Tema 1.4 → El método de mínimos cuadrados

☞ **Dónde estudiar el tema 1.4** (seguimos un procedimiento distinto que el libro de Taylor porque anteponemos este tema al capítulo 5 de dicho libro en el que se apoya el autor)

Apartado 2-6 y capítulos 8 y 9. **J.R. Taylor, "Error Analysis". Univ. Science Books, Sausalito, California 1997. O su versión en español " Introducción al análisis de errores" ed. Reverté 2014.**
Apartados 4-3, 5-1, 5-2, 6-5 a 6-9, 6-14 de D.C. Baird, "Experimentación. Una introducción a la teoría de mediciones y al diseño de experimentos" 2ª ed.,Ed. Pearson Educación. México, 1991. [1]

Nuestro problema: Uno de los más comunes e interesantes experimentos que se realizan conlleva la adquisición de varias medidas de dos diferentes magnitudes físicas, relacionadas entre sí, con el fin de averiguar cuál es la relación matemática que las liga. En muchos casos, esa relación es lineal. Con el método que se propone, se busca cuál es la ecuación (de la recta) que liga las dos variables y en qué grado los valores medidos se ajustan a esa ecuación. (Revisa el ejemplo 10 del tema I.)

Ejemplo 1(Baird 6-5): se miden los valores de corriente eléctrica I que pasa por un resistor cuando la ddp entre sus extremos es V y deseamos contrastar las observaciones con el Modelo $V = R\,I$ (ley de Ohm): ¿I =constante × V?

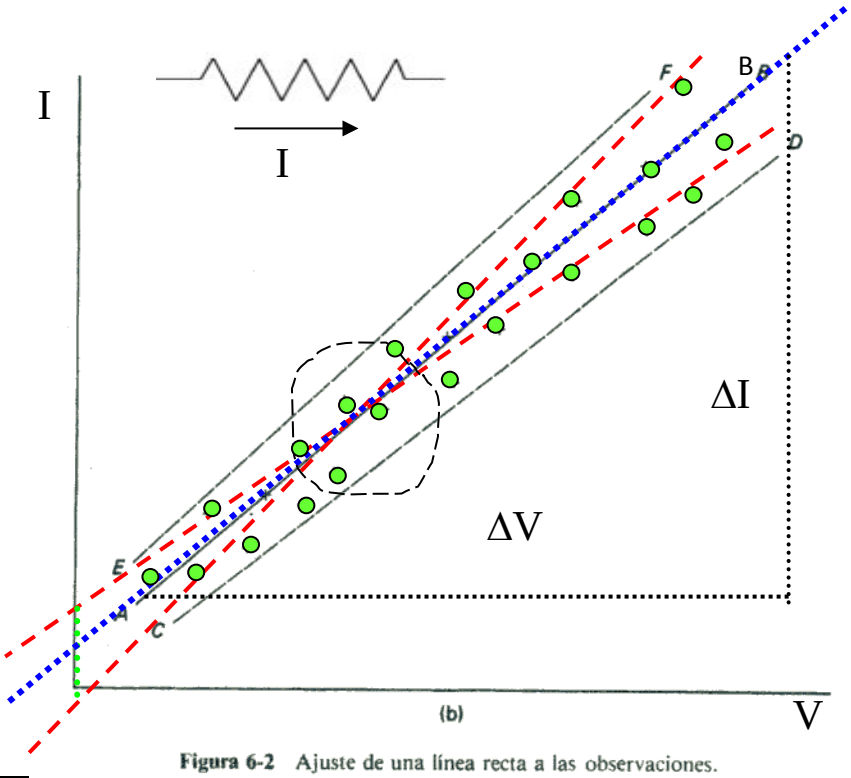


Figura 6-2 Ajuste de una línea recta a las observaciones.

[1] Las ilustraciones están tomadas de este libro.

La magnitud V, que se impone, es la variable independiente. La magnitud I es la respuesta a V y constituye la variable dependiente.

☞ Modelo lineal:

Un método aproximado: ajuste visual (ver la figura).

☞¿Están los puntos experimentales aproximadamente alineados**?***

1 Se delimita la región en donde están las medidas (banda experimental) con dos rectas EF y CD (**rectas en negro**).

☞¿Está **el origen** (V=0, I=0) dentro de la banda experimental?*

2 Se escoge, centrada en esa región, la recta AB (**recta en azul**) más próxima al conjunto de puntos, de manera visual.

3 Se trazan las rectas CF y ED (**rectas en rojo**) para acotar el valor de la pendiente y de la ordenada en el origen de la recta AB. Deben cortarse aproximadamente en el "centro de gravedad" ($\langle x\rangle$,$\langle y\rangle$) del conjunto de puntos (x_i, y_i) con $i = 1\ldots N$.

4 La pendiente de la recta AB se determina gráficamente y representa una estimación de la inversa de la resistencia eléctrica **R** basada en las medidas consideradas conjuntamente.

Dados dos puntos de la recta AB, $P_1 = (I_1, V_1)$ y $P_2 = (I_2, V_2)$

$$\text{Pendiente recta AB} = (I_2 - I_1)/(V_2 - V_1) = \Delta I/\Delta V = R^{-1}$$
$$R = (\text{Pendiente recta AB})^{-1}$$

5 La acotación de **R** se obtiene de las pendientes de las rectas CF y ED (**rectas en rojo**)

$$\pm\Delta R \approx \pm(R_{max} - R_{min})/2$$

siendo R_{max} = (Pendiente recta CF)$^{-1}$ y R_{min} = (Pendiente recta ED)$^{-1}$

* Una respuesta afirmativa sugiere que **hay consistencia o compatibilidad entre el modelo** (ley de Ohm) **y el comportamiento del sistema** (puntos experimentales), ya que no habría explicación, según el modelo, para que circule corriente eléctrica si no se aplica una diferencia de potencial al resistor.

