll D$ .

De esta fórmula hay varios casos particulares de interés

- Para una variable proporcional a otra

$$y = Kx \quad (1.26)$$

con  $K$  una constante *sin error* (por ejemplo, 2 ó  $\pi$ )

$$E_y = K E_x \quad (1.27)$$

- Para la función exponencial

$$y = ae^x \quad (1.28)$$

resulta

$$E_y = ae^x E_x = y E_x \quad (1.29)$$

o, lo que es equivalente

$$\epsilon_y = \frac{E_y}{y} = E_x \quad (1.30)$$

esto es, el error relativo de  $y$  coincide con el error absoluto de  $x$ .

- Para el logaritmo tenemos la relación inversa

$$y = \ln(x) \quad (1.31)$$

$$E_y = \frac{E_x}{x} = \epsilon_x \quad (1.32)$$

El error absoluto de  $y$  coincide con el relativo de  $x$ .

### **Función de varias variables**

Supongamos ahora que debemos hallar el volumen de un recipiente cilíndrico, del cual hemos medido por separado el diámetro de la base y la altura, con valores

$$D = 12.5 \pm 0.1 \text{ cm} \quad h = 10.2 \pm 0.2 \text{ cm} \quad (1.33)$$

El volumen valdrá

$$V = \frac{\pi D^2 h}{4} \sim 1251.7 \text{ cm}^3 \quad (1.34)$$

Para hallar el error de esta cantidad debemos tener en cuenta que tanto  $D$  como  $h$  están sujetos a error, por lo que el error de  $V$  es una combinación de ambos.

Dada una función  $z = f(x, y, \dots)$ , la fórmula para el error de  $z$  en el caso de que  $x, y, \dots$  sean variables independientes (esto es, que la medida de una no afecta a la medida de la otra) es una generalización del caso de una sola variable

$$E_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 E_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 E_y^2 + \dots} \quad (1.35)$$

Podemos ver como para el caso de una sola variable la expresión se reduce a la que ya conocemos.

Así para el caso del volumen tendríamos

$$\frac{\partial V}{\partial D} = \frac{\pi D h}{2} = 200.3 \text{ cm}^2 \quad \frac{\partial V}{\partial h} = \frac{\pi D^2}{4} = 122.7 \text{ cm}^2 \quad (1.36)$$

$$E_V = \sqrt{200.3^2 \times 0.1^2 + 122.7^2 \times 0.2^2} \text{ cm}^3 = 31.7 \text{ cm}^3 \quad (1.37)$$

y la expresión final del volumen es

$$V = 1250 \pm 30 \text{ cm}^3 \quad (1.38)$$

Una comparación interesante es la del error relativo de las tres cantidades

$$\epsilon_D = 0.8\% \quad \epsilon_h = 2.0\% \quad \epsilon_V = 2.4\% \quad (1.39)$$

Como indicación de si un resultado puede ser correcto, podemos observar que el error relativo del resultado final debe ser algo mayor que el mayor de los errores relativos individuales. Si resulta un error relativo mucho menor es que algo falla.

Como casos particulares de la fórmula del error de una función tenemos los siguientes.

- Si tenemos una función de la forma

$$u = x \pm y \pm z \pm \dots \quad (1.40)$$

esto es, formada exclusivamente por sumas o restas, el error absoluto de  $u$  vale

$$E_u = \sqrt{E_x^2 + E_y^2 + E_z^2 + \dots} \quad (1.41)$$

es decir, la suma cuadrática de los errores individuales.

- Para una función del tipo

$$u = \frac{x \cdot y \cdot \dots}{z \cdot t \cdot \dots} \quad (1.42)$$

formada solamente por productos y cocientes (pero ninguna suma o resta) operando en la expresión del error se llega a

$$\epsilon_u = \sqrt{\epsilon_x^2 + \epsilon_y^2 + \epsilon_z^2 + \dots} \quad (1.43)$$

esto es, el error *relativo* de  $u$  es la suma cuadrática de los errores *relativos* individuales. Es decir, para hallar el error absoluto de  $u$  (que es el que acompaña a la medida) de esta forma habría que calcular previamente los errores relativos de cada factor, computar el error relativo del producto y, multiplicando por la propia cantidad  $u$  llegaríamos a  $E_u$ , error absoluto de  $u$ .

## 4 (\*) Justificación de las fórmulas

### 4.1 Error estadístico

Antes hemos enunciado las expresiones para, dado un conjunto de medidas  $x_i$  de una misma magnitud  $x$ , establecer qué valor le asignamos a la medida y a su error como

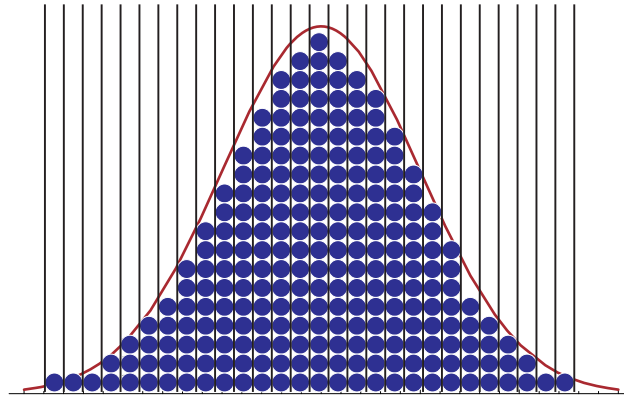
$$\bar{x} = \frac{1}{n} \sum_i x_i = \frac{S_x}{S} \quad (1.44)$$

$$E_x = 2\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}} = 2\sqrt{\frac{SS_{xx} - S_x^2}{S^2(S-1)}} = 2\sqrt{\frac{S_{XX}}{S(S-1)}} \quad (1.45)$$

con  $X_i = x_i - \bar{x}$  la variable relativa a la media.

La justificación de estas fórmulas proviene de una hipótesis básica: que existe una cierta distribución para las distintas medidas.

Supongamos que hiciéramos muchas medidas de esta magnitud y fuéramos clasificando los resultados según la frecuencia con que aparece cada valor. Cuando el número de medidas tiende a infinito obtenemos lo que se llama la *función de distribución*,  $p(x)$ .



Esta función nos permite obtener la probabilidad de un resultado concreto, de un intervalo de valores, o la probabilidad para una variable  $A$ , función de  $x$ .

Dada la función de distribución de la variable  $x$ ,  $p(x)$ , el *valor esperado* (para entendernos, el promedio) de la variable  $A(x)$  viene dado por

$$\langle A(x) \rangle = \int A(x)p(x)dx \quad (1.46)$$

que, en palabras, significa que el promedio de  $A$  no es más que la suma de todos los valores, teniendo en cuenta que algunos valores de  $x$  aparecen más a menudo que otros.

Las funciones de distribución,  $p(x)$ , pueden ser de diferentes tipos, pero la más común es la llamada *distribución gaussiana*, con forma de campana, como la curva continua de la figura anterior.

Dada la función de distribución de una magnitud, dos son sus características principales:

**Media:** o *esperanza*. Equivale a la media aritmética de los resultados, cuando el número de medidas tiende a infinito. Para una distribución gaussiana, la media corresponde al valor de  $x$  para el máximo de la campana. Representaremos la media de la distribución como

$$\langle x \rangle = \alpha \quad (1.47)$$

**Varianza:** Se define como

$$\sigma^2 = \langle (x - \alpha)^2 \rangle \quad (1.48)$$

y da una medida de la dispersión de los datos alrededor del valor medio, esto es, de la anchura de la distribución.

Para ser más precisos, la raíz de la varianza se denomina la *desviación cuadrática media*,  $\sigma$ , y, para una distribución gaussiana, es igual a la mitad de la anchura de la campana, medida a una altura  $e^{-1/2} = 0.61$  del máximo.

Para el caso de una distribución gaussiana, si tomamos una banda de anchura  $2\sigma$  alrededor de la media, la probabilidad de que nuestras medidas queden dentro de este intervalo es de un 95.5%. Se entiende entonces que, adoptando el criterio del 95% para la banda de error, una muy buena aproximación para el error de una cierta cantidad es

$$A = \langle A \rangle \pm 2\sigma_A \quad (1.49)$$

Como criterio supondremos que, dado un conjunto de datos, que suponemos que sigue una cierta distribución de probabilidad, el valor “real” de la medida es la media de la distribución correspondiente.

El problema que surge es que la media de la distribución (y la desviación cuadrática media) sólo se conoce tras *infinitas* medidas, lo cual la hace inalcanzable.

Se trata entonces de, con un número *finito* de datos, hallar la mejor aproximación posible a la media.

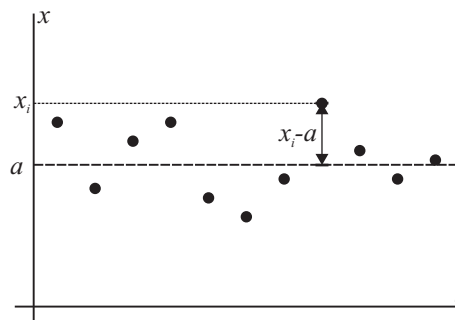
Una forma de obtenerla es mediante el *método de mínimos cuadrados*. Cuando se consideran infinitas medidas, se verifica que la cantidad

$$\chi^2(a) = \sum_{i=1}^{\infty} (x_i - a)^2 \quad (1.50)$$

es mínima cuando  $a = \alpha$ , esto es, para la media de la distribución. La idea es mantener esta expresión para un número finito de datos. La mejor estimación para la media de la distribución será aquel valor de  $a$  que haga mínima la cantidad

$$\chi^2(a) = \sum_{i=1}^n (x_i - a)^2 \quad (1.51)$$

Podemos interpretar este resultado gráficamente. Supongamos que anotamos las diferentes medidas como alturas en una escala. Se trata de hallar un valor que sea la mejor aproximación a todos los datos. Para cada valor tomamos la diferencia entre la medida y este valor. Será una cierta cantidad, que podemos considerar el error de cada dato individual. Elevándola al cuadrado (para que sea siempre positiva) y sumando para todos los datos, el valor mínimo nos dará la mejor aproximación simultánea a todos los datos, aunque puede que no coincida con ninguna de las medidas.



Aplicando la fórmula para el mínimo de una función tenemos

$$0 = \frac{d\chi^2}{da} = -2 \sum_i (x_i - a) = -2(S_x - aS) \quad (1.52)$$

donde hemos separado los dos sumandos y sacado factor común. El valor óptimo será entonces

$$a = \frac{S_x}{S} = \frac{1}{n} \sum_i x_i = \bar{x} \quad (1.53)$$

esto es, la media aritmética. Esto está en perfecto acuerdo con la idea de que la media de la distribución coincide con la media aritmética cuando el número de medidas tiende a infinito.

