Método de ajuste de cuadrados mínimos

La física estudia la interdependencia entre dos o mas magnitudes. Para establecer las leyes físicas que permitan predecir la evolución de un sistema, es necesario conocer en forma experimental el tipo de relación que hay entre las cantidades de las magnitudes involucradas y representarla matemáticamente.

En la práctica, las cantidades están afectadas por errores de medición o fluctuaciones intrínsecas, por lo que es necesario aplicar un algoritmo que permita determinar *la relación más probable* entre las magnitudes físicas vinculadas en forma casual por un mecanismo físico. El caso más sencillo es suponer que la relación entre las magnitudes es lineal, y es el que se estudiará en este curso.

Con el propósito de conocer, evidenciar y/o verificar la ley física que vincula las variables de una serie de N valores (X_i, Y_i) obtenida experimentalmente, se asume que estas son en principio independientes, de modo que sus incertezas también lo son. Para definir cuál es la variable dependiente y cuál la independiente, se observan los errores cometidos en la medición de cada una. La de menor error absoluto será considerada la variable independiente. Los N pares de datos, (X_i, Y_i), son representados en un gráfico que sugiere la forma de una curva, como se muestra en la figura más adelante.

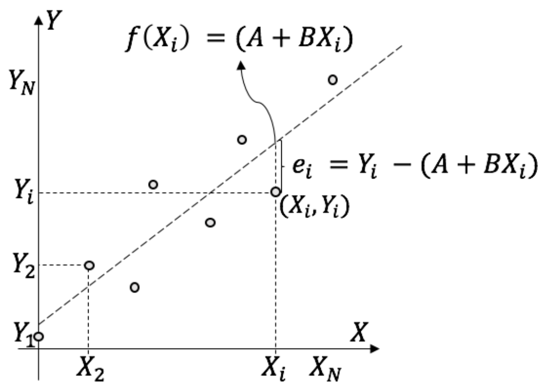
Si la relación entre estas magnitudes es lineal, los puntos experimentales representados en la gráfica $Y(X)$ parecen estar sobre una recta, la cual puede escribirse como:

$$Y = A + B\,X \tag{1}$$

donde A y B son dos parámetros a determinar, que representan la ordenada al origen y la pendiente de la recta respectivamente. El ***método de ajuste por cuadrados mínimos*** consiste en obtener el par de valores A y B que definen a la recta que mejor ajusta a los datos experimentales, que es aquella para la cual la distancia *vertical* de los valores experimentales a la misma sea mínima. Si todos los puntos estuviesen exactamente sobre una recta, se cumpliría que:

$$Y_i - (A + B\,X_i) = 0 \quad \forall\, i \tag{2}$$

Sin embargo, dado que los datos experimentales siempre están afectados de un error, la diferencia en la ecuación (2) no es cero, sino que para cada par de puntos (X_i, Y_i) se observa una diferencia e_i con el valor calculado por la ecuación (1),



La desviación vertical entre cada punto experimental Y_i y el correspondiente valor de función de la recta $f(X_i) = (A + B\,X_i)$ está indicada en el gráfico como e_i. De esta forma:

$$Y_i - (A + B\,X_i) = e_i \quad \forall\, i \tag{3}$$

De las N lecturas (X_i, Y_i) se obtienen N desviaciones e_i, y cada e_i depende del valor que se le asigne a los parámetros A y B.