Del ajuste visual a una recta de las medidas de I y V hemos extraído una estimación de la resistencia eléctrica del resistor

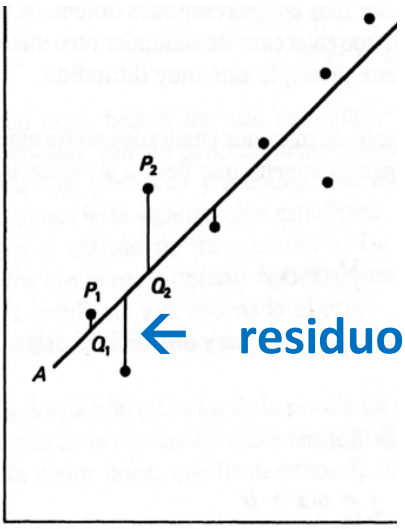
$$R \pm \Delta R$$

☞ Regresión lineal. Ajuste a una recta por el método de los mínimos cuadrados (método analítico)

- Se basa en el principio estadístico de "los mínimos cuadrados" para escoger la mejor recta.
- Admitimos que y es función lineal de x.
- Disponemos de N pares de valores (x, y) medidos experimentalmente.
- Admitimos que podemos despreciar los errores de x y que los errores de y son aleatorios (y obedece una distribución gaussiana, ver tema 1.5)

¿Cuál es la *mejor recta* $y = mx + n$ que ajusta las observaciones experimentales? ¿Qué quiere decir "*mejor recta*" en este contexto?

Mejor recta
es aquella para la cual la suma de los cuadrados de los residuos
$M \equiv \Sigma(P_i O_i)^2$
toma su valor mínimo
(este criterio minimiza la incertidumbre de la pendiente de la recta)



Si la *mejor recta* es

$$y = m\,x + b\,,$$

(línea de regresión de y sobre x)

el *residuo i-ésimo* es: $P_i O_i \equiv \delta y_i = y_i - (m\,x_i + b)$

y la condición impuesta es que la cantidad

$$M(m, b) \equiv \sum_{i=1}^{N} [y_i - (m x_i + b)]^2$$ sea mínima

Para obtener los parámetros m y b debe verificarse, por tanto, que:

$$\frac{\partial M}{\partial m} = 0 \quad \text{y} \quad \frac{\partial M}{\partial b} = 0$$

De esta dos ecuaciones se obtiene que

$$m = \frac{N \Sigma (x_i y_i) - \Sigma x_i \Sigma y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$b = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma (x_i y_i)}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

El método tiene significación estadística porque engloba todos los puntos a la vez para obtener m y b. La incertidumbre (o error estándar) de los parámetros m y b (S_m y S_b) es función de la desviación estándar[2] S_y de la distribución de los valores de y, según las expresiones[3]:

$$S_m = S_y \times \sqrt{\frac{N}{N \Sigma x_i^2 - (\Sigma x_i)^2}} \qquad S_y = \sqrt{\frac{\Sigma (\delta y_i)^2}{N - 2}}$$

$$S_b = S_y \times \sqrt{\frac{\Sigma x_i^2}{N \Sigma x_i^2 - (\Sigma x_i)^2}}$$

[2] No debe confundirse esta cantidad con la desviación estándar asociada con una muestra aleatoria de resultados obtenido de la repetición de una medida que fue definida en la lección 1.3.

[3] Demostración más adelante (Consultar apéndice 2, Baird)

Así queda definida la mejor recta $y = (m \pm S_m)\, x + (b \pm S_b)$

pendiente = $m \pm S_m$ y **ordenada en el origen = $b \pm S_b$**

En la desviación estándar S_y, el promedio de los cuadrados de los residuos de y_i parece que requiere dividir por N. Sin embargo, en el caso particular de que midiéramos solamente dos puntos experimentales (x_1, y_1) y (x_2, y_2), la *mejor recta* pasaría por ambos y los residuos, en ese caso, serían cero. Entonces, se obtendría el resultado absurdo $S_y = 0/2 = 0$ (distribución de las medidas y_i con error cero). Si en lugar de N, escribimos $N-2$, $S_y = 0/0$ quedaría indeterminado. Este resultado pone de manifiesto que el número de medidas (dos) ha sido del todo insuficiente para poder estimar la incertidumbre (asociada a la estadística) en y. Generalizando, el factor que debe aparecer dividiendo, más bien que el número de medidas, debe ser *el número de grados de libertad*. (Si se miden N puntos experimentales y se les imponen dos condiciones $\frac{\partial M}{\partial m} = 0$ y $\frac{\partial M}{\partial b} = 0$, el número de grados de libertad es $N-2$).

☞Cuando el modelo contiene relaciones no lineales entre las variables que se miden

Ejemplo 2:

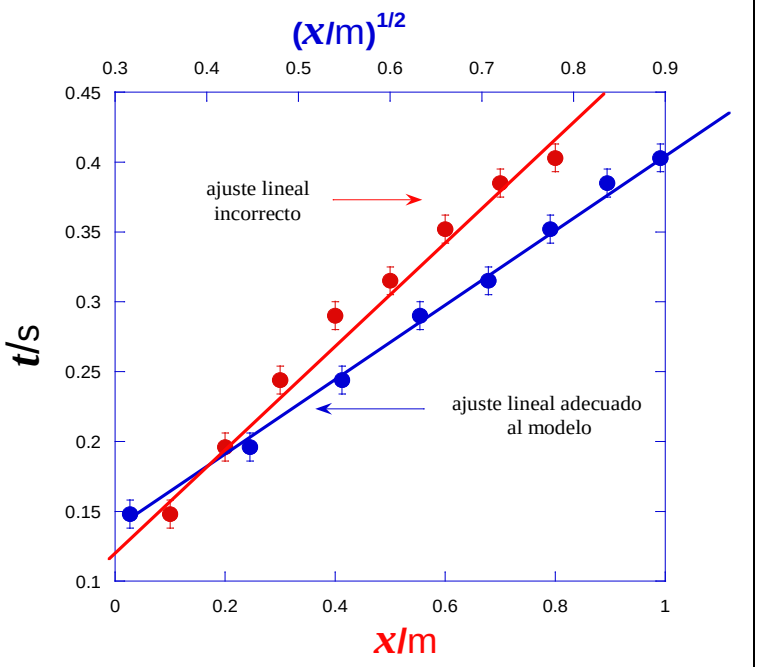
Tabla 1. Experimento real de caída libre: se deja caer una bola de acero desde diferentes alturas x bien determinadas y se miden los correspondientes tiempos t de caída con el fin de averiguar si las alturas y los tiempos están ligados por la ecuación del mov. uniformemente acelerado (modelo). Las alturas se fijan a priori (x es la variable independiente) y se pueden elegir justamente en los puntos mejor definidos de la regla que las mide (por ejemplo, en las marcas calibradas), de

manera que $\Delta x \approx 0$. Se registran las medidas experimentales de los pares de puntos (x, t) en una tabla.