Para hallar el error de esta estimación, partimos de que si tenemos un conjunto de medidas independientes entre sí, se verifica que la varianza de la suma es la suma de las varianzas

$$\sigma_{x+y+\dots}^2 = \sigma_x^2 + \sigma_y^2 \quad (1.54)$$



por lo que

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} \sum_i \sigma_{x_i}^2 = \frac{n\sigma_x}{n^2} = \frac{\sigma_x^2}{n} \quad (1.55)$$

donde  $\sigma_x^2$  es la varianza de la distribución de los datos, que también debemos estimar.

Para hallar la aproximación a la varianza, puede hacerse un razonamiento similar al de la media, pero ligeramente más complicado. Sabemos que, en el límite de infinitas medidas, la varianza verifica

$$\sigma^2 = \lim_{n \rightarrow \infty} \sum_i \frac{(x_i - \alpha)^2}{n} \quad (1.56)$$

Si intentamos extender esta fórmula a un número finito de datos, debemos tener en cuenta que la propia media aritmética está sometida a error, de forma que la mejor aproximación a la varianza de los datos es

$$\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1} = \frac{S_{XX}}{S-1} = \frac{SS_{xx} - S_x^2}{S(S-1)} \quad (1.57)$$

La razón de que aparezca  $n-1$  en lugar de  $n$  en el denominador se debe precisamente a que  $\bar{x}$  es una aproximación a la verdadera media de la distribución y no coincide exactamente con ella.

La estimación de la varianza de la media será entonces

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{\sum_i (x_i - \bar{x})^2}{n(n-1)} = \frac{S_{XX}}{S(S-1)} = \frac{SS_{xx} - S_x^2}{S^2(S-1)} \quad (1.58)$$

La aproximación a la desviación cuadrática media será la raíz de esta cantidad, por lo que la mejor aproximación a la magnitud y su error será

$$x = \bar{x} \pm 2s_{\bar{x}} \quad (1.59)$$

que es la expresión que dimos anteriormente en (1.12).

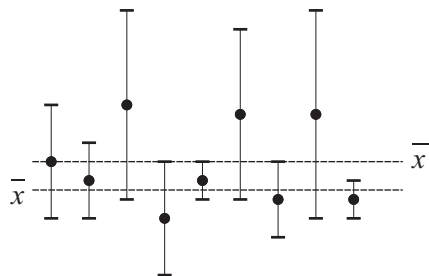
### Extensión al caso de datos con diferentes errores

Las fórmulas anteriores presuponen que todas las medidas pertenecen a la misma distribución y por tanto están sujetas al mismo error. Sin embargo, puede ocurrir que tengamos una serie de datos con errores diferentes y aun así estemos interesados en hallar la mejor aproximación a la media.

En este caso, al método de mínimos cuadrados aun es aplicable, pero la función que debemos minimizar es

$$\chi^2 = \sum_i \frac{(x_i - a)^2}{E_i^2} \quad (1.60)$$

siendo  $E_i$  el error del dato  $i$ . La interpretación de esta modificación es simple: tenemos más en cuenta aquellos datos que tienen menos error, mientras que los que tienen un error más grande aportan una contribución pequeña al sumatorio.



En la figura,  $\bar{x}'$  representa la media que obtendríamos si consideramos todos los datos con el mismo error, mientras que  $\bar{x}$  sería el resultado si se tiene en cuenta que los valores superiores son más inciertos que los inferiores.

La formula corregida para la media sigue siendo

$$\bar{x} = \frac{S_x}{S} \quad (1.61)$$

pero donde ahora los sumatorios valen

$$S = \sum_i \frac{1}{E_i^2} \quad S_x = \sum_i \frac{x_i}{E_i^2} \quad (1.62)$$

## 4.2 Error de una función

Anteriormente establecimos que si teníamos una variable  $x$  cuya medida y error eran de la forma

$$x = x_0 \pm E_x \quad (1.63)$$

y queríamos calcular, a partir de ella, una función  $y = f(x)$ , asignábamos como valor de la función a  $y_0 = f(x_0)$  y, como error,

$$E_y = \left| \frac{df}{dx} \right| E_x \quad (1.64)$$

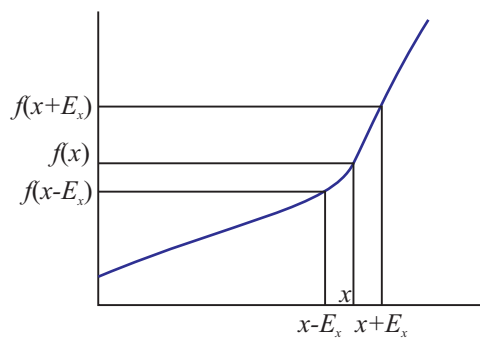
Para explicar esta expresión partimos, de que, con un 95% de probabilidad  $x$  está comprendida en el intervalo

$$x \in (x_0 - E_x, x_0 + E_x) \quad (1.65)$$

Por tanto,  $y = f(x)$  tendrá la misma probabilidad de encontrarse en el intervalo

$$y = f(x) \in (f(x_0 - E_x), f(x_0 + E_x)) \quad (1.66)$$

Sin embargo, como comentamos antes, no resulta una banda centrada en  $f(x_0)$



En la mayoría de los casos, no obstante, se cumple que el error es una cantidad pequeña y podemos hacer el desarrollo en serie de Taylor en torno al punto  $x_0$ . En este caso, sí resulta una banda simétrica. Aplicando esta función a los extremos del intervalo resulta aproximadamente

$$f(x_0 + E_x) \simeq f(x_0) + E_x \frac{df}{dx}(x_0) + \dots \quad f(x_0 - E_x) \simeq f(x_0) - E_x \frac{df}{dx}(x_0) + \dots \quad (1.67)$$

siendo el punto medio del intervalo

$$y_0 = \frac{1}{2} (f(x_0 + E_x) + f(x_0 - E_x)) = f(x_0) \quad (1.68)$$

y siendo el error la mitad de la distancia entre los extremos del intervalo de  $f(x)$

$$E_y = \frac{1}{2} |f(x_0 + E_x) - f(x_0 - E_x)| = \left| \frac{df}{dx} \right| E_x \quad (1.69)$$

que es la expresión que dimos antes.

Si ocurre que esta derivada es nula o el error es muy grande, habrá que retener más términos en el desarrollo en serie de Taylor o recurrir a la expresión original, admitiendo una barra de error asimétrica.

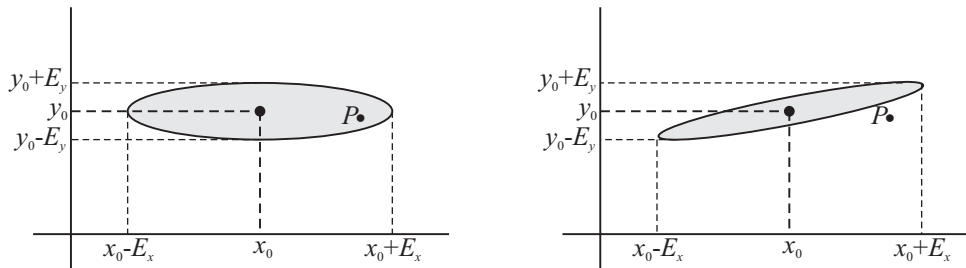
### 4.3 Funciones de varias variables. Importancia de la independencia

Para deducir la fórmula que nos da el error de una función de varias variables, es necesario hacer hipótesis adicionales.

Si tenemos una función de dos variables  $z = f(x, y)$  no nos basta con conocer los valores medios de  $x$  e  $y$  y los errores respectivos. Necesitamos además saber si estas variables son independientes o no.

Cuando tenemos dos variables, definimos una función de probabilidad conjunta  $p(x, y)$  que nos da la probabilidad de obtener simultáneamente los valores  $x$  e  $y$ .

Consideremos que dada la función de distribución, representamos en una gráfica la curva que nos da un 95% de probabilidad de obtener un valor de  $(x, y)$  próximo al valor "real". En la mayoría de los casos obtendremos una curva aproximadamente elíptica.



Supongamos las dos distribuciones de probabilidad de las figuras. En ambos casos las bandas de error de  $x$  e  $y$ , por separado, son las mismas. Sin embargo, en un caso es probable que se obtenga el punto  $P$ , mientras que en el otro caso no. La diferencia entre ambos casos es la dependencia entre  $x$  e  $y$ , esto es, el hecho que la probabilidad de los valores de  $y$  dependen de lo que se haya obtenido para  $x$ .

Para medir la dependencia entre las dos variables se define la *covarianza*

$$\sigma_{xy} = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle \quad (1.70)$$

y, a partir de esta, la correlación entre las dos variables

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (1.71)$$

Este coeficiente está comprendido en el intervalo  $(-1, 1)$ . Un valor nulo indica que las dos variables son independientes. El valor de  $y$  no depende en absoluto del valor de  $x$ , y la curva del 95% se convierte

en una elipse “horizontal”. Si  $r = +1$  las dos variables están absolutamente correlacionadas. Dado  $x$  tenemos determinado completamente el valor de  $y$  y la elipse se reduce a una recta oblicua.

Suponiendo que disponemos de alguna manera de medir la covarianza o la correlación entre dos variables, podemos determinar el error de una función de ambas como sigue:

En primer lugar consideramos de nuevo la serie de Taylor en torno al valor medido de  $x$  e  $y$ , en este caso aplicando la fórmula para dos variables

$$f(x, y) \simeq f(x_0, y_0) + (x - x_0) \frac{\partial f}{\partial x} + (y - y_0) \frac{\partial f}{\partial y} \quad (1.72)$$

El valor medio de esta cantidad es simplemente

$$\langle f(x, y) \rangle \simeq f(x_0, y_0) \quad (1.73)$$

mientras que la varianza nos da

$$\langle (f(x, y) - f(x_0, y_0))^2 \rangle = \langle (x - x_0)^2 \rangle \left( \frac{\partial f}{\partial x} \right)^2 + 2 \langle (x - x_0)(y - y_0) \rangle \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial f}{\partial y} \right) + \langle (y - y_0)^2 \rangle \left( \frac{\partial f}{\partial y} \right)^2 \quad (1.74)$$

esto es

$$\sigma_z = \sqrt{\sigma_x^2 \left( \frac{\partial f}{\partial x} \right)^2 + 2\sigma_{xy} \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial f}{\partial y} \right) + \sigma_y^2 \left( \frac{\partial f}{\partial y} \right)^2} \quad (1.75)$$

Introduciendo la proporcionalidad entre los errores y la desviación cuadrática media, así como la correlación nos queda finalmente la expresión completa

$$E_z = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 E_x^2 + 2 \left( \frac{\partial f}{\partial x} \right) \left( \frac{\partial f}{\partial y} \right) E_x E_y r_{xy} + \left( \frac{\partial f}{\partial y} \right)^2 E_y^2} \quad (1.76)$$

En el caso particular de que  $x$  e  $y$  sean independientes,  $r_{xy} = 0$  y la expresión se reduce a (1.35)

$$E_z = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 E_x^2 + \left( \frac{\partial f}{\partial y} \right)^2 E_y^2} \quad (1.77)$$

La extensión es inmediata al caso de más de dos variables, añadiendo las demás derivadas y términos cruzados.