Se considera como valor más acertado para atribuir a los parámetros A y B, aquel que haga mínima la suma de los cuadrados de las desviaciones:

$$S = \sum_{i=1}^{N} e_i^2 \quad \Rightarrow \quad \textit{Mínimo} \tag{4}$$

La condición de *valor mínimo* para la ecuación (4) se obtiene pidiendo que las derivadas parciales respecto de los parámetros A y B sean cero,

$$\frac{\partial S}{\partial A} = 0 \quad \text{y} \quad \frac{\partial S}{\partial B} = 0 \tag{5}$$

En el caso propuesto, esto sería:

$$\frac{\partial S}{\partial A} = \sum_{i=1}^{N} \frac{\partial (Y_i - (A + B\,X_i))^2}{\partial A} = 2 \cdot \sum_{i=1}^{N} [(Y_i - (A + B\,X_i))(-1)] = 0$$

$$\frac{\partial S}{\partial B} = \sum_{i=1}^{N} \frac{\partial (Y_i - (A + B\,X_i))^2}{\partial B} = 2 \cdot \sum_{i=1}^{N} [(Y_i - (A + B\,X_i))(-X_i)] = 0$$

de donde resulta el sistema de ecuaciones:

$$B \sum_{i=1}^{N} (X_i)^2 + A \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} (X_i \cdot Y_i) = 0 \tag{6}$$

$$B \sum_{i=1}^{N} X_i + A \cdot N - \sum_{i=1}^{N} Y_i = 0 \tag{7}$$

La solución de este par de ecuaciones da los valores buscados de A y B, es decir, el de aquellos que cumplen con la condición de la ecuación (4). En matemática esta manera de ajustar los valores experimentales obtenidos se denomina regresión lineal de Y en X. En una calculadora es el modo LR.

Resolviendo las ecuaciones (6) y (7), se obtienen los valores buscados para la ordenada al origen, A, y la pendiente B de la recta que mejor ajusta a los datos experimentales:

$$A = \frac{\sum_{i=1}^{N} (X_i)^2 \sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} (X_i Y_i)}{D} \tag{8}$$

$$B = \frac{N \sum_{i=1}^{N} (X_i Y_i) - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{D} \tag{9}$$

donde el denominador D también es un valor que da idea de la dispersión de los datos en X:

$$D = N \sum_{i=1}^{N} (X_i)^2 - \left(\sum_{i=1}^{N} X_i \right)^2$$

¿Qué intervalo de incerteza corresponde asignar a los valores de A y de B determinados?

La medida de la variación de los valores Y_i medidos, respecto de los valores calculados, $Y_i = A + B\,X_i$, es llamado *error típico de la estimación* σ_Y, y tiene un significado equivalente al de la desviación típica para una serie de medidas,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}}$$

Si tenemos en cuenta que la desviación estándar en los valores medidos de X es mucho menor que la desviación estándar en los valores medidos de Y ($\sigma_X << \sigma_Y$), podemos afirmar que la desviaciones en el valor de la pendiente, B, y de la ordenada al origen, A, calculadas, se deben casi exclusivamente a la desviaciones en los valores medidos de Y.

Nota: el valor de σ_Y puede obtener directamente de la calculadora en modo regresión lineal.

Luego, aplicando la relación que permite calcular la desviación estándar de una magnitud que está dada en función de otras variables medidas, se llega a la siguiente expresión:

$$\sigma_B \approx \sqrt{\sum_{i=1}^{N} \left(\frac{\partial B}{\partial Y_i}\right)^2 \sigma_Y^2} \quad \text{y} \quad \sigma_A \approx \sqrt{\sum_{i=1}^{N} \left(\frac{\partial A}{\partial Y_i}\right)^2 \sigma_Y^2}$$

Calculando las derivadas parciales de B y A, ecuaciones (8) y (9), se obtienen las siguientes expresiones:

$$\sigma_B \approx \sigma_Y \sqrt{\frac{N}{D}} \quad \text{y} \quad \sigma_A \approx \sigma_Y \sqrt{\frac{\sum X_i^2}{D}}$$

Los errores absolutos en la pendiente y ordenada al origen son:

$$E_B = \frac{\sigma_B}{\sqrt{N}} \qquad E_A = \frac{\sigma_A}{\sqrt{N}}$$

Luego, la **expresión final** para los parámetros de la recta será:

$$\text{Pendiente} \Rightarrow B \pm E_B$$

$$\text{Ordenada al origen} \Rightarrow A \pm E_A$$

Observar que el método de cuadrados mínimos, presentado de esta manera, se puede aplicar a funciones no lineales, siempre que la función que relacione los puntos se pueda linealizar (por ejemplo: $y = bx^a$ puede ser considerada como $y^{1/a} = b^{1/a}\,x$, donde $Y = y^{1/a}$ y $X = x$).