Modelo: $x = \frac{1}{2}gt^2 \rightarrow t^2 = \frac{2}{g}x$ o bien $t = \frac{2}{g}x^{1/2}$

la relación entre x y t es cuadrática, no es lineal

x /m $\Delta x \approx 0$	$(t \pm 0.010)$/s
.1	.148
.2	.196
.3	.244
.4	.290
.5	.315
.6	.352
.7	.385
.8	.403



En la gráfica adjunta, una recta NO ajusta bien los puntos en rojo, cuyos residuos presentan tendencia regular (no aleatoria) con x. Esto suele significar que existe una relación entre las variables x y t que no es de carácter lineal. Los puntos en azul ajustan mucho mejor a una recta y sus residuos muestran un comportamiento más bien aleatorio (como se espera) con x.

☞ Comportamientos no lineales que se pueden "linealizar"

Ejemplo 3: *(variables compuestas)*

1 Movimiento uniformemente acelerado $y = v_0 t + (1/2) a t^2$

2 $y = x^a$, a *es desconocido*

3 $y = ae^{bx}$

4 Periodo T de un péndulo físico:
D es la variable de entrada y T la de salida ($I = I_{cm} + MD^2$ momento de inercia, M masa y D distancia del punto de suspensión al CM del péndulo, g es la gravedad).

$$T = 2\pi\sqrt{\frac{I}{MgD}} \Rightarrow T^2 = \frac{4\pi^2}{Mg}\left(\frac{I}{D}\right)$$

caso	*v indep*	*v depen*	*Eje x*	*Eje y*	*Recta*
1	t_i	y_i	t_i	y_i / t_i	$y/t = v_0 + (1/2) a t$
2	x_i	y_i	$\ln x_i$	$\ln y_i$	$\ln y = a \ln x$
3	x_i	y_i	x_i	$\ln(y_i/a)$	$\ln y/a = b x$
4	D_i	T_i	$(I/D)_i$	T_i^2	$T_i^2 = (4\pi^2/Mg)(I/D)_i$

Ejemplo 4: Se carga un condensador mediante una fuente de alimentación y se miden los voltajes V en función del tiempo t con el fin de averiguar si voltajes y tiempos están ligados por la ecuación

$$V(t) = V_0(1 - e^{-t/RC})$$

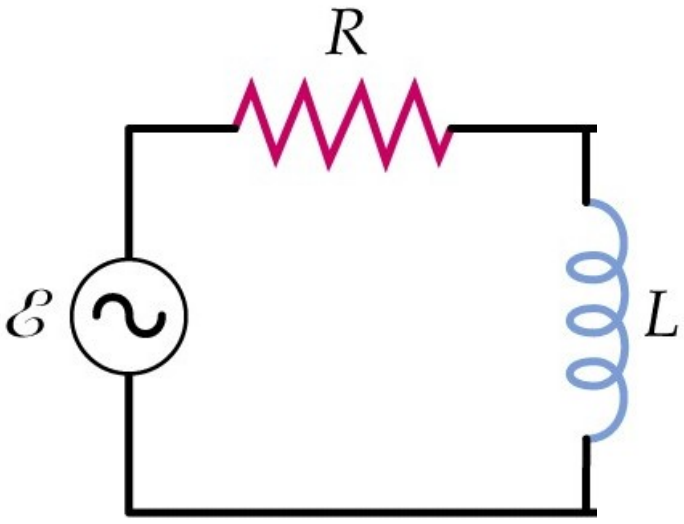
Ejemplo 5: (pg 144, Baird) (Kaleida: Se realizó un experimento para medir la impedancia de un circuito R-L en serie. La impedancia Z se da en función de la resistencia eléctrica R (se mide en *ohmios*), la frecuencia f (se mide en s^{-1}) de la señal de corriente alterna que suministra la fuente de alimentación y la inductancia de la bobina L (se mide en *henrios,* Hr=Ω.s) como

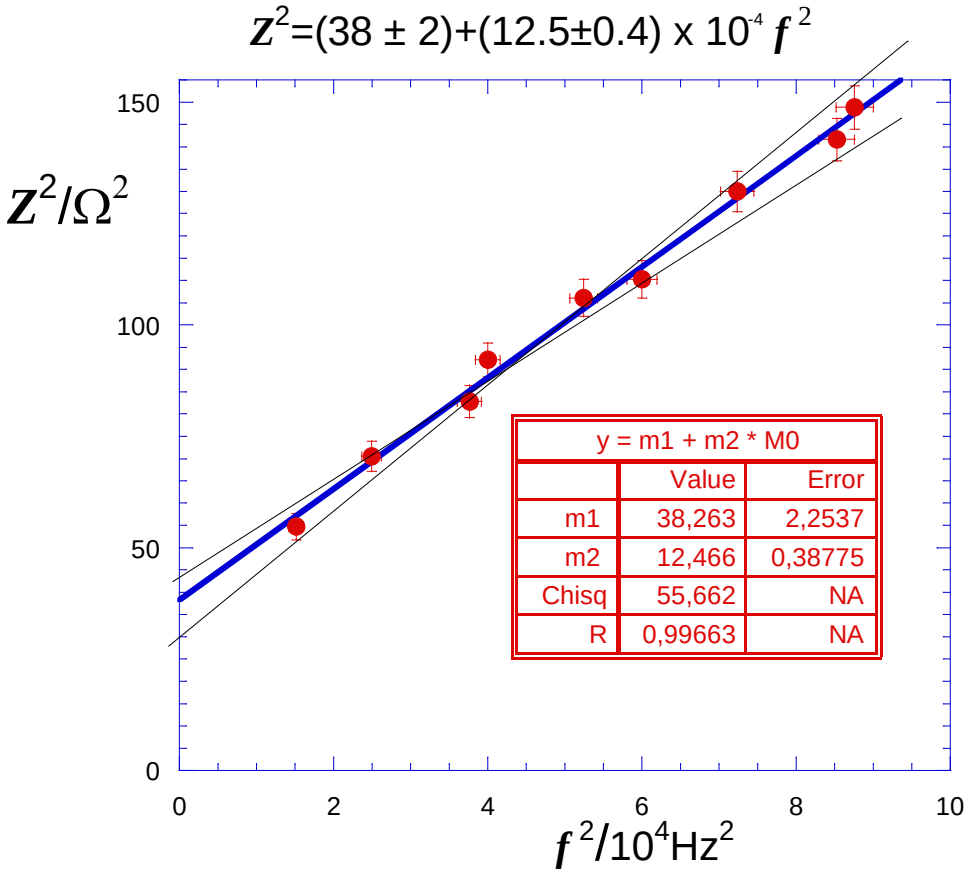
$$Z^2 = R^2 + 4\pi^2 f^2 L^2$$

En el experimento se midió Z como función de f, con la intención de graficar Z^2 en el eje vertical y f^2 en el horizontal para obtener L a partir de la pendiente y R a partir de la ordenada al origen. Las lecturas obtenidas se dan en la tabla.