Como ilustración trivial del problema de la independencia, consideremos la función

$$z = x^2 = x \cdot x \quad (1.78)$$

Si calculamos el error aplicando la fórmula de la derivada para una variable nos queda

$$E_z = 2xE_x \quad (1.79)$$

pero si se nos ocurriera hacerlo como la derivada de un producto y aplicáramos la fórmula que presupone variables independientes resultaría

$$E_z = \sqrt{x^2 E_x^2 + x^2 E_x^2} = \sqrt{2}x E_x \quad (1.80)$$

que es una expresión incorrecta.

En este caso es evidente el fallo, porque sabemos que las variables que se multiplican son la misma, pero en muchos casos, no es nada evidente la presencia de una correlación entre dos de las variables que aparecen en una función. Más adelante veremos un caso para las rectas de mejor ajuste.

La presencia de correlaciones afecta también al cálculo del error de una magnitud medida repetidas veces, tratado en la sección 4.1, ya que la media aritmética

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots) \tag{1.81}$$

es en sí misma una función de varias variables y si éstas no son independientes, esto es, si el proceso de una de las medidas afecta al resto, ya la fórmula para el error de la media pasa a ser incorrecta.

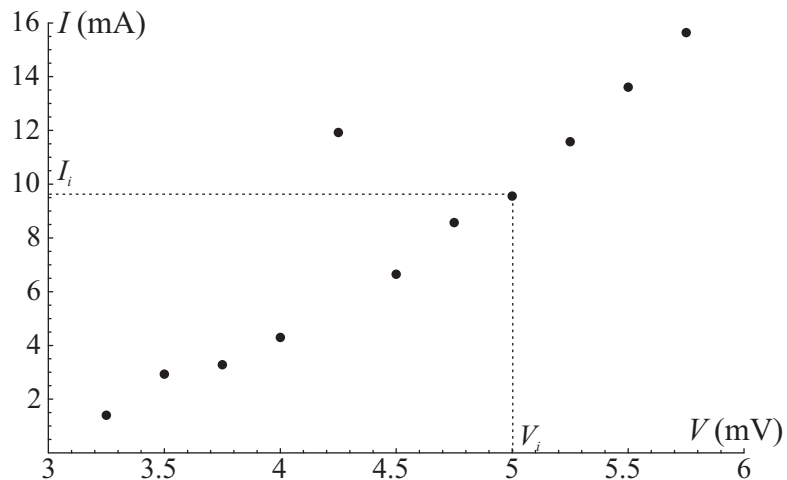
## 5 Rectas de mejor ajuste

En numerosas ocasiones no debemos determinar cuánto vale una cantidad directa o indirectamente, sino cómo depende una variable de otra.

A la hora de establecer una dependencia vamos variando una cierta magnitud  $x$  (a la que denominaremos *entrada*) en una serie de valores  $x_i$ . Para cada uno de estos valores medimos un valor  $y_i$  de una magnitud  $y$  (la *respuesta* o *salida*). El resultado puede tabularse en una tabla de dos columnas, y representarse gráficamente.

Supongamos que fijamos la tensión en los extremos de un cable y medimos la corriente que circula por él, obteniendo el siguiente resultado

$V$ (mV)	$I$ (mA)
3.25	1.40
3.50	2.93
3.75	3.28
4.00	4.30
4.25	11.92
4.50	6.65
4.75	8.57
5.00	9.56
5.25	11.58
5.50	13.61
5.75	15.64



Es evidente, de este comportamiento, que existe una relación funcional entre  $x$  e  $y$  ( $I$  y  $V$  en este caso). Ahora bien, ¿qué función representa correctamente esta conducta, una recta, una parábola, una exponencial? Y, si se trata de una parábola, ¿cuáles son los coeficientes  $a$ ,  $b$ ,  $c$  de la función  $y = ax^2 + bx + c$ ?

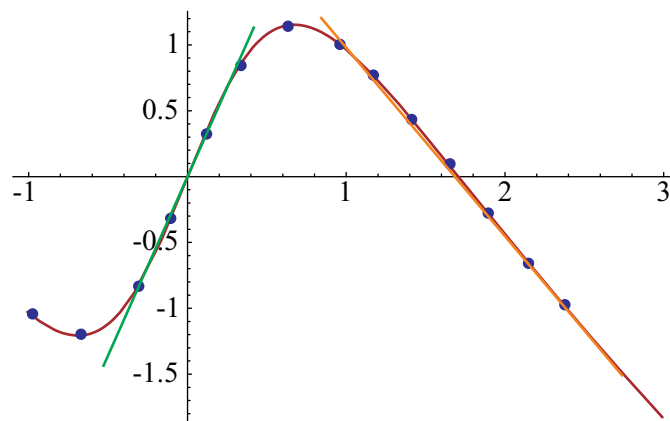
Además, sospechamos que el punto que se aparta mucho del resto es un error experimental y no debe ser tenido en cuenta. Sin embargo, ¿cómo podemos estar seguros de ello? Quizás la dependencia entre  $x$  e  $y$  sea muy compleja y el punto separado sea correcto.

Este tipo de problema se denomina *ajuste de una función* y como vemos puede ser extremadamente complicado, dependiendo del tipo de función y del número de datos y parámetros empleados para describirla.

Aquí nos limitaremos al caso más simple: el ajuste de una recta por el *método de mínimos cuadrados*. Partimos de una serie de datos más o menos alineados y trataremos de buscar la recta que pasa más cerca de ellos.

Este tipo de ajuste es de interés en varias situaciones diferentes

- Cuando tenemos una serie de entradas y salidas, para las cuales no sabemos si existe una relación funcional, pero lo sospechamos, como en la figura anterior. En este caso, la *bondad* de la recta de regresión (esto es, cuánto se aproxima a los datos) nos informa de si esta relación existe o no.
- Cuando tenemos un conjunto de datos, de los cuales sabemos con seguridad que deberían estar alineados y queremos emplear la recta de mejor ajuste para determinar alguna cantidad indirecta. Por ejemplo, supongamos que conocemos una serie de pares de valores de tensión  $V$  frente a intensidad de corriente  $I$  para un cable y queremos determinar la resistencia de acuerdo con la ley de Ohm,  $V = IR$ . En lugar de determinar una resistencia para cada dato, buscaremos la recta de mejor ajuste, cuya pendiente será la resistencia buscada.
- Las rectas de mejor ajuste también pueden usarse para hallar el valor de la salida para entradas que no hayamos medido experimentalmente. Esto es lo que se conoce como *interpolación* si  $x$  se encuentra en el mismo intervalo que los datos experimentales y *extrapolación* si está fuera de éste.
- Cuando tenemos un conjunto de datos para los cuales existe una relación funcional complicada, como la ilustrada en la figura, pero para la cual sabemos que en determinadas regiones, más o menos grandes, la conducta es aproximadamente lineal. En este caso, si nos restringimos a una de estas regiones, podemos aproximar la función por la recta de mejor ajuste, lo que simplifica los cálculos.



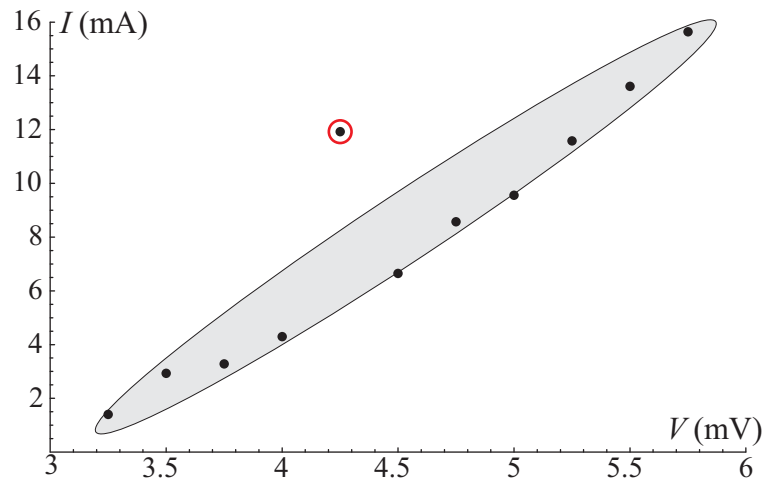
Por supuesto, en este caso la recta de mejor ajuste dependerá de la región en que nos encontremos, por lo que deberemos ser conscientes en todo momento de los límites de validez de la aproximación que empleemos. En estas aproximaciones también podemos emplear la recta de mejor ajuste para interpolar. Sin embargo, las extrapolaciones son mucho más arriesgadas, ya que podemos salirnos de los límites de validez.

## 5.1 Antes de hacer ningún cálculo: gráfica de los puntos

Antes incluso de determinar la pendiente y la ordenada en el origen de la recta de mejor ajuste, podemos establecer si los puntos están más o menos alineados.

Para ello, antes de hacer ningún cálculo, situamos los puntos en una gráfica, sea por ordenador o sobre papel milimetrado. Una simple inspección de su posición nos indicará si es conveniente ajustar una recta, si necesitaríamos una función más complicada o si no se ve ninguna relación entre las entradas y las salidas. Igualmente en esta fase podemos descartar algún punto que se aparte excesivamente del resto. En este caso, dicho punto no se incluirá en los cálculos, aunque si debe indicarse en la gráfica. Debe rodearse por un pequeño círculo, como indicación de que no ha sido considerado.

Aplicando esto a nuestra lista de datos, eliminamos el punto (1.25 mV, 11.92 mA), que en lo sucesivo no será incluido en los cálculos, salvo que indiquemos lo contrario.



La nube de puntos resultantes quedarán cubiertos por una elipse (que no hay que trazar). Los puntos estarán tanto más alineados cuanto más se aproxima esta elipse a una recta, esto es, cuanto más “estrecha” es.

### Ejes e intervalos en una gráfica

La representación de una serie de datos deberá hacerse de forma que los puntos ocupen el máximo posible de papel. Para ello se elegirán los intervalos desde ligeramente por debajo del dato menor a un poco por encima del dato mayor. Así, para la tabla anterior, intervalos posibles serían de 3 mV a 6 mV para las  $x$  y de 0 mA a 16 mA para las  $y$ . No importa que el origen de coordenadas quede fuera de la gráfica, ni que los intervalos de cada escala tengan diferente extensión.

