

Introducció al tractament de dades

1 Càlcul d'errors

Les lleis de la Física expressen relacions entre magnituds com ara *longitud*, *temps*, *força*, *energia* o *temperatura*. Per aquest motiu, la capacitat de mesurar aquestes magnituds és un requisit de la Física.

Per mesurar una magnitud, s'ha de comparar amb un cert valor de la mateixa que prenem com a referència. A aquest valor se l'anomena *unitat*. Per exemple, són unitats el *metre*, el *kilogram* o el *segon*.

No obstant, el valor exacte d'una magnitud (que anomenarem X) és, en realitat, una abstracció matemàtica. Mai podrem conèixer amb una precisió absoluta quant val aquest valor. Això es degut, en primer lloc, a que els aparells sempre tenen una resolució limitada. També, factors que estan fora de control (condicions ambientals, els reflexos de la persona que fa la mesura, ...) afecten la lectura que ens dona l'instrument i, fins i tot, la pròpia magnitud que volem mesurar.

A la pràctica, doncs, sempre mesurarem un valor (x) que no coincideix amb el valor real de la magnitud (X). Definirem error absolut com la diferència entre aquests dos nombres

$$\epsilon_a = x - X \quad (1)$$

i error relatiu com la relació entre l'error absolut i el valor que hem mesurat

$$\epsilon_r = \frac{x - X}{x} = \frac{\epsilon_a}{x}. \quad (2)$$

L'error absolut d'una magnitud té les mateixes unitats que aquesta. En canvi, l'error relatiu és un nombre adimensional, que serveix per donar-nos una idea de la qualitat de la mesura.

Aquestes definicions matemàtiques no es poden emprar per calcular l'error perquè desconexem X . Per aquesta mateixa raó, mai podrem calcular l'error d'una manera exacta. Sí és possible, però, fer una estimació aproximada d'aquest error.

Estimar l'error vol dir que, a la pràctica, sempre expressarem el resultat de qualsevol mesura en forma d'entorn

$$x \pm \epsilon \quad (3)$$

i amb això indicarem que és "molt probable" que el valor veritable X es trobi a dintre d'aquest entorn.

El procediment per estimar l'error depèn del tipus de mesura:

- Mesures directes que es fan una sola vegada.
- Mesures directes que es fan N vegades.
- Mesures indirectes.

De fet, el significat de "molt probable" dependrà de la tècnica concreta que utilitzem per estimar l'error. Aquest és un dels punts que discutirem especialment.

Abans de continuar, parlem breument de l'anomenat *error de zero*. Es dona quan l'escala d'un instrument està desplaçada i, per aquest motiu, totes les mesures tenen una quantitat (la mateixa) de més o de menys. No es tracta d'un error en el sentit estricte de la paraula doncs si ens adonem de la seva existència després d'haver realitzat les mesures, és possible recuperar els valors correctes. Tan sols caldrà restar la lectura de l'instrument quan estem mesurant una magnitud zero a totes les dades.

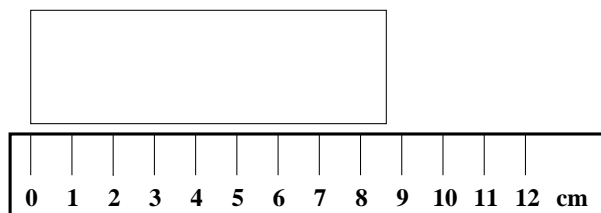


Figura 1: Exemple de mesura que es fa una vegada. El resultat és $8 \pm 1 \text{ cm}$

1.1 Mesures que es fan una sola vegada

En principi, sempre que sigui possible, s'intentarà fer més d'una mesura. Es pot donar, però, el cas en què una mesura sigui destructiva i només es pugui fer una sola vegada. També podria passar que el resultat de totes les mesures realitzades fos idèntic. En ambdós casos, aplicarem la següent tècnica.

Considerarem que l'error es degut exclusivament a les limitacions de l'instrument. En concret, agafarem com a error la *resolució* de l'instrument, es a dir, la diferència entre magnituds més petita que es pot distingir amb l'ajuda de l'instrument

$$X = x \pm \epsilon_{res}. \quad (4)$$

Per exemple, en el cas d'un cronòmetre digital que indiqui els temps fins a les centèsimes de segon, l'error de resolució serà de 0.01 s . Un altre exemple es presenta a la Fig. 1. La resolució del regle és igual a la distància més petita entre les línies de l'escala (1 cm).