f, Hz	Z, Ω	f^2	$f(\delta f)$	$\delta f^2 = 2f(\delta f)$	Z^2	$Z(\delta Z)$	$\delta Z^2 = 2Z(\delta Z)$
123 ± 4	7.4 ± 0.2						
158	8.4						
194	9.1						
200	9.6						
229	10.3						
245	10.5						
269	11.4						
292	11.9						
296	12.2						

Las incertidumbres que se dan en el primer renglón se refieren a todas las mediciones en cada columna.





A) Linealizar la función $Z=F(f)$
B) Construir la tabla: f, Z, f^2, Z^2, $\Delta(f^2)$, $\Delta(Z^2)$
C) Representar los puntos (f^2 , Z^2) con sus barras de error
D) Verificar si es adecuado un ajuste lineal
E) Obtener, por el método visual, la **mejor recta.** Calcular las pendientes máxima y mínima para estimar el error que, por este método, se asocia a la **mejor pendiente**. Estimar el valor de la **ordenada en el origen** y su correspondiente error.
F) Obtener, por el método de mínimos cuadrados la **mejor recta,** con calculadora y utilizando kaleidagraph, y escribir los parámetros que la definen con sus cifras significativas.
G) Comparar entre sí los resultados obtenidos por los dos métodos (**precisiones, discrepancias y compatibilidad**).
H) Calcular el mejor valor de R con su incertidumbre
I) Calcular el mejor valor de L con su incertidumbre

RESUMEN: Los dos últimos apartados contienen la información buscada, R y L a partir de las medidas experimentales dadas inicialmente. Se ha admitido que estas medidas estaban sometidas a **errores aleatorios (no hay errores sistemáticos)**. Que los errores en la frecuencia al cuadrado se podían ignorar (si bien, se han utilizado para la obtención de las incertidumbres de los parámetros por el método visual) y se ha buscado una expresión que liga Z con f, o mejor una función de Z con otra función de f en forma lineal. Así, se hapodido utilizar el método de los mínimos cuadrados para encontrar la línea de regresión (mejor recta) de Z^2 sobre f^2. A partir de los parámetros de tal recta, se han obtenido los mejores valores de R y L y se han propagado los errores de los parámetros para expresar, según este método, R y L con sus respectivas incertidumbres.

☞ Obtención de las expresiones de los parámetros de la mejor recta

La condición impuesta es $\Sigma\,[y_i - (\,m\,x_i + b)]^2 = M\ (mínimo)$

$$\frac{\partial M}{\partial m} = 0 \Rightarrow \sum 2\left[y_i - (mx_i + b\right](-x_i) = 0 \Rightarrow \sum \left(x_i y_i - m x_i^{\,2} - b x_i\right) = 0$$

$$\Rightarrow \sum\ (x_i y_i) - m \sum\ x_i^{\,2} - b \sum\ x_i = 0 \qquad (*)$$

$$\frac{\partial M}{\partial b} = 0 \Rightarrow \sum 2\left[y_i - (mx_i + b\right](-1) = 0 \Rightarrow \sum \left(y_i - mx_i - b\right) = 0$$

$$\Rightarrow \sum\ y_i - m \sum\ x_i - \sum\ b = 0 \qquad (\Rightarrow y = mx + b)$$

el "centro de gravedad" de todos los puntos pertenece a la mejor recta

$$b = y - mx \ \text{junto con}\ (*) \Rightarrow \sum \left(x_i y_i\right) - m \sum\ x_i^{\,2} - \left(y - mx\right) \sum x_i = 0$$

$$\Rightarrow \sum \left(x_i y_i\right) - m\left[\sum\ x_i^{\,2} - x \sum\ x_i\right] - y \sum x_i = 0$$

$$\Rightarrow \sum \left(x_i y_i\right) - y \sum x_i = m\left[\sum\ x_i^{\,2} - x \sum x_i\right]$$

$$\Rightarrow m = \frac{\sum \left(x_i y_i\right) - y \sum x_i}{\sum\ x_i^{\,2} - x \sum x_i} = \frac{N \sum \left(x_i y_i\right) - \sum y_i \sum x_i}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2} = \frac{\sum \left(x_i - x\right) y_i}{\sum \left(x_i - x\right)^2}$$

$$b = \bar{y} - \frac{N\sum(x_i y_i) - \sum y_i \sum x_i}{N\sum x_i^{\,2} - \left(\sum x_i\right)^2}\bar{x} =$$

$$= \frac{\sum y_i \sum x_i^{\,2} - \bar{y}\left(\sum x_i\right)^2 - N\sum(x_i y_i)\bar{x} + \sum y_i \sum x_i \bar{x}}{N\sum x_i^{\,2} - \left(\sum x_i\right)^2} =$$

$$= \frac{\sum y_i \sum x_i^{\,2} - \sum x_i \sum(x_i y_i)}{N\sum x_i^{\,2} - \left(\sum x_i\right)^2}$$

☞ Si la recta debe pasar por el origen, uno de los parámetros es cero: $b=0$

Ejemplo 6 ejercicio 8.5 Taylor: La condición impuesta es

$$\Sigma\,[y_i - m\,x_i)]^2 = M\ (mínimo)$$

$$\frac{dM}{dm} = 0 \Rightarrow \sum 2[y_i - mx_i](-x_i) = 0 \Rightarrow \sum\left(x_i y_i - mx_i^{\,2}\right) = 0$$

$$\Rightarrow \sum(x_i y_i) - m\sum x_i^{\,2} = 0 \Rightarrow m = \frac{\sum(x_i y_i)}{\sum x_i^{\,2}}$$

$$S_m^2 = \left(\frac{\partial m}{\partial y_1}\right)^2 S_y^2 + \left(\frac{\partial m}{\partial y_2}\right)^2 S_y^2 + \ldots \Rightarrow S_m = \frac{S_y}{\sqrt{\sum x_i^{\,2}}}$$