Los ejes se dividirán en segmentos adecuados, de forma que haya 5 ó 6 en cada uno (en nuestro caso podrían ser de 0.5 mV en 0.5 mV en las  $x$  y de 2.5 mA en 2.5 mA en las  $y$ ). No obstante, debe procurarse que las escalas y sus divisiones sean “normales”, en cuanto a que puede dividirse una unidad en medios o en cuartos, pero no es normal dividirla en séptimos, por ejemplo.

En los extremos de los ejes se indicarán las magnitudes y las unidades en las que mide entre paréntesis, ( $I$  (A) y  $V$  (mV) en este caso).

**No** hay que escribir las coordenadas de los datos, ni en los ejes ni cerca de cada punto. Esto es redundante y dificulta la observación de la gráfica.

### Ubicación

Las gráficas que se hagan, sean exclusivamente de los datos experimentales, de rectas de mínimos cuadrados, u otras solicitadas explícitamente, deben adjuntarse al final de la memoria, estando convenientemente etiquetadas.

Toda gráfica deberá ir numerada y con un título explicativo. Si hay varias curvas o rectas trazadas en la misma hoja se indicará cuál es cada una.

## 5.2 Coeficiente de correlación

Como medida matemática del grado de alineación (o *correlación lineal*), la magnitud relevante es el llamado *coeficiente de correlación lineal*,  $r$ , definido como

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{\sum_i X_i Y_i}{\sqrt{(\sum_i X_i^2) (\sum_i Y_i^2)}} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \quad (1.82)$$

Esta definición es análoga a la medida de la correlación entre dos variables, definida en la sección 4.3, con la única diferencia de que aquella se suponía calculada conocida la distribución completa (esto es, infinitos datos), mientras que ésta se limita a los datos de que disponemos.

A partir de esta definición puede demostrarse que el coeficiente de correlación es una cantidad comprendida en el intervalo  $(-1, 1)$ , siendo su signo el mismo de la pendiente de la recta de mejor ajuste (positivo si la recta va “hacia arriba” y negativo en caso contrario).

Podemos caracterizar la alineación dividiendo los valores de  $r$  en intervalos:

$|r| = +1$ : Los puntos se encuentran perfectamente alineados.

$0.75 < |r| < 1$ : Los puntos muestran una clara alineación, que será mejor cuanto más se acerque  $r$  a la unidad.

$0.25 < |r| < 0.75$ : Existe una cierta tendencia lineal entre los puntos, que disminuye al reducirse  $r$ .

$0 < |r| < 0.25$ : No puede decirse que exista una relación lineal entre  $x$  e  $y$ , ya que los puntos forman una nube casi informe.

$|r| = 0$ : no existe correlación entre la entrada y la salida, esto es, se trata de variables independientes entre sí.

**Ejemplo:** En el caso que estamos considerando obtenemos

$$S_{XX} = 6.80625 \text{ mV}^2 \quad S_{XY} = 37.5395 \text{ mV} \cdot \text{mA} \quad S_{YY} = 211.857 \text{ mA}^2$$

$$r = \frac{37.5395 \text{ mV} \cdot \text{mA}}{\sqrt{6.80625 \text{ mV}^2 \cdot 211.857 \text{ mA}^2}} = 0.988582$$

Vemos que resulta una cantidad adimensional próxima a la unidad, lo que indica el hecho ya conocido a partir de la gráfica de que los datos están muy alineados.

Las calculadoras que incluyen cálculos de regresión o las hojas de cálculo, como *Excel*, incorporan entre sus funciones el coeficiente de correlación, con lo que basta introducir los datos y aplicar dicha función.

Caso de que esto no sea posible, puede recurrirse a la definición indicada antes, o a la expresión matemáticamente equivalente

$$r = \frac{SS_{xy} - S_x S_y}{\sqrt{(SS_{xx} - S_x^2)(SS_{yy} - S_y^2)}} \quad (1.83)$$



donde

$$S = \sum_i 1 = n \quad S_x = \sum_i x_i \quad S_{xx} = \sum_i x_i^2 \quad (1.84)$$

$$S_y = \sum_i y_i \quad S_{xy} = \sum_i x_i y_i \quad S_{yy} = \sum_i y_i^2 \quad (1.85)$$

Aunque teóricamente esta expresión da el mismo resultado que la definición, es posible que al calcularla en un ordenador o calculadora, se obtenga un valor diferente, debido a la presencia de errores de redondeo, como los comentados en la sección 3.2. Por ello, si esta operación se hace con una hoja de cálculo, es preferible emplear la definición, con los valores medidos respecto a la media, tal como aparece en la ecuación (1.82), que presenta un menor error.

El coeficiente de correlación es una propiedad de la distribución de la nube de puntos y no es preciso hallar el error de esta cantidad (que se debería a la incertidumbre de cada dato). Al no disponer de un error para truncar las cifras significativas, adoptaremos el criterio de retener todos los 9's consecutivos que haya tras el punto decimal, más la primera cifra distinta de 9 que les siga, sin redondear esta última cifra. Por ejemplo, si  $r = 0.999816$  lo redondeamos a  $r = 0.9998$ , si  $r = 0.994923$  lo dejamos en  $r = 0.994$  y si  $r = 0.989$  como  $r = 0.98$ . La razón de este extraño convenio es que estamos más interesados en ver cuánto se aproxima  $r$  a la unidad, como medida de buena alineación, por lo que la información más significativa es la cantidad de 9's que hay tras el punto decimal. En nuestro caso sería  $r = 0.98$ .

A partir de  $r$  podemos definir una variable angular como

$$r = \cos \phi \quad \phi = \arccos r \quad (1.86)$$

que no tiene interpretación geométrica, pero de la que podemos decir que estará más próxima a 0 (o a  $\pi$ ) cuanto más alineados estén los puntos (en nuestro caso sería  $\phi = 0.151$ ).

Un coeficiente de correlación bajo para una lista de puntos en la que sabemos con seguridad que debería existir una relación lineal sugiere que existe algún dato aislado que se separa del resto. La gráfica de estos puntos, como indicamos en la sección anterior, puede decirnos si debemos descartar algún dato, lo que suele redundar en la mejora del valor de  $r$  (esto es, que se acercará más a la unidad). Así, en nuestro ejemplo, si hubiéramos retenido el dato que hemos descartado, habría resultado un valor de  $r = 0.922556$ , mientras que sin él resulta  $r = 0.988582$ , como ya hemos visto.

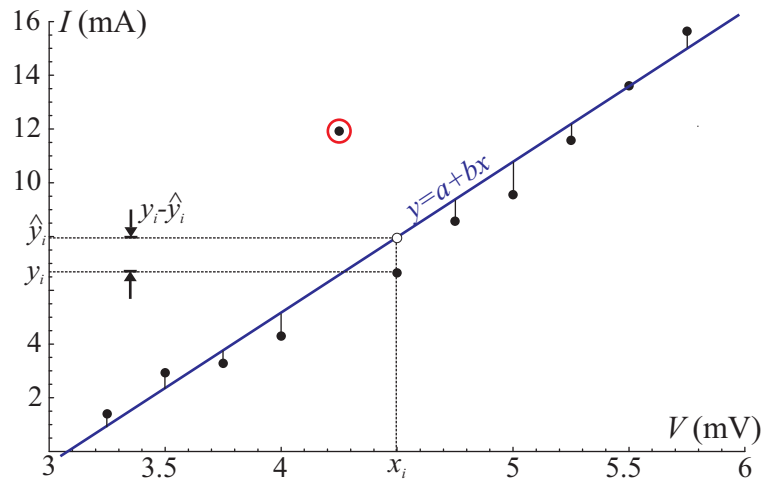
### 5.3 Pendiente y ordenada en el origen

Una vez que conocemos el coeficiente de correlación y sabemos que está próximo a la unidad en valor absoluto, esto es, que los puntos están prácticamente alineados, determinaremos los parámetros de la recta de mejor ajuste o de regresión.

Para ello, empleamos el método de *mínimos cuadrados*, ya comentado en el cálculo de la media (sección 4.1). La idea de este método es sencilla: Dados los puntos en la gráfica, suponemos una recta que pase entre ellos, con ecuación

$$y = a + bx \quad (1.87)$$

con  $a$  la ordenada en el origen y  $b$  la pendiente.



Para cada valor de la entrada,  $x_i$ , esta recta nos dará una predicción

$$\hat{y}_i = a + bx_i \tag{1.88}$$

La diferencia entre el valor real de la salida,  $y_i$ , y esta predicción será

$$y_i - \hat{y}_i = y_i - (a + bx_i) \tag{1.89}$$

La idea es minimizar el conjunto de estas diferencias mediante una sola recta. Para ello construimos la función

$$\chi^2 = \sum_i (y_i - (a + bx_i))^2 \tag{1.90}$$

El cuadrado del exponente se incluye para garantizar que tenemos una suma de valores positivos y no se cancelan unos términos a otros. Esta función depende de dos variables  $a$  y  $b$ . Aplicando el cálculo del mínimo de una función de varias variables se llega a que los valores óptimos de estos parámetros son

$$b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{XY}}{S_{XX}} \quad a = \bar{y} - b\bar{x} \tag{1.91}$$

El segundo resultado implica que la recta puede escribirse como

$$y - \bar{y} = b(x - \bar{x}) \quad Y = bX \tag{1.92}$$

esto es, que la recta de mejor ajuste siempre pasa por el punto  $(\bar{x}, \bar{y})$  definido por la media de las entradas y la de las respuestas.

**Ejemplo:** En nuestro caso podemos aprovechar algunos de los valores que calculamos para hallar el coeficiente de correlación

$$\bar{x} = \frac{S_x}{S} = \frac{45.25 \text{ mV}}{10} = 4.525 \text{ mV} \quad \bar{y} = \frac{S_y}{S} = \frac{77.52 \text{ mA}}{10} = 7.752 \text{ mA}$$

$$b = \frac{S_{XY}}{S_{XX}} = \frac{37.5395 \text{ mA} \cdot \text{mV}}{6.80625 \text{ mV}^2} = 5.51545 \text{ S} \quad a = \bar{y} - b\bar{x} = -17.20539027 \text{ mA}$$

(la S que aparece en la expresión de  $b$  es la abreviatura de *Siemens*, la unidad de conductancia eléctrica). Obsérvese que  $a$  tiene las mismas unidades que  $y$ , mientras que  $b$  tiene unidades de  $y/x$ .

Las funciones para calcular la pendiente  $b$  y la ordenada en el origen suelen estar directamente disponibles en las calculadoras y hojas de cálculo (aunque debe tenerse en cuenta qué notación se usa, ya que dependiendo del modelo a  $a$  se le puede llamar  $b$  y viceversa).