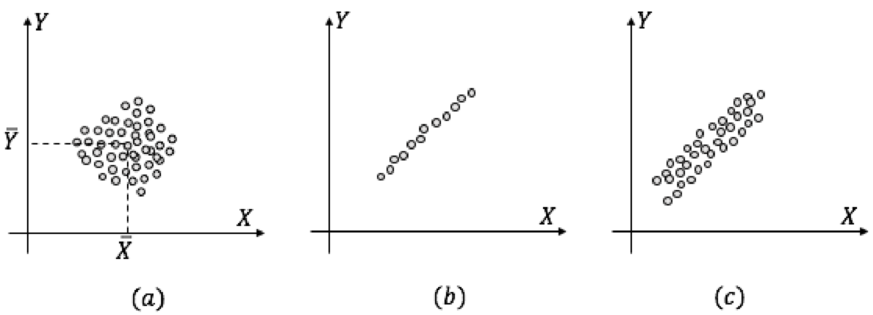
Correlación

Para averiguar si existe o no correlación entre las dos variables se construye un diagrama de puntos (X, Y) en base a un sistema de ejes cartesianos. Se obtiene así gráficamente un conjunto de puntos llamado *diagrama de dispersión*. En general se presentan tres situaciones:

(a) Los valores están distribuidos simétricamente alrededor de los valores de los promedios de X y de Y o sea, del punto de coordenadas $(\overline{X}, \overline{Y})$, formando una nube en torno a dicho punto.

(b) La distribución de puntos (valores) se aproxima a una curva con desviaciones de poca entidad.

(c) Los puntos forman una nube alrededor de una curva con fluctuaciones de cierta importancia.



En el caso (a), las variables no tienen correlación, es decir, no hay dependencia entre ellas y se puede observar que las distancias de estos puntos a una recta hipotética es muy grande, por lo que no se cumple con la hipótesis establecida.

En el caso (b), existe una dependencia casi total entre X e Y. Se dice que existe una correlación fuerte entre las variables, y hay una relación funcional $Y = f(X)$. Las fluctuaciones se deben a las incertezas casuales y de apreciación, y pueden ser compensadas con un análisis estadístico.

En el caso (c), existe una zona de dispersión con pendiente positiva o negativa. La correlación es estadística, es decir, las variables están vinculadas por factores conocidos o ignorados pero muy difíciles de aislar. No es interesante desde el punto de vista de la experimentación, ya que el diseño de la experiencia incorpora excesivas incertezas: permitiría encontrar más de una recta posible, siendo imposible determinar cual de ellas es la correcta.

La gráfica de la curva a la que se aproximan los datos del caso (b) es la llamada *curva de regresión de Y sobre X*. Cuando la curva es una recta se llama *recta de regresión de Y sobre X*.

Una medida de la exactitud con que se realizó la regresión lineal, o el ajuste de los datos experimentales a la recta, está dada por el ***coeficiente de correlación***, $\boldsymbol{R}$, definido como:

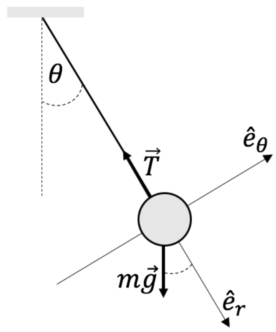
$$R = \frac{\sum\limits_{i=1}^{N} X_i Y_i}{\sqrt{\sum\limits_{i=1}^{N} X_i^2 \sum\limits_{i=1}^{N} Y_i^2}}$$

Una correlación *perfecta* entre la recta y los valores experimentales equivale a $R = 1$, en tanto que un ajuste malo, o nada de correlación entre los datos, equivale a $R \approx 0$.