S_m es la desviación estándar de la pendiente, función de la desviación estándar S_y de las medidas y_i

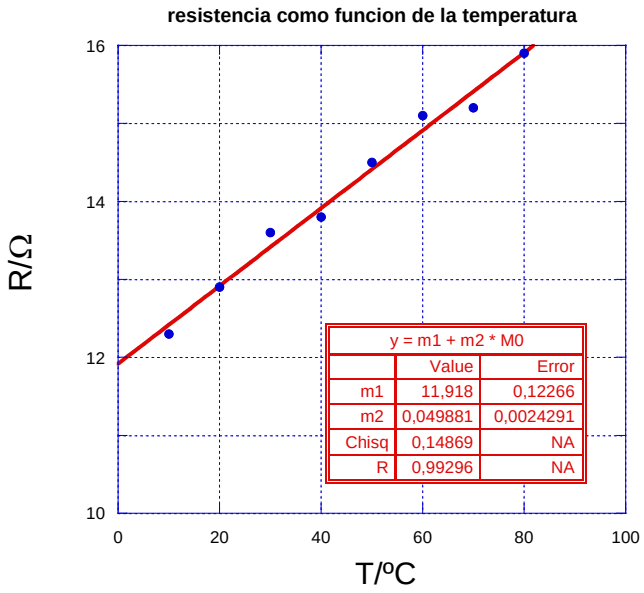
Ejemplo 7 (pg 146,Baird): Se estudia la dependencia de la resistencia de un alambre de cobre con la temperatura. El modelo común viene dado por $R = R_0 (1+\alpha T)$ donde R_0 y R son las resistencias a 0 y T grados centígrados, respectivamente, y α el coeficiente de temperatura (que representa la variación relativa de resistencia por unidad de temperatura) que consideramos constante. Las mediciones de R (variable dependiente) y T (variable independiente) se dan en la tabla siguiente (considérese que los valores de temperatura carecen de error y que las medidas de resistencia tienen un error de 0.1 Ω):

T/ºC	10	20	30	40	50	60	70	80
R/Ω	12.3	12.9	13.6	13.8	14.5	15.1	15.2	15.9

a) Añade un pie de tabla explicativo. b) Utiliza el método de mínimos cuadrados para obtener R_0 y α (mejor estimación y error estándar) expresados correctamente (cifras significativas)

1- utilizando las expresiones de m y b, S_m y S_b, y haciendo operaciones con columnas en kaleidagraph

2- utilizando el ajuste lineal de kaleidagraph



$R_0 = (11.92 \pm 0.12)\ \Omega$

$\alpha = (4.2 \pm 0.2)\ 10^{-3}\ (ºC)^{-1}$

Nota: En este ajuste a una recta, los dos parámetros de la recta m y b no son independientes ya que R_0 aparece en ambos.

$$b = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma (x_i y_i)}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$m = \frac{N \Sigma (x_i y_i) - \Sigma x_i \Sigma y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$S_y = \sqrt{\frac{\Sigma (\delta y_i)^2}{N-2}}$$

$$S_m = S_y \times \sqrt{\frac{N}{N \Sigma x_i^2 - (\Sigma x_i)^2}}$$

$$S_b = S_y \times \sqrt{\frac{\Sigma x_i^2}{N \Sigma x_i^2 - (\Sigma x_i)^2}}$$

☞Obtención de las expresiones de las desviaciones estándar S_m y S_b de los parámetros de la mejor recta

El valor calculado de m depende de los valores y_i: $m = f(y_1, y_2, \ldots y_N) \Rightarrow$

$$S_m^2 = \left(\frac{\partial f}{\partial y_1}\right)^2 S_y^2 + \left(\frac{\partial f}{\partial y_2}\right)^2 S_y^2 + \ldots. \left(recordar \quad que \quad S_y = \sqrt{\frac{\sum (\delta y_i)^2}{N-2}} \right)$$

$$m = \frac{\left[N x_1 y_1 - y_1 \sum x_i + N x_2 y_2 - y_2 \sum x_i + \ldots\right]}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2} =$$

$$\frac{\partial f}{\partial y_k} = \frac{N x_k - \sum x_i}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2} \Rightarrow \left(\frac{\partial f}{\partial y_k}\right)^2 = \frac{N^2 x_k^2 + \left(\sum x_i\right)^2 - 2 N x_k \sum x_i}{\left[N \sum x_i^{\,2} - \left(\sum x_i\right)^2\right]^2}$$

$$\sum_k \left(\frac{\partial f}{\partial y_k}\right)^2 = \frac{N^2 \sum x_i^{\,2} + N\left(\sum x_i\right)^2 - 2N\left(\sum x_i\right)^2}{\left[N \sum x_i^{\,2} - \left(\sum x_i\right)^2\right]^2} = \frac{N}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2}$$

$$S_m = S_y \sqrt{\frac{N}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2}}$$

que es la expresión de S_m (adelantada en la pg.5) .

Para obtener S_b hacemos algo semejante:

$$b = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma (x_i y_i)}{N \Sigma x_i^2 - (\Sigma x_i)^2}$$