Si estas funciones no están disponibles pueden calcularse a partir de las fórmulas anteriores, o bien empleando la expresión alternativa

$$b = \frac{SS_{xy} - S_x S_y}{SS_{xx} - S_x^2} \quad a = \bar{y} - b\bar{x} = \frac{S_y S_{xx} - S_x S_{xy}}{SS_{xx} - S_x^2} \quad (1.93)$$

Como en casos anteriores, esta expresión alternativa está sujeta a errores de redondeo, por lo que si es posible hay que aplicar las expresiones relativas a las medias.

Tanto la pendiente como la ordenada se obtienen a partir de una cierta distribución estadística de puntos alrededor de la recta de mejor ajuste y como tales están sujetas a error. El error de estas cantidades puede hallarse aplicando métodos estadísticos y resultan los siguientes:

$$E_b = \frac{2b}{r} \sqrt{\frac{1-r^2}{n-2}} \quad E_a = E_b \sqrt{\sigma_x^2 + \bar{x}^2} = E_b \sqrt{\frac{S_{xx}}{S}} \quad (1.94)$$

En estas expresiones  $b$  es la pendiente y  $E_b$  su error (que se necesita para calcular el error de la ordenada),  $\sigma_x$  es la llamada desviación estándar de la población para la variable  $x$ , definida en (1.13)

$$\sigma_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}} = \sqrt{\frac{S_{XX}}{S}} \quad (1.95)$$

Suele aparecer en las calculadoras como  $x\sigma_n$  (y no hay que confundirla con  $x\sigma_{n-1}$ ).

**Ejemplo:** Siguiendo en nuestro caso tenemos

$$E_b = \frac{2b}{r} \sqrt{\frac{1-r^2}{n-2}} = 0.5944 \text{ S}$$

$$E_a = E_b \sqrt{\sigma_x^2 + \bar{x}^2} = E_b \sqrt{\frac{S_{XX}}{S} + \bar{x}^2} = 2.734238701 \text{ mA}$$

Obsérvese como, al tratarse de errores absolutos, tienen las mismas unidades que las magnitudes correspondientes.

Una vez calculados los errores, podemos expresar la pendiente y la ordenada como

$$b = 5.5 \pm 0.6 \text{ S} \quad a = -17 \pm 3 \text{ mA}$$

y la ecuación de la recta de mejor ajuste será

$$I = -17(3) + 5.5(6)x \quad (\text{mA})$$

La fórmula para los errores de  $a$  y  $b$  no suelen aparecer explícitamente en las calculadoras y hojas de cálculo, aunque son fáciles de implementar. La fórmula para  $b$  puede presentar errores de redondeo, que se pueden evitar empleando la expresión alternativa equivalente (pero más difícil de implementar):

$$E_b = 2\sqrt{\frac{\sum_i (y_i - \bar{y} - b(x_i - \bar{x}))^2}{(n-2)\sum_i (x_i - \bar{x})^2}} = 2\sqrt{\frac{\sum_i (Y_i - bX_i)^2}{(n-2)S_{XX}}} \quad (1.96)$$

Una forma alternativa de hallar  $E_b$  consiste en emplear la variable angular  $\phi$  definida en (1.86)

$$r = \cos \phi \quad \Rightarrow \quad E_b = \frac{2b \operatorname{tg} \phi}{\sqrt{n-2}} \quad (1.97)$$

#### 5.4 Otras rectas de regresión

Una vez que se ha definido y calculado la recta de regresión básica, es inmediato generalizarla a otras situaciones.

En muchos casos, la dependencia funcional entre la entrada y la salida no es lineal pero, con alguna transformación matemática en los datos puede reducirse a un ajuste lineal. Veamos algunos ejemplos:

##### Dependencia exponencial

Supongamos que la relación funcional supuesta es de la forma

$$y = Ae^{\lambda x} \quad (1.98)$$

En este caso, tomando logaritmos en los dos miembros queda

$$\ln(y) = \ln(A) + \lambda x \quad (1.99)$$

que podemos reescribir como

$$z = a + bx \quad z = \ln(y) \quad a = \ln(A) \quad b = \lambda \quad (1.100)$$

esto es, que si a partir de las salidas  $y_i$  hallamos sus respectivos logaritmos,  $z_i$  y calculamos la recta de mejor ajuste de  $z_i$  frente a los  $x_i$ , el resultado es una recta que nos da los parámetros  $A$  y  $\lambda$ . El factor del exponente coincide con la pendiente de la recta, mientras que el prefactor  $A$  vale

$$A = e^a \quad (1.101)$$

Los errores de  $\lambda$  y  $A$  serán

$$E_\lambda = E_b \quad E_A = \left| \frac{dA}{da} \right| E_a = AE_a \quad (1.102)$$

Para obtener gráficamente una recta en este caso deberemos representar  $\ln(y)$  frente a  $x$ .

**Dependencia potencial**

En ocasiones debemos suponer que entre la entrada y la salida existe una relación *potencial* de la forma

$$y = kx^\alpha \tag{1.103}$$

y el objetivo es determinar el exponente  $\alpha$  y el prefactor  $k$ . Esto lo conseguimos hallando el logaritmo de cada miembro

$$\ln(y) = \ln(k) + \alpha \ln(x) \tag{1.104}$$

que se puede escribir como

$$z = a + bu \tag{1.105}$$

$$z = \ln(y) \quad u = \ln(x) \quad a = \ln(k) \quad b = \alpha \tag{1.106}$$

es decir, que si se representa gráficamente  $\ln(y)$  frente a  $\ln(x)$  (lo que se denomina una *gráfica log-log*) deberemos obtener un comportamiento aproximadamente lineal.

Los errores del prefactor y del exponente serán

$$E_\alpha = E_b \quad E_k = kE_a \tag{1.107}$$

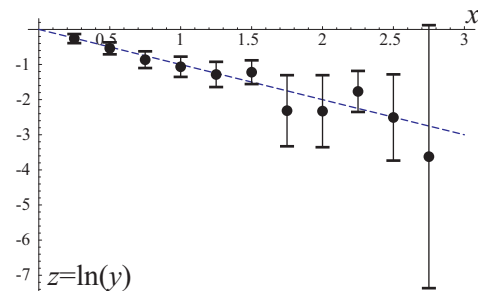
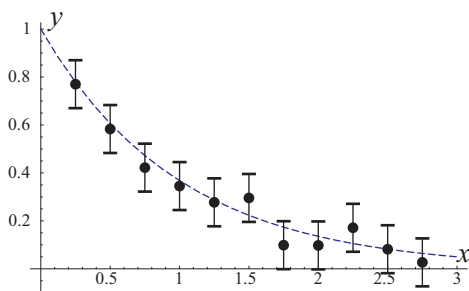
**(\*) Observación sobre los cálculos de la recta logarítmica y potencial**

Las dos rectas anteriores presentan un problema que debe señalarse. A la hora de medir la magnitud  $y$ , suele emplearse un aparato cuya precisión es la misma para todos los datos, esto es,  $E_{y_i}$  es una constante.

Sin embargo, cuando se calcula la recta, no se emplea la variable  $y$ , sino  $z = \ln(y)$ , cuyo error depende de  $y$

$$E_{z_i} = \left| \frac{dz}{dy} \right| E_{y_i} = \frac{E_{y_i}}{y_i}$$

esto es, cuanto más pequeño sea  $y_i$ , mayor es el error de  $z_i$ . Esto es especialmente delicado en comportamientos exponenciales que tienden a cero (que es un caso que aparece a menudo, por ejemplo, en las curvas de carga o descarga de un circuito RC, o en un oscilador amortiguado). En este caso los valores finales de  $z_i$  poseen un error mayor que los iniciales y deberían pesar menos en los cálculos.



La solución consiste, como en el caso de una sola variable, en pesar los distintos datos de acuerdo con su error. Se trata entonces de minimizar la función

$$\chi^2 = \sum_i \frac{(z_i - (a + bx_i))^2}{E_{z_i}^2} = \frac{1}{E_y^2} \sum_i y_i^2 (z_i - (a + bx_i))^2$$

La constante  $E_y$  no afecta al cálculo del mínimo de esta función.

Operando se obtienen fórmulas análogas al caso ya estudiado

$$b = \frac{S_{XZ}}{S_{XX}} \quad a = \bar{z} - b\bar{x}$$

donde ahora los sumatorios incluyen los pesos

$$S = \sum_i y_i^2 \quad S_x = \sum y_i^2 x_i \quad S_z = \sum y_i^2 z_i$$

$$\bar{x} = \frac{S_x}{S} \quad \bar{z} = \frac{S_z}{S}$$

$$S_{XX} = \sum_i y_i^2 (x_i - \bar{x})^2 \quad S_{XZ} = \sum_i y_i^2 (z_i - \bar{z})^2$$

Las fórmulas para el error de  $a$  y de  $b$  son exactamente las mismas que en (1.94), simplemente teniendo en cuenta que tanto  $r$  como  $\sigma_x$  también deben calcularse con los sumatorios ponderados.

Dado que las calculadoras, a menos que se programen específicamente, no hallan estas pendientes y ordenadas corregidas, daremos como válido el resultado aproximado mediante las fórmulas habituales (sin tener en cuenta los errores individuales). Sin embargo, en una hoja de cálculo es posible implementar las funciones correctas, que proporcionarán un valor más ajustado. En la página *web* de la asignatura de Campos Electromagnéticos hay herramientas ya programadas para realizar estos cálculos.

### Otros comportamientos linealizables

Técnicas parecidas a las de las rectas anteriores pueden aplicarse a otras situaciones en que mediante alguna transformación la relación funcional se reduzca a una lineal.

Por ejemplo, si se mide la corriente en un circuito alimentado por una fuente de continua de f.e.m  $\mathcal{E}$  y resistencia interna  $R_0$ , para diferentes valores de una resistencia externa, se tiene la relación

$$I = \frac{\mathcal{E}}{R + R_0} \quad (1.108)$$

se trata de hallar  $\mathcal{E}$  y  $R_0$ . Conseguimos esto invirtiendo cada miembro

$$\frac{1}{I} = \frac{R_0 + R}{\mathcal{E}} = \frac{R_0}{\mathcal{E}} + \frac{1}{\mathcal{E}}R \quad (1.109)$$

de forma que si se representa  $1/I$  frente a  $R$  se obtiene una recta cuya pendiente y ordenada podemos identificar con

$$a = \frac{R_0}{\mathcal{E}} \quad b = \frac{1}{\mathcal{E}} \quad \Rightarrow \quad \mathcal{E} = \frac{1}{b} \quad R_0 = \frac{a}{b} \quad (1.110)$$

### (\*) Dependencia entre la pendiente y la ordenada

A la hora de realizar un cálculo que implique simultáneamente a  $a$  y a  $b$ , como en el valor de  $R_0$  del ejemplo anterior, hay que tener en cuenta que estas dos variables no son completamente independientes entre sí.