Pel que fa a la probabilitat de que efectivament la magnitud X sigui a dins de l'entorn indicat a l'Eq. 4, depèn de la facilitat per fer una lectura correcta. En un instrument ben dissenyat hauria de ser gairebé el 100%.

1.2 Mesures que es fan N vegades

Pensem ara en un cas en que ha estat possible fer vèries mesures

$$x_1, x_2, x_3, \dots, x_N \quad (5)$$

on N és el nombre de mesures que s'han realitzar.

El més probable és que aquests valors x_i siguin diferents entre ells, per molts motius. És molt difícil controlar completament tots els factors que poden afectar el resultat d'un experiment. També cal tenir present l'error humà. Per tenir en compte tots els factors que varien entre mesura i mesura, introduïrem l'*error accidental*.

A més, continuarem tenint un *error de resolució* degut a les limitacions de l'instrument. Per no complicar innecessàriament l'explicació, suposem (de moment!) que el nostre instrument no té problemes en aquest aspecte. Quin valor cal assignar a l'error accidental en aquest cas?

En primer lloc s'ha de calcular la *mitja*

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i \quad (6)$$

i la *desviació estàndar de la mitja*

$$\sigma_{\langle x \rangle} = \sqrt{\frac{\sum_{i=1}^N (x_i - \langle x \rangle)^2}{N - 1}} \quad (7)$$

Aquestes dues magnituds es representen a les calculadores científiques pels símbols \bar{x} i σ_{n-1} , respectivament. Considerem ara l'error accidental expressat per

$$\epsilon_{acc} = \frac{\sigma_{\langle x \rangle}}{\sqrt{N}} = \sqrt{\frac{\sum_{i=1}^N (x_i - \langle x \rangle)^2}{N(N-1)}} \quad (8)$$

Si el nombre de mesures compleix $N \leq 15$ hi haurà una probabilitat entre el 60% i el 70% de que el valor veritable sigui a dins de l'interval determinat per aquest error accidental.

$$X \in \langle x \rangle \pm \epsilon_{acc} \quad (9)$$

Aquest interval s'anomena *interval de confiança* i es caracteritza per la probabilitat de que l'expressió 9 es compleixi. L'expressió 8 és adient la major part de les vegades. No obstant, si ens cal una expressió de l'interval de confiança més acurada podem afegir un factor que depèn del nombre de mesures i de la probabilitat de que el valor veritable hi pertanyi

$$\epsilon_{acc} = F_p(N) \frac{\sigma_{\langle x \rangle}}{\sqrt{N}} \quad (10)$$

Podem trobar aquest factor tabulat als llibres de taules més emprats. Per exemple, si volem un interval de confiança del 97.5%, el factor que utilitzarem és:

N	$F_{0.975}(N)$	N	$F_{0.975}(N)$	N	$F_{0.975}(N)$
2	12.71	9	2.31	16	2.13
3	4.30	10	2.26	17	2.12
4	3.18	11	2.23	18	2.11
5	2.78	12	2.20	20	2.09
6	2.57	13	2.18	30	2.04
7	2.45	14	2.16	40	2.02
8	2.36	15	2.14	∞	1.96

Ara sí, cal tenir en compte que la resolució de l'instrument és limitada. Estimarem l'error de resolució de la mateixa manera que en el cas d'una sola mesura. Al final, obtindrem l'error total a partir de l'expressió

$$\epsilon = \sqrt{\epsilon_{res}^2 + \epsilon_{acc}^2} \quad (11)$$

i indicarem el resultat de la mesura com

$$X = \langle x \rangle \pm \epsilon \quad (12)$$

Si un dels dos errors, l'accidental o el de resolució, és molt més gran que l'altre llavors l'error total serà aproximadament igual a l'error més gran.