Para ejemplificar un caso de aplicación del método de cuadrados mínimos, consideraremos a continuación el experimento de oscilación de un péndulo puntual.

Ejemplo de aplicación: Período de oscilación de un péndulo puntual

Utilizando conceptos de dinámica del cuerpo puntual, y con el fin de evidenciar la dependencia entre período de oscilación de un péndulo puntual y la longitud del mismo, se desarrolló un experimento en el laboratorio.



Repasando lo visto en clase, consideremos el diagrama de cuerpo aislado de un péndulo puntual de masa m, suspendido del techo mediante una cuerda de longitud L, y analicemos las fuerzas que actúan sobre la masa en el momento en el que la cuerda está separada de la vertical por un ángulo θ.

Ubicando el sistema de coordenadas polares como se muestra en la figura, y teniendo en cuenta que las fuerzas sobre m son la tensión de la cuerda $\vec{T}$ y el peso de la misma $m\vec{g}$, las sumatorias de fuerzas de la segunda ley de Newton resultan:

$$\sum F_{\hat{e}_\theta}) \longrightarrow -m\, g\, sen\theta = m\, a_\theta = m\, L\, \ddot{\theta}$$

$$\sum F_{\hat{e}_r}) \longrightarrow m\, g\, cos\theta - T = m\, a_r = m\,(-L\,\dot{\theta}^2)$$

Bajo la consideración de oscilaciones a pequeños ángulos, es decir donde el ángulo θ no supere los 10°, el seno de θ podrá aproximarse simplemente como θ, y entonces:

$$-m\, g\, \theta \approx m\, L\, \ddot{\theta}$$

que reordenando nos sirve para demostrar que el péndulo cumple con la ecuación diferencial correspondiente a un movimiento armónico simple (M.A.S.):

$$\ddot{\theta} + \frac{g}{L}\,\theta = 0 \quad \longrightarrow \quad \ddot{\theta} + \omega^2\,\theta = 0$$

La frecuencia angular de la oscilación del péndulo (ω) dependerá entonces únicamente de la longitud del mismo. Utilizando la relación entre el período de oscilación T y la frecuencia angular ω resulta:

$$T = \frac{2\pi}{\omega} \quad \longrightarrow \quad T = 2\pi\sqrt{\frac{L}{g}}$$

Las magnitudes físicas que pueden medirse de forma directa en el laboratorio, son el período de oscilación del péndulo y la longitud de la cuerda con la cual se construye. Entonces, debemos buscar una relación lineal entre estas variables, para visualizar la utilidad del método de ajuste de cuadrados mínimos discutido antes. Obtenemos que:

$$T^2 = \frac{4\pi^2}{g}L$$

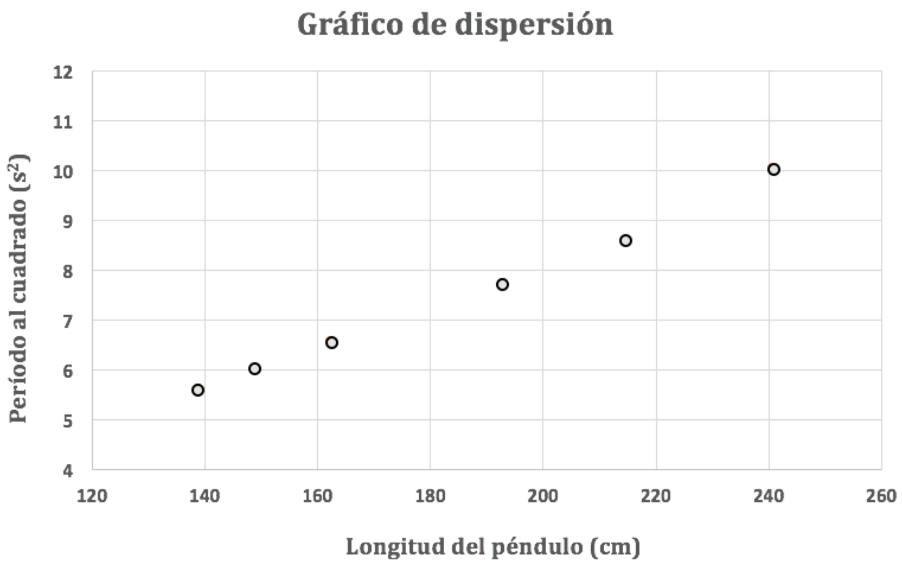
y esta es la relación que nos va a servir para el tratamiento de los datos experimentales, identificando a la variable independiente (X_i) como la longitud de la cuerda del péndulo L y a la variable dependiente como el período de oscilación al cuadrado ($Y_i = T_i^2$). Como veremos más adelante, la longitud del péndulo se midió con un error relativo menor al que se obtuvo en la determinación del período de oscilación. Esto influye en la elección de quién es la variable independiente y cuál la dependiente, como vimos en la teoría del método.

Dentro de la pendiente de la recta que vincula T^2 con L está el valor de la gravedad g, cuya magnitud esperada es de $9.8\,m/s^2 = 980\,cm/s^2$. Este experimento nos servirá entonces para determinar de manera indirecta un valor experimental de la aceleración de la gravedad.

Supongamos que se tomaron registros de los períodos de oscilación del péndulo para $N = 6$ valores de longitud diferentes, como se muestra en la siguiente tabla:

Longitud L (X_i) [$\pm 0.1cm$]	Período T [$\pm 0.15s$]	T^2 (Y_i) [s^2]
138.7	2.37	(5.62 ± 0.71)
148.8	2.46	(6.03 ± 0.74)
162.3	2.56	(6.56 ± 0.77)
192.7	2.78	(7.73 ± 0.83)
214.6	2.93	(8.60 ± 0.88)
240.7	3.17	(10.04 ± 0.95)

Recordemos que el error en el período estará gobernado por el tiempo de reacción del operador y el error en el período al cuadrado se calculará por propagación de errores, por tratarse de una medición indirecta ($E_{T_i^2} = \left|\frac{\partial T_i^2}{\partial T_i}\right| |E_{T_i}| = 2T_i|E_{T_i}|$).



Como vemos en el gráfico de dispersión anterior, donde se representaron los valores experimentales, existe una relación lineal entre las magnitudes medidas, y será factible utilizar el método de ajuste de cuadrados mínimos.

Si se desea realizar el cálculo de A y B, debemos recurrir a las expresiones obtenidas en las ecuaciones (8) y (9), realizando las diferentes sumatorias involucradas en las mismas:

$$\sum_{i=1}^{6} X_i = 1097.8\,[cm]$$

$$\sum_{i=1}^{6} Y_i = 44.58\,[s^2]$$

$$\sum_{i=1}^{6} X_i Y_i = 8493.205\,[cm\,s^2]$$

$$\sum_{i=1}^{6} (X_i)^2 = 208843.36\,[cm^2]$$

$$D = 6 \cdot \sum_{i=1}^{6} (X_i)^2 - \left(\sum_{i=1}^{6} X_i\right)^2 = 47895.32\,[cm^2]$$

Reemplazaremos estas sumatorias en las ecuaciones (8) y (9) para obtener los parámetros buscados:

$$A = \frac{\sum_{i=1}^{6} (X_i)^2 \sum_{i=1}^{6} Y_i - \sum_{i=1}^{6} X_i \sum_{i=1}^{6} (X_i Y_i)}{D}$$

$$A = \frac{208843.36\,[cm^2] \cdot 44.58\,[s^2] - 1097.8\,[cm] \cdot 8493.205\,[cm\,s^2]}{47895.32\,[cm^2]} = -0.284024831\,[s^2]$$

$$B = \frac{6 \sum_{i=1}^{6} (X_i Y_i) - \sum_{i=1}^{6} X_i \sum_{i=1}^{6} Y_i}{D}$$

$$B = \frac{6 \cdot 8493.205\,[cm\,s^2] - 1097.8\,[cm] \cdot 44.58\,[s^2]}{47895.32\,[cm^2]} = 0.04216082 \left[\frac{s^2}{cm}\right]$$