La razón es evidente. Sabiendo que la recta de mejor ajuste pasa por el punto medio de la distribución de datos, un cambio en la pendiente hace que el punto de corte con el eje de ordenadas se desplace hacia arriba o hacia abajo. La correlación no es absoluta, porque la posición del punto medio tiene su

propia incertidumbre y la recta “real” podría estar más arriba o abajo, lo que modificaría la ordenada independientemente de la pendiente.

La correlación entre  $a$  y  $b$  puede determinarse también por métodos estadísticos y resulta ser

$$r_{ab} = -\frac{S_x}{\sqrt{SS_{xx}}} \quad (1.111)$$

de forma que el error de una función que depende de  $a$  y  $b$  es de forma que el error de una función que depende de  $a$  y  $b$  es

$$\begin{aligned} E_f &= \sqrt{E_a^2 \left(\frac{\partial f}{\partial a}\right)^2 + 2r_{ab}E_aE_b \left(\frac{\partial f}{\partial a}\right) \left(\frac{\partial f}{\partial b}\right) + E_b^2 \left(\frac{\partial f}{\partial b}\right)^2} = \\ &= E_b \sqrt{\sigma_x^2 \left(\frac{\partial f}{\partial a}\right)^2 + \left(\bar{x} \frac{\partial f}{\partial a} - \frac{\partial f}{\partial b}\right)^2} \end{aligned} \quad (1.112)$$

Esta fórmula contiene como caso particular el error de la propia ordenada en el origen  $a$ , dada en (1.94) y el error de la extrapolación,  $\hat{y}$  (que veremos en (1.114)).

### Cálculo de la recta inversa

Un problema común consiste en, dado un conjunto de datos  $(x_i, y_i)$  y la recta de mejor ajuste  $y = a + bx$ , hallar la recta inversa, esto es, despejar la  $x$  y obtener una nueva recta

$$x = my + n$$

Despejando de la recta original obtenemos

$$x = \frac{y - a}{b} \Rightarrow m = \frac{1}{b} \quad n = -\frac{a}{b}$$

El cálculo es inmediato. El único problema está en la determinación del error de  $m$  y  $n$ . Para la pendiente aplicamos la fórmula del error de una función de una sola variable

$$E_m = \left| \frac{dm}{db} \right| E_b = \frac{E_b}{b^2}$$

Para el error de la ordenada debemos tener cuidado, ya que  $n$  es función de  $a$  y  $b$  que, según acabamos de ver, no son variables independientes, por lo que no podemos aplicar directamente la fórmula del error de un cociente. En su lugar tenemos

$$E_n = E_b \sqrt{\sigma_x^2 \left(\frac{\partial n}{\partial a}\right)^2 + \left(\bar{x} \frac{\partial n}{\partial a} - \frac{\partial n}{\partial b}\right)^2} = \frac{E_b}{b^2} \sqrt{b^2 \sigma_x^2 + \bar{y}^2}$$

Hay que señalar que, cuando se calcula la recta inversa por este método *no* se obtiene el mismo resultado que si se halla a partir de los puntos  $(y_i, x_i)$ . No obstante, si los datos están muy alineados, los dos resultados se parecen bastante.

Así, para el ejemplo que estamos considerando, a partir de la inversión de los datos  $(y_i, x_i)$  se obtiene

$$m = 0.177 \pm 0.020 \Omega \quad n = 0.15 \pm 0.17 \text{ mV}$$

mientras que despejando de la recta  $y = a + bx$  resulta

$$m = 0.181 \pm 0.020 \Omega \quad n = 0.11 \pm 0.17 \text{ mV}$$

Para la mayoría de las situaciones prácticas, el segundo método es perfectamente aceptable y evita el tener que calcular una nueva recta de mejor ajuste.

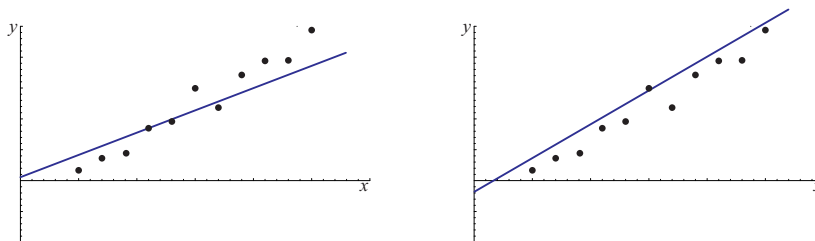
## 5.5 Trazado de la recta

Una vez que se ha calculado la pendiente y la ordenada en el origen, debe procederse al trazado de la recta correspondiente. Debe recordarse que **todo cálculo de regresión de mínimos cuadrados implica el trazado de la gráfica correspondiente**, salvo que se indique específicamente lo contrario.

Esta recta se trazará sobre la gráfica que ya hemos efectuado con los datos experimentales (que **siempre deben aparecer**) y para la que suponemos que ya hemos elegido los ejes, intervalos y unidades adecuados.

La posible recta de mínimos cuadrados se indicará junto con los puntos y en un color diferente, de forma que puedan apreciarse ambos. La forma de trazar la recta es tomar la ecuación  $\hat{y} = a + bx$  sustituir  $x$  por dos valores que estén dentro del intervalo de representación (en el ejemplo que dimos antes, entre 0 y 6 A), hallar la salida  $\hat{y}$  y trazar la recta que pasa por los dos puntos. Al final se borrarán los dos puntos, pues no son datos experimentales.

A menudo es fácil apreciar cuando una recta de mínimos cuadrados está mal. Por ejemplo, cuando todos los puntos están al mismo lado de la recta (ya que entonces un cambio en la ordenada en el origen daría una recta de mejor ajuste), o cuando todos los puntos del principio están a un lado y los del final a otro (ya que entonces se mejoraría cambiando la pendiente). Así, estas dos rectas serían incorrectas:



En ocasiones podrán trazarse varias curvas o rectas en la misma gráfica. Esto se podrá hacer siempre que rangos de los datos de las diferentes curvas sean similares. En este caso deberán emplearse colores o signos diferentes para las distintas rectas y para los puntos correspondientes a cada una.

## 5.6 Interpolación y extrapolación

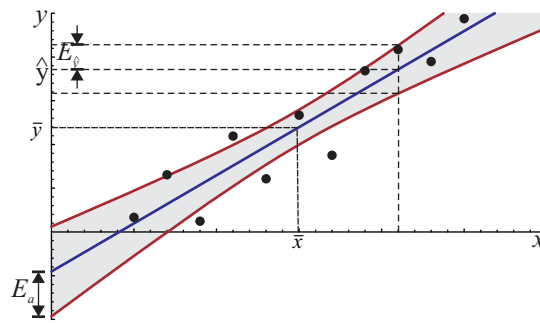
Una vez que hemos determinado la pendiente y la ordenada de la recta de mejor ajuste, podemos emplear estos valores para calcular valores de la salida para entradas que no hayamos medido, esto es, si  $x_0$  es una entrada, podemos estimar la salida correspondiente como

$$\hat{y} = a + bx_0 \quad (1.113)$$

Esta estimación será una *interpolación* si  $x_0$  está dentro del intervalo del resto de medidas y una *extrapolación* si está fuera de él.

Las interpolaciones y extrapolaciones se encuentran sometidas a error por dos razones relacionadas: primero, porque existe una desviación estadística alrededor de la recta de mejor ajuste, ya que los puntos nunca estarán perfectamente alineados; segundo, porque al existir un error en la pendiente, una modificación en su valor hace que la recta “apunte” en la dirección incorrecta.





Si representamos las curvas que nos dan el 95% de probabilidad de que la extrapolación sea correcta obtendremos una gráfica como la de la figura. El error será mínimo en la posición del punto medio (ya que todas las posibles rectas de mejor ajuste pasan por este punto, solo tendremos el error estadístico) y aumenta a medida que nos alejamos de él. Una extrapolación será tanto más incierta cuanto más lejos nos vayamos.

La expresión para el error de una interpolación o extrapolación viene dado por la expresión

$$E_{\hat{y}} = E_b \sqrt{\sigma_x^2 + (\bar{x} - x_0)^2} \tag{1.114}$$

Nótese que la fórmula para el error en la ordenada (1.94) es un caso particular de ésta, haciendo  $x_0 = 0$ .

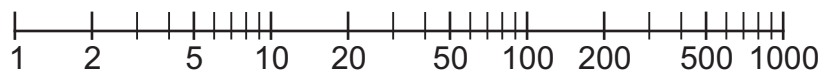
### 5.7 Representación logarítmica y semilogarítmica

Cuando se tienen datos de los que se sabe, o se sospecha, que poseen una conducta exponencial o potencial, interesa usar como eje el logaritmo de una o de las dos cantidades. Sin embargo, al indicar en los ejes dichos logaritmos, las gráficas son más difíciles de interpretar. Es mucho más fácil entender una gráfica en la que los puntos corresponden a “2” y a “3” que una en que corresponden a “0.301” y “0.477” (los logaritmos decimales de 2 y 3).

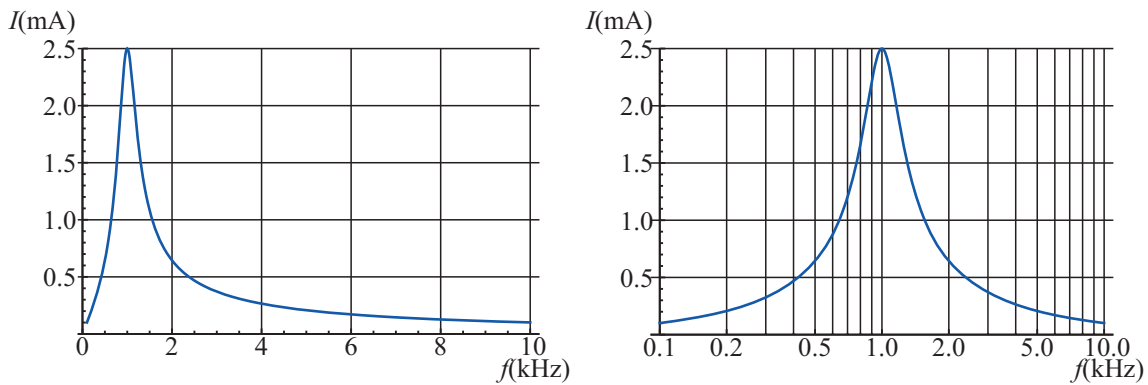
Nos interesa entonces una representación que, aun estando las marcas espaciadas según los logaritmos de 1, 2, 3, ..., las etiquetas corresponden a “1”, “2”, “3”, ... de forma que sabemos a qué valor original corresponde cada logaritmo.