Per fixar idees considerem el següent exemple: volem mesurar el temps que tarda un mòbil en caure des de 10 m d'alçada i amb aquesta finalitat prenem 3 mesures:

t (s)	1.43	1.38	1.47
-------	------	------	------

En primer lloc calculem la mitja dels valors

$$\langle t \rangle = \frac{1.43 + 1.38 + 1.47}{3} = 1.42667 \text{ s} \quad (13)$$

Amb aquest resultat podem calcular l'error accidental

$$\epsilon_{acc} = \sqrt{\frac{(1.43 - 1.42667)^2 + (1.38 - 1.42667)^2 + (1.47 - 1.42667)^2}{3 \cdot 2}} \quad (14)$$

El resultat és

$$\epsilon_{acc} = 0.02603 \text{ s} \quad (15)$$

Ara ja tenim els dos errors

error accidental = 0.02603 s

error de resolució = 0.01 s

Calculem l'error total i l'arrodonim a una xifra significativa (recordem que es tracta d'una estimació i no d'un càlcul exacte).

$$\epsilon = \sqrt{0.02603^2 + 0.01^2} = 0.027789 \simeq 0.03 \text{ s} \quad (16)$$

Finalment, presentem el resultat, arrodonint la mitja de manera que l'ordre de magnitud del darrer decimal sigui el mateix que el de l'error

$$t = 1.43 \pm 0.03 \text{ s} \quad (17)$$

Finalment, un parell de comentaris sobre els arrodoniments.

Com que el càlcul de l'error es tracta d'una estimació, els errors s'acostumen a arrodonir sempre a l'alça, per augmentar la seguretat de l'interval de confiança.

Per una altra banda, en algunes situacions es considera convenient expressar l'error amb dues xifres significatives. Com a conveni, expressarem l'error amb dues xifres significatives quan, si només l'arrodoníssim amb una, aquesta seria un 1 o un 2. En qualsevol cas, han de coincidir l'ordre de magnitud de la darrera xifra significativa de la magnitud i de l'error. Per exemple, si el resultat a arrodonir és 1.4278 ± 0.0176 podem escriure 1.428 ± 0.018 per comptes de 1.43 ± 0.02 .

1.3 Mesures indirectes

Sovint una magnitud no es mesura directament sinó que es calcula a partir d'altres magnituds, obtingudes experimentalment. Aquest procés s'anomena *mesura indirecta*.

Considerem una magnitud Z que es pot obtenir a partir de dues altres magnituds X i Y mitjançant una relació teòrica coneguda

$$Z = f(X, Y) \quad (18)$$

Si x i y han estat mesurades experimentalment, després de calcular el seu error (pel mètode que pertorqui) tindrem

$$X = \langle x \rangle \pm \epsilon_x \quad (19)$$

$$Y = \langle y \rangle \pm \epsilon_y \quad (20)$$

Denotarem per $\langle z \rangle$ el resultat de la mesura indirecta de Z i per ϵ_z el seu error.

$$Z = \langle z \rangle \pm \epsilon_z \quad (21)$$

El valor de $\langle z \rangle$ es pot calcular a partir de la relació teòrica que coneixem, agafant com a variables el resultat de les mesures directes

$$\langle z \rangle = f(\langle x \rangle, \langle y \rangle) \quad (22)$$

És més complicat veure quin error cal assignar al resultat de la mesura indirecta. Es pot demostrar que l'error de Z ve donat per

$$\epsilon_z = \sqrt{\left(\frac{\partial f}{\partial X}\right)^2 \epsilon_x^2 + \left(\frac{\partial f}{\partial Y}\right)^2 \epsilon_y^2} \quad (23)$$

Aquesta darrera expressió es pot aproximar per

$$\epsilon_z = \left|\frac{\partial f}{\partial X}\right| \epsilon_x + \left|\frac{\partial f}{\partial Y}\right| \epsilon_y \quad (24)$$

sempre els errors de X i Y siguin relativament petits.

Apliquem-ho a un exemple: volem conèixer el volum d'un cilindre, un cop hem mesurat el seu radi i la seva alçada

$$r = 1.00 \pm 0.01 \text{ m} \quad (25)$$

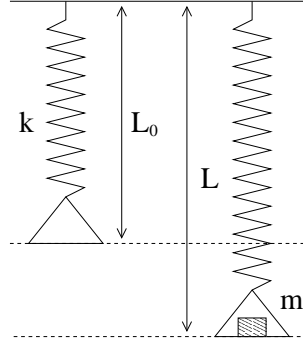


Figura 2: Exemple de dues magnituds relacionades linealment (m i L). Els paràmetres de la molla, k i L_0 , es poden determinar per regressió lineal

$$h = 2.00 \pm 0.01 \text{ m} \quad (26)$$

La relació entre el volum, el radi i l'alçada és

$$V(r, h) = \pi r^2 h \quad (27)$$

Per tant, el volum és

$$V = \pi \cdot 1^2 \cdot 2 = 6.28 \text{ m}^3 \quad (28)$$

i el seu error és

$$\epsilon_V = \sqrt{\left(\frac{\partial V(r, h)}{\partial r}\right)^2 \epsilon_r^2 + \left(\frac{\partial V(r, h)}{\partial h}\right)^2 \epsilon_h^2} \quad (29)$$