**Nota:* El redondeo de las sumatorias puede llevar a leves discrepancias entre el valor de A obtenido a mano, con calculadora y con algún software como Excel.

Para poder expresar correctamente los resultados obtenidos de A y B debemos calcular sus errores absolutos. Para esto, debemos calcular los residuos, esto es, los valores de los e_i de cada medición y de ahí el error de estimación σ_Y.

Completemos la siguiente tabla:

Longitud L (X_i) [cm]	T^2 (Y_i) [s^2]	$e_i = Y_i - (BX_i + A)$ [s^2]	e_i^2 [s^4]
138.7	5.62	0.0496	0.00246
148.8	6.03	0.0339	0.00115
162.3	6.56	−0.1100	0.01211
192.7	7.73	−0.1648	0.02717
214.6	8.60	−0.1640	0.02690
240.7	10.04	0.1749	0.03057

resultando:

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{6} e_i^2}{6}} = \sqrt{\frac{0.07346}{6}} = 0.110649597\,[s^2]$$

Recordando que:

$$\sigma_B \approx \sigma_Y \sqrt{\frac{N}{D}} \quad \text{y} \quad \sigma_A \approx \sigma_Y \sqrt{\frac{\sum X_i^2}{D}}$$

reemplazamos:

$$\sigma_B \approx \sigma_Y \sqrt{\frac{N}{D}} \quad \longrightarrow \quad \sigma_B \approx (0.110649597) \cdot \sqrt{\frac{6}{47895.32}} = 1.24 \times 10^{-3} \left[\frac{s^2}{cm}\right]$$

$$\sigma_A \approx \sigma_Y \sqrt{\frac{\sum X_i^2}{D}} \quad \longrightarrow \quad \sigma_A \approx (0.110649597) \cdot \sqrt{\frac{208843.36}{47895.32}} = 0.23\,[s^2]$$

de forma que:

$$E_B = \frac{\sigma_B}{\sqrt{N}} \quad \longrightarrow \quad E_B = \frac{1.24 \times 10^{-3}}{\sqrt{6}} \left[\frac{s^2}{cm}\right] = 5.06 \times 10^{-4} \approx 5 \times 10^{-4} \left[\frac{s^2}{cm}\right]$$

$$E_A = \frac{\sigma_A}{\sqrt{N}} \quad \longrightarrow \quad E_A = \frac{0.23}{\sqrt{6}} = 9.40 \times 10^{-2} [s^2] \approx 9 \times 10^{-2} [s^2]$$

Por último, analicemos las unidades que tienen las constantes A y B y, consecuentemente, sus correspondientes errores:

La pendiente B es igual a $\frac{4\pi^2}{g}$, con lo cual tiene unidad de la inversa de la aceleración, esto es s^2/cm.

La ordenada al origen A se espera que sea nula, pero en caso de presentarse un corrimiento experimental en la recta de ajuste, sería un valor correspondiente al eje Y de nuestro sistema de coordenadas, y en dicho eje se representa el período de oscilación al cuadrado. Con lo cual, la unidad de A será s^2.