El valor calculado de b depende de los valores y_i: $b = g(y_1, y_2, \ldots y_N) \Rightarrow$

$$S_b^2 = \left(\frac{\partial g}{\partial y_1}\right)^2 S_y^2 + \left(\frac{\partial g}{\partial y_2}\right)^2 S_y^2 + \ldots . \left(\text{recordar} \quad \text{que} \quad S_y = \sqrt{\frac{\sum (\delta y_i)^2}{N-2}} \right)$$

$$b = \frac{\left[y_1 \sum x_i^{\,2} - x_1 y_1 \sum x_i + y_2 \sum x_i^{\,2} - x_2 y_2 \sum x_i + \ldots \right]}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2} =$$

$$\frac{\partial g}{\partial y_k} = \frac{\sum x_i^{\,2} - x_k \sum x_i}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2} \Rightarrow \left(\frac{\partial g}{\partial y_k}\right)^2 = \frac{\left(\sum x_i^{\,2}\right)^2 + x_k^{\,2} \left(\sum x_i\right)^2 - 2\left(\sum x_i^{\,2}\right) x_k \sum x_i}{\left[N \sum x_i^{\,2} - \left(\sum x_i\right)^2 \right]^2}$$

$$\sum_k \left(\frac{\partial g}{\partial y_k}\right)^2 = \frac{N\left(\sum x_i^{\,2}\right)^2 + \left(\sum x_i^{\,2}\right)\left(\sum x_i\right)^2 - 2\left(\sum x_i^{\,2}\right)\left(\sum x_i\right)^2}{\left[N \sum x_i^{\,2} - \left(\sum x_i\right)^2 \right]^2} =$$

$$= \frac{N\left(\sum x_i^{\,2}\right)^2 - \left(\sum x_i^{\,2}\right)\left(\sum x_i\right)^2}{\left[N \sum x_i^{\,2} - \left(\sum x_i\right)^2 \right]^2} = \frac{\left(\sum x_i^{\,2}\right)}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2}$$

$$S_b = S_y \sqrt{\frac{\sum x_i^{\,2}}{N \sum x_i^{\,2} - \left(\sum x_i\right)^2}}$$

que es la expresión de S_b (adelantada en la pg.5) .

COVARIANZA Y CORRELACIÓN

Nuestro problema: a) ¿Cómo podemos valorar si dos variables tienen sus respectivas incertidumbres correlacionadas o no, es decir, si podemos o no considerarlas independientes? b) ¿Cómo podemos saber si dos magnitudes físicas están **linealmente** correlacionadas y en qué medida lo están?

☞ Intoducción

Al final del **Tema II Propagación de errores**, obtuvimos las incertidumbres de una función $q(x)$ de una gráfica. Efectivamente, una vez medida en el laboratorio una cantidad en forma estándar $<x> \pm \Delta x$, si queremos obtener alguna función q(x), a partir de una gráfica, q(x) en función de x , buscamos $q_0 \equiv q(<x>)$ en la gráfica, es decir , el punto [<x>, q(<x>)] sobre la curva y los puntos $[<x> - \Delta x,\ q(<x> - \Delta x)]$ y $[<x> + \Delta x,\ q(<x> + \Delta x)]$ también sobre la curva, cuyas ordenadas delimitan el intervalo de incertidumbre de q_0, es decir, para identificar el error

$$\Delta q_0 \cong \left| q(<x> - \Delta x) - q(<x> + \Delta x) \right| /2,$$

de manera que a $<x> \pm \Delta x$ en la gráfica le corresponde $q_0 \pm \Delta q_0$.

Tomamos valor absoluto ya que la pendiente puede ser positiva o negativa y Δq_0 va a tomarse por encima (+) y por debajo (–) de q_0.

Dado que $\pm \Delta x$ y $\pm \Delta q_0$ son cantidades pequeñas respecto de $<x>$ y q_0, respectivamente, una aproximación fundamental de cálculo afirma que

$$\Delta q = q(x + \Delta x) - q(x) \cong (dq/dx)\, \Delta x$$

el cociente $\Delta q / \Delta x \cong dq/dx$, es decir la derivada de q(x).

Así, tuvimos $\Delta q_0 \cong q'(x)]_{<x>}\, \Delta x$: el error de la mejor estimación de q es la derivada de q respecto de x evaluada en $<x>$ y multiplicada por el error de x.

Similarmente, si q es función de más variables, $q(x, y, z, \ldots)$, la expresión anterior se generaliza:

$$\Delta q = q(x + \Delta x, y + \Delta y, z + \Delta z, \ldots) - q(x, y, z, \ldots)$$

$$\cong (\partial q/\partial x)\, \Delta x + (\partial q/\partial y)\, \Delta y + (\partial q/\partial z)\, \Delta z + \ldots \qquad (1)$$

Cuando las incertidumbres Δx , Δy , Δz , … son independientes y aleatorias esta suma directa es sustituida por la suma en cuadratura, según veremos a continuación.

Recuerda (se demostrará) que la expresión (1), suma directa, constituye un límite superior a la suma en cuadratura [la suma de los catetos (suma directa) de un triángulo rectángulo es mayor que la hipotenusa (suma en cuadratura de los catetos)]

Hasta ahora, hemos considerado que los modelos utilizados son satisfactorios, altamente contrastados…*caída libre, ley de Ohm, dependencia de la resistencia con la temperatura, dependencia de la impedancia de un circuito con la frecuencia de la señal eléctrica, dependencia del periodo de un*

péndulo físico con su momento de inercia y su centro de masas, …No siempre las relaciones entre las variables son tan claras. Muchas veces los efectos que buscamos pueden quedar parcial o totalmente encubiertos por perturbaciones que no sabemos separar o neutralizar. Ya mencionamos los procesos que enmascaraban o falseaban la presencia del presunto bosón de Higss en los innumerables eventos registrados. En otras ramas del conocimiento esto es muy frecuente, por ejemplo, es muy difícil poder inferir que del hábito de fumar se deriva un cáncer de pulmón, porque hay innumerables causas que intervienen y enmascaran sin que sea fácil, o posible, eliminarlas. Sin embargo, se acepta sin discusión que ese cáncer tiene tendencia a seguir dicho hábito.
En estos casos, aun descubriendo una correlación, es decir, que una variable tiende a seguir a otra, no se prueba que haya una relación de causalidad. El análisis de regresión proporciona una medida numérica del grado de correlación y se puede evaluar, para un conjunto de medidas, un coeficiente de correlación.