Desenvolupant aquesta expressió

$$\epsilon_V = \sqrt{(2\pi r h)^2 \epsilon_r^2 + (\pi r^2)^2 \epsilon_h^2} \quad (30)$$

Finalment, substituint per les dades

$$\epsilon_V = \sqrt{(2 \cdot \pi \cdot 1 \cdot 2)^2 0.01^2 + (\pi \cdot 1^2)^2 0.01^2} = 0.130 \text{ m}^3 \quad (31)$$

El resultat de la mesura indirecta és, després de fer els arrodoniments

$$V = 6.3 \pm 0.2 \text{ m}^3 \quad (32)$$

Fixem-nos que l'error s'ha arrodonit per excès. Això es fa per augmentar la fiabilitat del nostre resultat.

2 Regressió lineal

Tot el que s'ha vist fins a aquest punt tenia com a finalitat mesurar *una* magnitud, amb el seu corresponent error. Ara ens plantejem un problema diferent. Es tracta de trobar quina relació hi ha entre *dues* magnituds.

A l'hora de fer l'experiment, una d'aquestes magnituds és controlada per nosaltres mateixos. Per això, utilitzarem la notació X per referir-nos a aquesta variable (variable independent). L'altra magnitud varia quan nosaltres canviem el valor de X . La notarem com a Y i l'anomenarem variable dependent. Un exemple podria ser una massa que penja d'una molla (variable independent) i l'estirament que produeix aquesta massa a la molla (variable dependent). Aquest exemple està representat a la Fig. 2

Durant l'experiment, donem valors x_i a la magnitud que controlem (la massa, a l'exemple) i mesurem els valors y_i que assoleix l'altra magnitud (l'estirament, en aquest exemple). El resultat es pot expressar en forma de conjunt de N punts

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N) \quad (33)$$

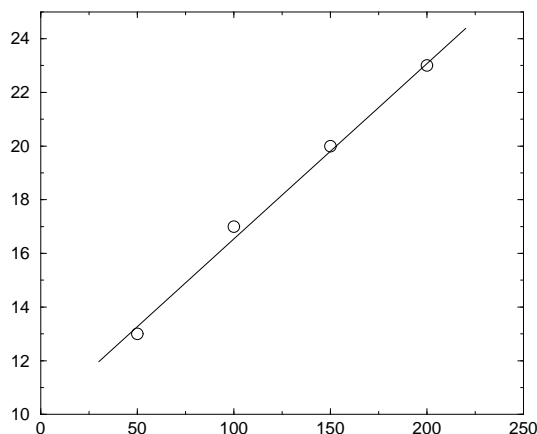


Figura 3: Representació dels punts i de la recta de regressió

El proper pas és representar aquests punts en un diagrama cartesià. Normalment, representarem la variable independent a l'eix d'abscises i la dependent a l'eix d'ordenades. Un cop representats els punts podrem inferir qualitativament quina forma té la dependència de Y respecte X .

Un dels casos més importants té lloc quan la dependència de Y respecte X és *lineal*. Ens n'adonarem perquè al diagrama cartesià els punts tenen tendència a formar una recta. En aquest cas, podrem *ajustar* una recta per *regressió lineal* al conjunt de punts. Es a dir, podrem trobar una recta que sigui la que representi d'una manera més acurada tots els punts. Aquesta recta s'anomena *recta de regressió*. Més endavant donarem una definició més rigorosa, juntament amb el procediment per fer aquest tipus d'ajust. El resultat de l'ajust per regressió lineal és un parell de coeficients, a i b , que determinen la recta de regressió mitjançant l'expressió

$$Y = a + bX \quad (34)$$

Aquesta expressió analítica de la recta de regressió és interessant perquè si la comparem amb alguna fórmula teòrica que relacioni les dues variables, podrem obtenir el valor de paràmetres rellevants.