Para construir esta escala logarítmica se emplea usualmente la base 10. Se sitúa la marca de “1” en el origen (pues su logaritmo es 0) y “10” a una distancia unitaria (por ejemplo, 1 cm). Los valores correspondientes a “2”, “3”, etc., se situarán a 0.301 cm, 0.477 cm, etc. del origen.

Esto produce una escala no lineal, en la que las marcas se van acumulando. Así, la distancia entre 100 y 10 es la misma que entre 10 y 1, y la marca del 20 dista del 10, lo mismo que la del 2 de la del 1.



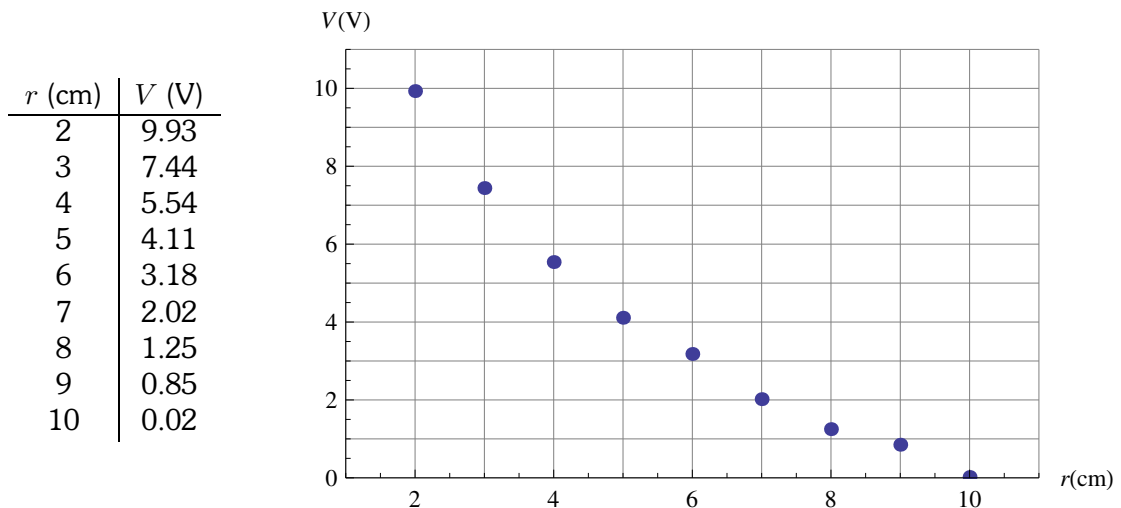
El uso de estas escalas es especialmente útil cuando se tienen un rango de datos muy amplio, como por ejemplo, al hacer un barrido en frecuencias. Empleando una escala logarítmica se le da la misma importancia a las bajas frecuencias que a las altas. Por ejemplo, para la respuesta un circuito RLC, la representación en una escala logarítmica muestra la simetría del comportamiento para altas y bajas frecuencias:



Combinando los diferentes tipos de escalas, podemos tener gráficas semilogarítmicas, cuando una de las dos escalas es logarítmica y la otra lineal (útil para comportamientos exponenciales y logarítmicos), y logarítmicas (o log-log), cuando las dos escalas son logarítmicas (apropiado para comportamientos potenciales).

No es necesario diseñar manualmente este tipo de escalas: los programas de representación gráfica, como Excel, permiten seleccionar escalas logarítmicas. También existen papeles milimetrados con una o dos escalas logarítmicas.

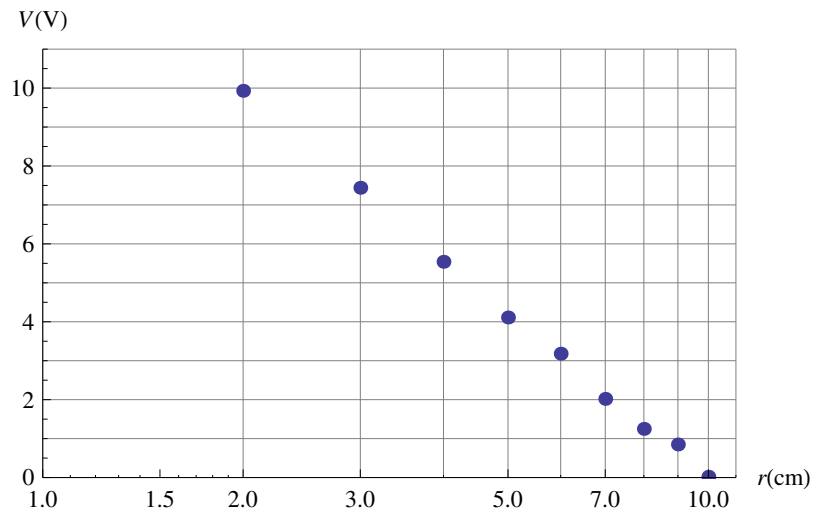
Por ejemplo, supongamos que tenemos los datos



Si representamos el voltaje respecto a la distancia, obtenemos una gráfica como la de la figura. En esta apreciamos un decrecimiento con  $r$  pero que no es totalmente lineal. Esa curvatura podría obedecer a una exponencial, una parábola, un logaritmo,...

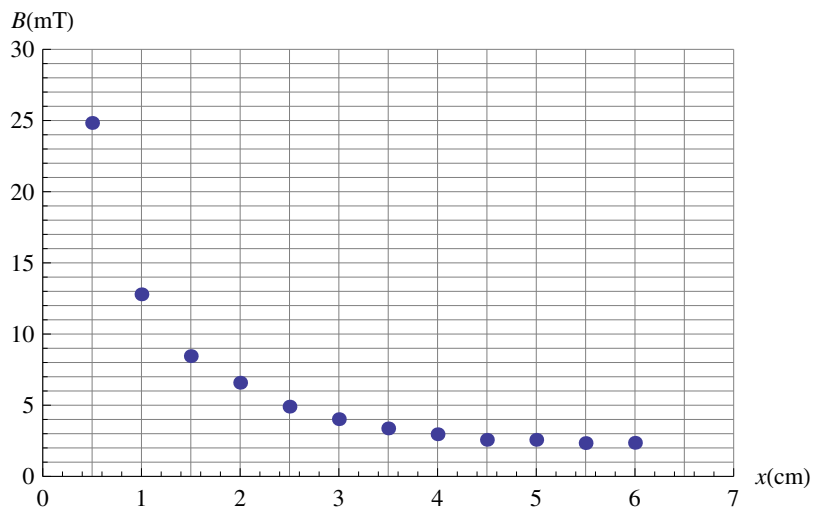
Si la tabla anterior la representamos en una gráfica semilogarítmica, en la que una de las dos escalas (la de la distancia) es logarítmica, mientras que la del voltaje es lineal, el resultado es claramente rectilíneo, por lo que debemos suponer una dependencia de la forma  $V = a + b \ln(r)$

$r$ (cm)	$V$ (V)
2	9.93
3	7.44
4	5.54
5	4.11
6	3.18
7	2.02
8	1.25
9	0.85
10	0.02



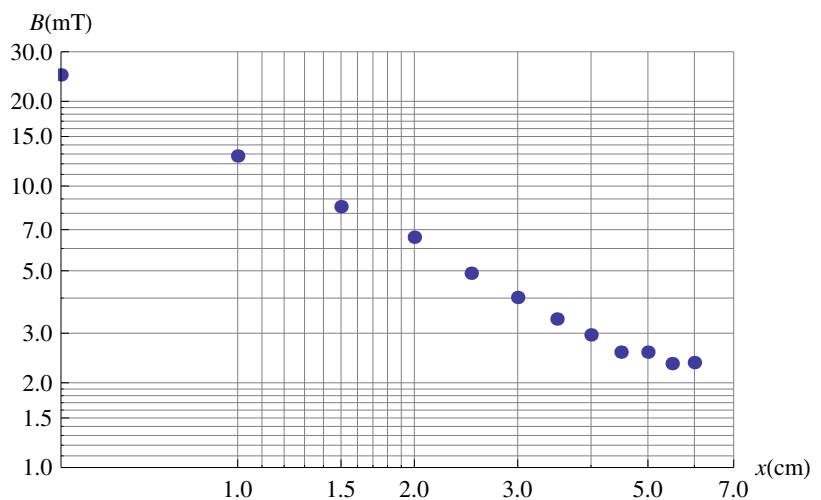
Como otro ejemplo, supongamos la siguiente tabla de datos

$x$ (cm)	$B$ (mT)
0.5	24.83
1.0	12.79
1.5	8.45
2.0	6.58
2.5	4.90
3.0	4.02
3.5	3.37
4.0	2.96
4.5	2.57
5.0	2.57
5.5	2.34
6.0	2.36



En una representación con escalas lineales, la gráfica podría ser un decaimiento exponencial, o una hipérbola, u otros comportamientos.

$x$ (cm)	$B$ (mT)
0.5	24.83
1.0	12.79
1.5	8.45
2.0	6.58
2.5	4.90
3.0	4.02
3.5	3.37
4.0	2.96
4.5	2.57
5.0	2.57
5.5	2.34
6.0	2.36



Si hacemos la representación en una gráfica log-log, vemos la ley teórica es una potencial, pues en esta

gráfica muestra un comportamiento lineal. Es más, midiendo la pendiente en esta representación resulta un valor próximo a -1, lo que indica que sigue una ley  $B = K/x$ .

## 6 (\*) Generalización del ajuste

El método de mínimos cuadrados puede extenderse a otras funciones que no sean lineales, o a casos de más variables.

Por ejemplo, supongamos que tenemos una colección de salidas  $z_i$ , para un conjunto de datos  $(x_i, y_i)$  y deseamos establecer la dependencia  $z(x, y)$  mediante una cierta función  $f(x, y, \dots, a, b, \dots)$ , siendo  $a, b, \dots$  los parámetros que deseamos determinar. El método para hacerlo consistirá simplemente en minimizar la cantidad

$$\chi^2 = \sum_i (z_i - f(x_i, y_i, \dots, a, b, \dots))^2 \quad (1.115)$$

lo que nos lleva al sistema de ecuaciones

$$\frac{\partial \chi^2}{\partial a} = 0 \quad \frac{\partial \chi^2}{\partial b} = 0 \quad \dots \quad (1.116)$$

En general resultará un sistema imposible de resolver, ya que implica ecuaciones no lineales.

La excepción se da cuando  $f(x, y, \dots, a, b, \dots)$  es una combinación lineal de ciertas funciones base, esto es

$$f(x, y, \dots, a, b, \dots) = a\phi_1(x, y, \dots) + b\phi_2(x, y, \dots) + \dots \quad (1.117)$$

En este caso el sistema de ecuaciones resultante es uno lineal, que puede resolverse, por ejemplo, matricialmente.