☞ Covarianza en la propagación de errores

Supongamos que buscamos el valor de $q=f(x,y)$ midiendo x e y varias veces para obtener N pares de resultados (x_i, y_i). (Por ejemplo los lados de un aula para determinar su superficie) Podemos calcular el valor medio y la desviación estándar de x e y ($<x>$, S_x, $<y>$, S_y) (recordar sus expresiones en Tema IIIA) y podemos calcular N valores de q,
$q_i=q(x_i, y_i)$ $i= 1…N$.
Dados $q_i \ldots q_N$, obtenemos el valor medio $<q>$ que asumimos como la mejor estimación de q y S_q la mejor estimación de su desviación estándar, como la incertidumbre de los q_i. Los obtenemos a continuación.
Asumiendo, como es usual, que las medidas realizadas x_i , y_i, son próximas a $<x>$, $<y>$, respect., es decir, las desviaciones de las medidas son pequeñas comparadas con las medidas mismas, $\delta x_i << x_i$, $\delta y_i << y_i$, podemos aproximar

$$q_i = q(x_i, y_i) \approx q(< x >, < y >) + \frac{\partial q}{\partial x}(x_i - < x >) + \frac{\partial q}{\partial y}(y_i - < y >)$$

Las funciones derivadas parciales están tomadas en el punto $(<x>, <y>)$ y $<q>$ es, por tanto,

$$< q >= \frac{1}{N}\sum_{i=1}^{N} q_i \approx$$
$$\frac{1}{N}\sum_{i=1}^{N}\left[q(< x >, < y >) + \frac{\partial q}{\partial x}(x_i - < x >) + \frac{\partial q}{\partial y}(y_i - < y >)\right] =$$
$$= q(< x >, < y >)$$

Recuerda, para entender el resultado anterior, que esas derivadas no dependen de i y que la suma de los residuos respecto del valor medio es cero, tal como se demostró al principio del Tema IIIA.

La desviación estándar S_q viene dada por

$$S_q^2 = \frac{1}{N-1}\sum_{i=1}^{N}(q_i - < q >)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{\partial q}{\partial x}(x_i - < x >) + \frac{\partial q}{\partial y}(y_i - < y >)\right]^2 =$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 \frac{1}{N-1}\sum_{i=1}^{N}(x_i - < x >)^2 + \left(\frac{\partial q}{\partial y}\right)^2 \frac{1}{N-1}\sum_{i=1}^{N}(y_i - < y >)^2 +$$

$$+ 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{N-1}\sum_{i=1}^{N}(x_i - < x >)(y_i - < y >) = \left(\frac{\partial q}{\partial x}\right)^2 S_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 S_y^2 +$$

$$+ 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}S_{xy}; \quad \text{siendo} \quad S_{xy} \equiv \frac{1}{N-1}\sum_{i=1}^{N}(x_i - < x >)(y_i - < y >)$$

S_{xy} es la covarianza de x e y (concepto en paralelo con el de la varianza de x, S_x^2, o de y, S_y^2). Puede ser positiva o negativa. Sus dimensiones son las de xy, como las de S_x son las de x.

S_q es la desviación estándar de q cuando las medidas x e y son o no son independientes. Si son independientes, la cantidad $(x_i - <x>)$ es + ó − con igual probabilidad para cualquier valor de y_i, el último sumando es cero o muy pequeño $S_{xy} \approx 0$, al hacer muchas medidas, y se encuentra, en ese caso, que S_q es la suma en cuadratura de S_x y S_y.

Cuando la covarianza $S_{xv} \neq 0$ se dice que los errores en x e y están correlacionados

Ejemplo 8 (pg 213 T): dos ángulos con covarianza negativa

Cinco estudiantes miden los ángulos α y β. Encuentra la media y la desviación estándar de cada ángulo. Calcula $q = \alpha + \beta$. Encuentra el valor de q y su incertidumbre S_q, asumiendo (incorrectamente) que los errores de los ángulos son independientes. Calcula la covarianza S_{xy} y la desviación estándar según la correcta expresión. Compara ambas desviaciones estándar e interpreta los resultados.

La suma en cuadratura da mayor desviación estándar S_q. Debido a que un error + en α va acompañado de un error − en β, hace que en la suma $\alpha+\beta$ los errores cancelen en parte debido a esta correlación y el error asociado sea de una fracción de grado.(ej. 9.2 y 9.3Taylor, Pg222)

Table 9.1. Five measurements of two angles α and β (in degrees).

Student	α	β	$(\alpha - \bar{\alpha})$	$(\beta - \bar{\beta})$	$(\alpha - \bar{\alpha})(\beta - \bar{\beta})$
A	35	50	2	−2	−4
B	31	55	−2	3	−6
C	33	51	0	−1	0
D	32	53	−1	1	−1
E	34	51	1	−1	−1

☞ Se puede probar que $\left|S_{xy}\right| \leq S_x S_y$ (desigualdad de Schwarz)(ej. 9.7 T, pg 224) y por tanto

$$S_q^2 \leq \left(\frac{\partial q}{\partial x}\right)^2 S_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 S_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y} S_x S_y$$

$$= \left[\left|\frac{\partial q}{\partial x}\right| S_x + \left|\frac{\partial q}{\partial y}\right| S_y\right]^2$$

Este resultado, nos permite entender la expresión de la propagación de errores para q que utilizamos cuando sospechamos la posibilidad de que los errores de x e y **puedan no ser** independientes

La suma $$\delta q \approx \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial y}\right| \delta y$$ es un límite superior a la desviación estándar.

☞ Coeficiente de correlación lineal

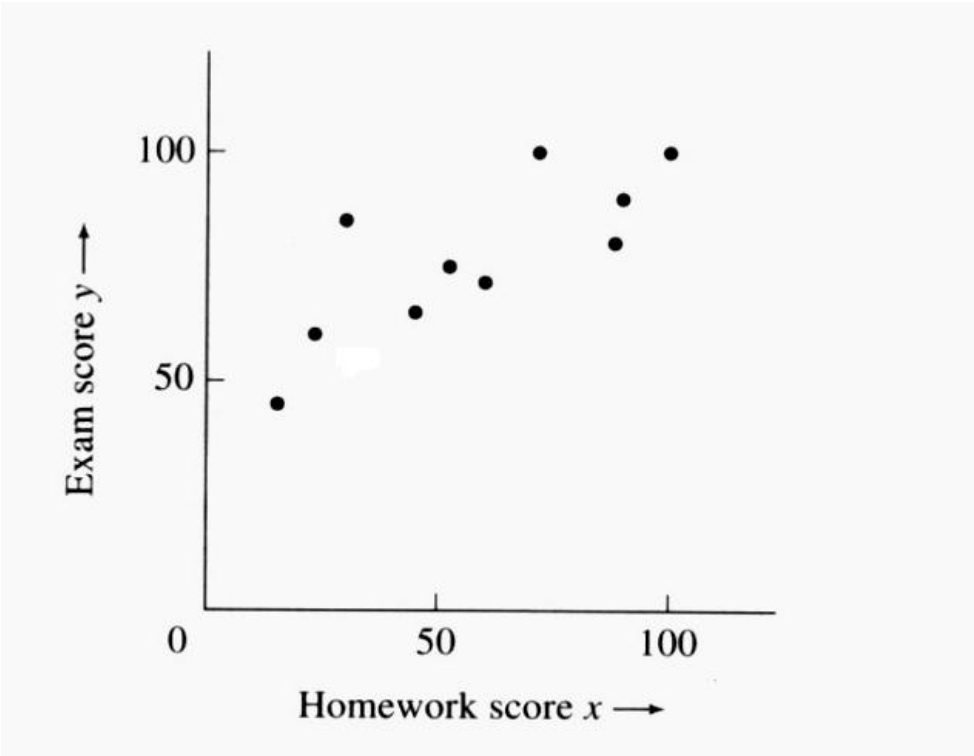
No suele presentarse, en un curso básico de laboratorio, la necesidad ni la posibilidad de establecer una medida fidedigna de una covarianza, pero quizá sí establecer un coeficiente de correlación lineal. Este coeficiente cuantifica en qué medida se cumple la hipótesis de que dos variables están linealmente relacionadas (vale como ejemplo cualquier linealización realizada hasta ahora).