Tornem a l'exemple de la massa i de l'estirament de la molla. Imaginem que el nostre objectiu és averiguar el valor de la constant k de la molla i la seva longitud natural L_0 . Suposem que, amb aquesta finalitat, hem penjat diverses masses conegudes i hem mesurat els estiraments corresponents

m (g)	50	100	150	200
L (cm)	13	17	20	23

Aquests punts estan representats a la Fig. 3.

La llei de Hooke ens diu que l'estirament d'una molla és proporcional a la força a la que està sotmesa

$$F = k(L - L_0) \quad (35)$$

Per una altra banda, el pes d'una massa ve donat per

$$F = mg \quad (36)$$

Combinant aquestes dues equacions obtenim una relació teòrica de la dependència de L respecte m

$$L = \frac{g}{k}m + L_0 \quad (37)$$

Seguint el procediment que s'explicarà al següent apartat, s'arriba a la conclusió de que la recta de regressió és

$$L(\text{cm}) = 0.0654m(\text{g}) + 10 \quad (38)$$

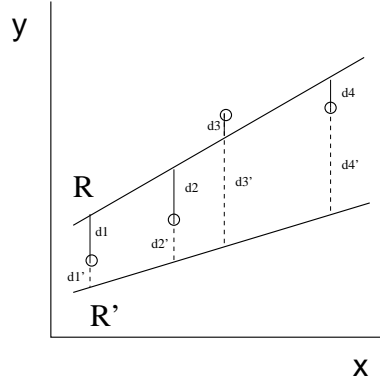


Figura 4: Distàncies d'un conjunt de punts a dues rectes

Aquesta recta també està representada a la Fig. 3.

Comparant les Eqs. 37 i 38 es dedueix que

$$0.0654 \text{ cm} \cdot g^{-1} = \frac{g}{k} \quad (39)$$

$$10 \text{ cm} = L_0 \quad (40)$$

Tenint en compte que g val $981 \text{ cm} \cdot \text{s}^{-2}$, al final aconseguim el valor de la constant k i de la longitud natural L_0

$$k = 15000 g \cdot \text{s}^{-2} \quad (41)$$

$$L_0 = 10 \text{ cm} \quad (42)$$

Ara que s'ha vist per a què pot servir una regressió lineal, veiem la definició matemàtica del que és ajustar una recta per regressió lineal i quin és el procediment per fer-ho.

2.1 Ajust d'una recta per regressió lineal

A la Fig. 4 hi ha la representació d'un conjunt de punts. També s'hi poden veure les distàncies (en vertical) de cada punt a dues rectes. Pensem en la següent quantitat: la suma, per a tots els punts, dels quadrats d'aquestes distàncies a una recta

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + \dots \quad (43)$$

Aquesta quantitat variarà si es calcula per a rectes diferents. La regressió lineal ens dóna els paràmetres (pendent i ordenada a l'origen) de la recta per a la qual aquesta quantitat és *mínima*. Intuitivament, podem pensar en la recta de regressió com en la recta que "representa millor" la posició de tots els punts.

Per trobar els paràmetres presents a l'Eq. 34 s'han d'aplicar les següents expressions

$$a = \frac{S_{xx}S_y - S_xS_{xy}}{D} \quad (44)$$

$$b = \frac{SS_{xy} - S_xS_y}{D} \quad (45)$$

a on s'ha emprat la notació

$$S = N, \quad S_x = \sum_{i=1}^N x_i, \quad S_y = \sum_{i=1}^N y_i$$

$$S_{xx} = \sum_{i=1}^N x_i^2, \quad S_{xy} = \sum_{i=1}^N x_i y_i$$

$$D = SS_{xx} - (S_x)^2$$

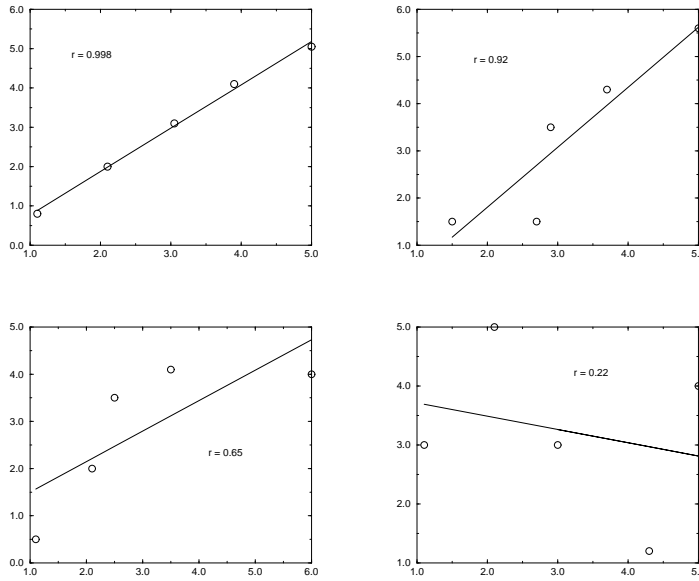


Figura 5: Rectes de regressió amb el seus coeficients de correlació

2.2 Coeficient de correlació

Donat un conjunt de punts

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

podem calcular la quantitat coneguda com a coeficient de correlació:

$$r = \frac{\sum_{i=1}^N (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^N (x_i - \langle x \rangle)^2 \sum_{i=1}^N (y_i - \langle y \rangle)^2}}$$

on

$$\langle x \rangle = \frac{S_x}{S}, \quad \langle y \rangle = \frac{S_y}{S}$$

El coeficient de correlació té un valor absolut entre 0 i 1. Com es pot veure a la Fig. 5, un valor allunyat d'1 indica manca de linealitat. Això pot ser degut a que l'experiment no ha estat realitzat correctament o a que la relació entre les dues magnituds no és lineal. En canvi, un coeficient de correlació proper a 1 assenyalava una bona linealitat.

2.3 Relacions no lineals

A vegades la relació entre X i Y pot ser més complicada que una relació lineal. Per exemple, ens podem trobar que sigui exponencial

$$Y = ce^{kX} \tag{46}$$

Encara podem utilitzar el mètode de la regressió lineal per trobar c i k . Només cal trobar dues variables, dependents de X i Y , que tinguin una relació lineal. En l'exemple de l'Eq. 46 podem aplicar logaritmes a les dues bandes per obtenir una relació lineal

$$\ln Y = \ln c + kX \tag{47}$$

Aquesta relació és lineal si considerem que les nostres magnituds són X i $\ln Y$. Per tant si fem una regressió lineal dels punts

$$(x_1, \ln y_1), (x_2, \ln y_2), \dots, (x_N, \ln y_N) \tag{48}$$

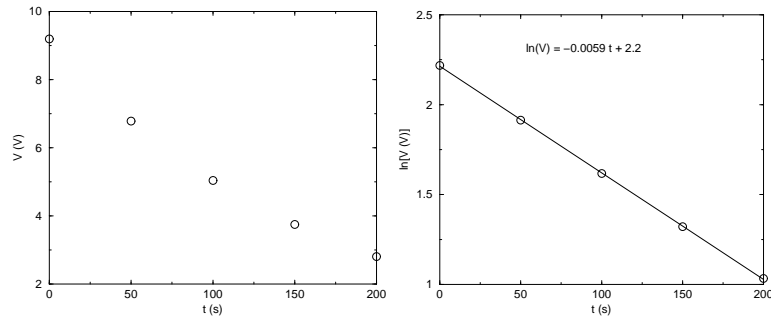


Figura 6: Diferència de potencial entre les plaques d'un condensador a un circuit RC

obtidrem els parametres a i b que caracteritzen la millor recta

$$\ln Y = a + bX \quad (49)$$

Comparant el resultat de la regressió amb el comportament que esperàvem, calculem c i k

$$\ln c = a \quad (50)$$

$$k = b \quad (51)$$

Aquest procediment es pot aplicar a altres tipus de dependència (quadràtica, proporcionalitat inversa, ...). El pas més difícil, però, sempre es trobar dues magnitud lineals, calculables a partir de les que ja coneixem.

Veiem un exemple: un fenomen que segueix una llei de tipus exponencial és la descàrrega d'un condensador en un circuit RC. La diferència de potencial entre les plaques del condensador evoluciona amb el temps segons una llei del tipus

$$V = V_0 e^{kt} \quad (52)$$

Es prenen experimentalment les següents dades

t (s)	0.00	50.0	100.0	150.0	200.0
V (V)	9.19	6.78	5.04	3.75	2.81

que estan representades a la Fig. 6. També s'ha representat el logaritme neperià de V en front de t , que és la gràfica lineal. Es fa, doncs, una regressió amb els punts d'aquesta segona gràfica. Aplicant les Eqs. 50 i 51 obtenim

$$k = -0.0059 \text{ s}^{-1}$$

$$V_0 = e^{2.2} V = 9.03 \text{ V}$$