La expresión correcta de los parámetros de la regresión lineal es:

$$\text{Pendiente} \Rightarrow B \pm E_B \quad \longrightarrow \quad B = (0.0422 \pm 0.0005) \left[\frac{s^2}{cm}\right]$$

que en notación científica puede expresarse como:

$$B = (422 \pm 5) \times 10^{-4} \left[\frac{s^2}{cm}\right]$$

$$\text{Ordenada al origen} \Rightarrow A \pm E_A \quad \longrightarrow \quad A = (-0.28 \pm 0.09)[s^2]$$

que en notación científica puede expresarse como:

$$A = (-28 \pm 9) \times 10^{-2}\,[s^2]$$

Sabiendo que la pendiente B obtenida es igual a $\frac{4\pi^2}{g}$:

$$B = \frac{4\pi^2}{g} \quad \longrightarrow \quad g = \frac{4\pi^2}{B}$$

obtenemos:

$$g = \frac{4\pi^2}{0.0422 \left[\frac{s^2}{cm}\right]} \quad \longrightarrow \quad g = 935.5075262 \left[\frac{cm}{s^2}\right]$$

Como g se obtuvo de una medición indirecta, para conocer su error asociado tenemos que utilizar la fórmula de cálculo de propagación de errores:

$$E_g = \left|\frac{\partial g}{\partial B}\right| |E_B| \quad \longrightarrow \quad E_g = \left|\frac{-4\pi^2}{B^2}\right| |E_B|$$

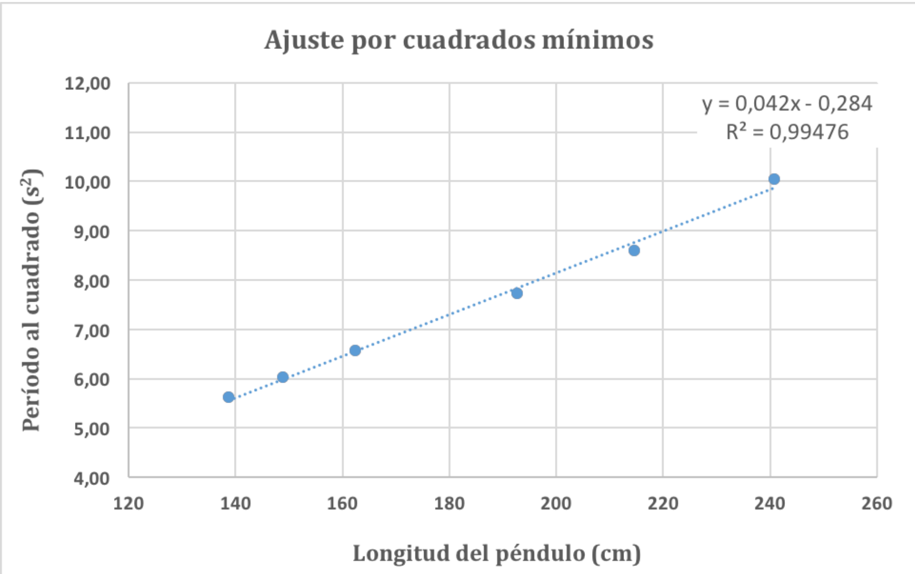
que en función de los valores ya obtenidos resulta:

$$E_g = \left|\frac{-4\pi^2}{(0.0422)^2}\right| |0.0005| \quad \longrightarrow \quad E_g \approx 11\left[\tfrac{cm}{s^2}\right]$$

La aceleración de la gravedad obtenida expresada correctamente es:

$$\mathbf{g = (935 \pm 11)}\left[\tfrac{\mathbf{cm}}{\mathbf{s^2}}\right]$$

A continuación, se muestra la gráfica de los datos experimentales, una vez que se realizó la regresión lineal de los mismos. En la gráfica aparece tanto la ecuación de la recta con los parámetros A y B que mejor ajustan los datos experimentales, como el valor del coeficiente de correlación R.



**Nota*: Los parámetros A y B pueden obtenerse directamente utilizando Excel (o con la calculadora en modo Regresión lineal). Sus errores deberán calcularse de forma manual en el marco de esta materia.

Por último, podemos notar que el valor de R mostrado en la figura es muy cercano a 1, lo que indica una muy buena correlación entre las magnitudes medidas y la recta de ajuste.