Como caso particular, supongamos la relación funcional

$$\hat{z} = a + bx + cy \quad (1.118)$$

que nos da el *plano de mejor ajuste* de  $z$  frente a  $x$  e  $y$ . Se trata de hallar los valores de  $a$ ,  $b$  y  $c$  que hacen mínima la cantidad

$$\chi^2 = \sum_i (z_i - (a + bx_i + cy_i))^2 \quad (1.119)$$

Tras derivar, resulta un sistema de tres ecuaciones con tres incógnitas que, escrito matricialmente, es

$$\mathbf{M} \cdot \mathbf{b} = \mathbf{z} \quad (1.120)$$

$$\mathbf{M} = \begin{pmatrix} S & S_x & S_y \\ S_x & S_{xx} & S_{xy} \\ S_y & S_{xy} & S_{yy} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} S_z \\ S_{xz} \\ S_{yz} \end{pmatrix} \quad (1.121)$$

La solución de este sistema será

$$\mathbf{b} = \mathbf{M}^{-1} \cdot \mathbf{z} \quad (1.122)$$

El cálculo de los errores requiere igualmente el cálculo matricial. El resultado es, en este caso

$$\mathbf{E} = 2\mathbf{M}^{-1} \sqrt{\frac{\sum_i (z_i - (a + bx_i + cy_i))^2}{n - 3}} \quad (1.123)$$

donde  $\mathbf{E}$  es una matriz cuadrada que contiene los errores de cada parámetro y las correlaciones respectivas, esto es

$$\mathbf{E} = \begin{pmatrix} E_a^2 & E_a E_b r_{ab} & E_a E_c r_{ac} \\ E_a E_b r_{ab} & E_b^2 & E_b E_c r_{bc} \\ E_a E_c r_{ac} & E_b E_c r_{bc} & E_c^2 \end{pmatrix} \quad (1.124)$$

Una vez ajustada la función, podremos emplearla para hallar interpolaciones y extrapolaciones, así como calcular el error de éstas empleando métodos estadísticos.

## 7 Sobre la presentación de las unidades y las tablas

Todos los datos que aparezcan en una práctica, sean medidas efectuadas, cálculos intermedios o resultados finales **deberán ir acompañados de sus unidades** (salvo que sean adimensionales). Asimismo habrá que indicar la precisión de los aparatos con que fueron medidos y la escala en la que se estaba trabajando..

La regla general es que dado un dato, su error absoluto y su unidad, deberá expresarse como

$$M \pm E U \quad (1.125)$$

**Ejemplo:** Para una longitud  $l$  de valor 16 mm con un error de 1 mm será

$$l = 16 \pm 1 \text{ mm} \quad (1.126)$$

Es importante entonces que tanto el dato como su error tengan las mismas unidades. Caso de que fueran diferentes habría que indicarlos por separado, p. ej.

$$l = 1.6 \text{ cm} \pm 1 \text{ mm} \quad (1.127)$$

pero siempre es preferible que coincidan.

Cuando se tienen una serie de medidas de un mismo tipo, puede colocarse la unidad a cada una de ellas o indicarlas colectivamente al final entre paréntesis.

**Ejemplo:** Supóngase que se tienen 5 medidas de una longitud. Estas pueden expresarse como:

$$17 \text{ mm} - 15 \text{ mm} - 18 \text{ mm} - 16 \text{ mm} - 16 \text{ mm}$$

o como

$$17 - 15 - 18 - 16 - 16 \quad (\text{mm})$$

En una tabla de valores las unidades deben colocarse preferentemente en la parte superior de la misma, como en el ejemplo siguiente:

$I \text{ (A)} \pm 0.1 \text{ A}$	$V \text{ (mV)} \pm 0.1 \text{ mV}$
1.0	2.5
2.0	4.3
3.8	7.9

En general, todo lo que se indique en la cabecera de una tabla se supone que afecta a todos los valores de la columna correspondiente. Así, en la tabla anterior se entiende que todas las intensidades se miden en amperios y poseen un error de 0.1 A, mientras que las tensiones se miden en milivoltios y tienen un error de 0.1 mV.

Si los valores de una tabla poseen unidades o errores diferentes, habrá que indicarlo junto a cada valor, en lugar de en la cabecera, como por ejemplo:

$I(\text{A}) \pm 0.1 \text{ A}$	$V$
1.0	$410 \pm 10 \text{ mV}$
2.0	$840 \pm 10 \text{ mV}$
3.8	$1.2 \pm 0.1 \text{ V}$

En una gráfica las unidades se indicarán de forma similar: en los extremos de los ejes se indicarán las magnitudes y sus unidades entre paréntesis.

Es importante señalar que los errores absolutos tienen unidades, mientras que los relativos no. También tienen unidades las pendientes y las ordenadas de las rectas de mínimos cuadrados, sus errores respectivos y los sumatorios empleados para calcularlos. El coeficiente de correlación lineal,  $r$ , no tiene unidades.

Recuérdese que la mayoría de las funciones matemáticas (como el logaritmo, la exponencial o el seno) no tienen unidades, como tampoco los tienen sus argumentos. Es tan absurdo hablar de una medida de “ $\ln(2 \text{ m})$ ” como de “ $\ln(2 \text{ m})$ ”, si este valor ha sido obtenido tomando el logaritmo de una medida de longitud.

**Nota:** Entre todas las funciones, la que presenta más problemas es el logaritmo, ya que a menudo se pide representar una cantidad frente al logaritmo de otra (a la hora de obtener una curva potencial o exponencial). ¿Qué se hace entonces con las unidades que acompañan a los datos? ¿Cuánto es el  $\ln(1 \text{ m})$ ? La respuesta es que se supone que antes de hallar el logaritmo se han dividido todos los datos por la unidad correspondiente del SI (en este caso 1 m) resultando datos adimensionales. Por ello, debemos tomar la precaución de pasar todos los datos a las mismas unidades. Por ejemplo, supongamos que tenemos la siguiente tabla

$V$	
2.1 V	
1.1 V	(1.128)
508 mV	
247 mV	

y queremos hallar el logaritmo. Podemos hacerlo de estas dos formas:

$V \text{ (V)}$	$\ln(V)$	$V \text{ (mV)}$	$\ln(V)$	
2.1	0.741	2100	7.650	(1.129)
1.1	0.095	1100	7.003	
0.508	-0.677	508	6.230	
0.247	-1.398	247	5.509	

La diferencia entre los logaritmos de las dos tablas es una constante ( $\ln(1000) = 6.908$ ) que, en la mayoría de los casos de interés (rectas de mejor ajuste, sobre todo) no tiene consecuencias prácticas.

Las unidades que deben emplearse preferentemente son las del Sistema Internacional (SI) que se basa en el kilogramo, el metro, el segundo, el amperio y el Kelvin. Así, es preferible, expresar la masa en kilogramos a hacerlo en gramos.

Las unidades del SI tienen unas abreviaturas estándar que se indican sin punto. Así, es “3 s”, no “3 seg.”. Igualmente, los kilogramos se abrevian como “kg”, los gramos como “g”, los newtons, como “N”, los voltios como “V” y los amperios como “A”. Una lista completa puede encontrarse en multitud de libros.

Las unidades derivadas del SI, como el centímetro o el microfaradio pueden usarse pero con moderación. Así, por ejemplo, una frecuencia puede medirse en “GHz” y una distancia en “cm”, pero no es lógico medir una velocidad en “GHz·cm”.

El uso de unidades que no son del SI ni derivadas del mismo debe evitarse al máximo. No se debe expresar la energía en calorías cuando puede hacerse en Julios. Lo mismo ocurre con los Angstroms (que deben pasarse a nanómetros), los Gauss (a Teslas), los kilopondios (a Newtons) y otras unidades parecidas.

En cuanto a los múltiplos y submúltiplos la tabla de los prefijos es:

Factor	Nombre	Símbolo	Factor	Nombre	Símbolo
$10^1$	deca	da	$10^{-1}$	deci	d
$10^2$	hecto	h	$10^{-2}$	centi	c
$10^3$	kilo	k	$10^{-3}$	mili	m
$10^6$	mega	M	$10^{-6}$	micro	$\mu$
$10^9$	giga	G	$10^{-9}$	nano	n
$10^{12}$	tera	T	$10^{-12}$	pico	p
$10^{15}$	peta	P	$10^{-15}$	femto	f
$10^{18}$	exa	E	$10^{-18}$	atto	a
$10^{21}$	zetta	Z	$10^{-21}$	zepto	z
$10^{24}$	yotta	Y	$10^{-24}$	yocto	y

## Resumen de fórmulas

- Error relativo

$$\epsilon_x = \frac{E_x}{x}$$

- Media aritmética de una muestra

$$\bar{x} = \frac{1}{n} \sum_i x_i$$

- Error de la media

$$E_{\bar{x}} = 2\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}}$$

- Error de una función de una variable  $f(x)$

$$E_f = \left| \frac{\partial f}{\partial x} \right| E_x$$

- Error de una función de varias variables independientes  $f(x, y, \dots)$

$$E_f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 E_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 E_y^2 + \dots}$$

- Error de una suma

$$E_{x+y} = \sqrt{E_x^2 + E_y^2}$$

- Error relativo de un producto

$$\epsilon_{x \cdot y} = \sqrt{\epsilon_x^2 + \epsilon_y^2}$$

- Pendiente de la recta de mínimos cuadrados

$$b = \frac{SS_{xy} - S_x S_y}{SS_{xx} - S_x^2}$$

- Ordenada de una recta de mínimos cuadrados

$$a = \bar{y} - b\bar{x}$$

- Coeficiente de correlación lineal

$$r = \frac{SS_{xy} - S_x S_y}{\sqrt{(SS_{xx} - S_x^2)(SS_{yy} - S_y^2)}}$$

- Sumatorios

$$S = n \quad S_x = \sum_i x_i \quad S_y = \sum_i y_i$$

$$S_{xx} = \sum_i x_i^2 \quad S_{xy} = \sum_i x_i y_i \quad S_{yy} = \sum_i y_i^2$$

- Varianza de  $x$

$$\sigma_x^2 = (\sigma_n)^2 = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{SS_{xx} - S_x^2}{S^2}$$

- Error de la pendiente

$$E_b = \frac{2b}{r} \sqrt{\frac{(1-r^2)}{n-2}}$$

- Error de la ordenada

$$E_a = E_b \sqrt{\sigma_x^2 + \bar{x}^2} = E_b \sqrt{\frac{S_{xx}}{S}}$$

- Extrapolación de un punto para un valor  $x_0$

$$\hat{y} = a + bx_0$$

- Error de la extrapolación

$$E_{\hat{y}} = E_b \sqrt{\sigma_x^2 + (\bar{x} - x_0)^2}$$