No siempre es fácil decidir si dos variables x e y se muestran linealmente correlacionadas. Las parejas (x_i, y_i) de N medidas pueden sugerir una relación entre las dos variables, pero las dispersiones de las medidas ser tan grandes que carece de sentido tratar de ajustar x e y a una función teórica. Tal es el caso del

Ejemplo 9 (pg 216 T, fig. 9.1 con la tabla 9.3): Entre las calificaciones de los trabajos hechos en casa y las calificaciones de los exámenes, ¿hay correlación? No sabemos nada de las incertidumbres de las "medidas", más bien, ocurre que las calificaciones se conocen exactamente.

Table 9.3. Students' scores.

Student i	1	2	3	4	5	6	7	8	9	10
Homework x_i	90	60	45	100	15	23	52	30	71	88
Exam y_i	90	71	65	100	45	60	75	85	100	80



En qué medida N puntos (x_i, y_i) ajustan a una recta está indicado cuantitativamente por el

coeficiente de correlación lineal r que se define así,

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\sum (x_i - < x >)(y_i - < y >)}{\sqrt{\sum (x_i - < x >)^2 \sum (y_i - < y >)^2}}$$

→ Se verifica que $-1 \le r \le 1$ porque $|S_{xy}| \le S_x S_y$

2→Si los puntos (x_i, y_i) están sobre la mejor recta, se verifica $y_i - <y> = m(x_i - <x>_i)$
(recuerda que el centro de gravedad de los puntos medidos pertenece siempre a la mejor recta)

$\Rightarrow$ $r = \pm 1$

Por tanto, si los puntos (x_i, y_i) se encuentran próximos a la mejor recta,
$r \rightarrow \pm 1$ (buena correlación)

3→Si no hay correlación entre x e y el numerador en la expresión de r, que es la covarianza, tiende a cero $($ya que cualquiera que sea y_i, cada x_i podría estar con igual probabilidad por encima o por debajo de $<x>)$ y el denominador $S_x\, S_y$ es siempre positivo. Entonces, después de muchas medidas, $r \rightarrow 0$

Por tanto, si la correlación lineal es débil r será cercano a cero (pequeña o nula correlación)

☞ Una medida más afinada de la correlación lineal de dos variables x e y se obtiene considerando además de *el coeficiente de correlación* también el número N de medidas $(x_i\ ,y_i)$ que se han utilizado para determinar r . Así, para cualquier valor observado r_o,

$\text{Prob}_N\ (|r\ | \geq r_o)$ es la probabilidad de que N medidas de dos variables NO correlacionadas produzcan un coeficiente de correlación $r \geq r_o$

Por tanto, si obtenemos un r_o para el cual esta probabilidad es pequeña, esto quiere decir que es improbable que nuestras variables estén no correlacionadas, y, por tanto, hay indicios de correlación.

Si $\text{Prob}_N\ (|r\ | \geq r_o) \leq 5\%$ se dice que la correlación es significativa

Si $\text{Prob}_N\ (|r\ | \geq r_o) \leq 1\%$ se dice que la correlación es altamente significativa.

La siguiente tabla (Tabla 9.4, pg 219T) muestra estas probabilidades estadísticas calculadas para diferentes valores de N, que nos dicen cómo debemos interpretar los valores de r dependiendo del número N de medidas realizadas.

Ejemplo: la probabilidad de que 20 medidas (N=20) de dos variables no correlacionadas diera un $r \geq 0.5$, nos dice la tabla que es 2.5%. De manera que si 20 medidas dieran $r_0 = 0.5$ tendríamos evidencia significativa de una correlación lineal entre las dos variables.

Table 9.4. The probability $Prob_N(|r| \geq r_o)$ that N measurements of two uncorrelated variables x and y would produce a correlation coefficient with $|r| \geq r_o$. Values given are percentage probabilities, and blanks indicate values less than 0.05%.

	r_o									
N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
3	100	94	87	81	74	67	59	51	41	29	0
6	100	85	70	56	43	31	21	12	6	1	0
10	100	78	58	40	25	14	7	2	0.5		0
20	100	67	40	20	8	2	0.5	0.1			0
50	100	49	16	3	0.4						0

Resumen: A la pregunta ¿en qué grado N pares de valores (x, y) de dos variables soportan una relación lineal?

Primero calculamos el coeficiente de correlación de las medidas, r_o.
Después, usando la Tabla 9.4, averiguamos, para nuestro valor de N, la probabilidad de que dos variables no correlacionadas tengan un $|r| \geq r_o$. Si es suficientemente pequeña, ≤5%, concluimos que es muy improbable que x e y no estén correlacionadas y , por tanto, muy probable que sí lo estén (la correlación es significativa).
En ese caso, no tenemos una respuesta definitiva sobre si hay o no correlación sino sólo una medida cuantitativa de cuán improbable es que no estén correlacionadas.

Ejemplo 9 (pg 216 T, fig. 9.1 con la tabla 9.3): Calcula el coeficiente de correlación lineal r para los resultados de la tabla 9.3 y decide si las calificaciones de los alumnos están o no están correlacionadas y, si lo están, en qué medida.