

Detección de los errores sistemáticos en los ajustes mediante el análisis de los residuos



Octavio Calzadilla, Julio Vidal, Héctor Borroto

*Facultad de Física, Universidad de La Habana.
San Lázaro y L. 10400 Plaza, Cuba.*

E-mail: calza@fisica.uh.cu

(Recibido el 9 de agosto de 2017, aceptado el 28 de noviembre de 2017)

Resumen

En los valores medidos pueden estar presentes los errores aleatorios y sistemáticos. Cuando los valores son ajustados por una regresión simple el error aleatorio influye directamente en la pendiente de la recta y el sistemático puede estar relacionadas con la pérdida de linealidad de la función de ajuste, desplazamiento del intercepto, etcétera. En general, los errores sistemáticos que influyen en los valores que se utilizaron para la regresión pueden ser detectados mediante el análisis de los residuos. En este trabajo se presenta como mediante el análisis de los residuos se pueden determinar la pérdida de la linealidad de la función de ajuste en una regresión lineal. También se mostrará la vigencia del análisis de los residuos para una regresión no lineal.

Palabras clave: Medición y errores sistemáticos, ajustes de curvas, Enseñanza en Física.

Abstract

In the measured values can be present the random and systematic errors. When the values are adjusted by a linear regression the random error influences directly in the slope of the straight line and the systematic can be related with the loss of linear regression function, the displacement of the intercept, etc. In general, the systematic errors that influence in the values that were used for the regression can be detected by means of the analysis of the residuals. In this work the analysis of the residuals is used to determine that a quadratic terms needs to be incorporated into the model. The validity of the analysis of the residuals will also be shown for a non linear regression.

Keywords: Measurement and systematic error, curve fitting, Research in Physics Education.

PACS: 06.20.Dk, 02.60.Ed, 01.40.Fk

ISSN 1870-9095

I. INTRODUCCIÓN

En la Física Experimental se utiliza el análisis de regresión, una técnica estadística, para comprobar las leyes físicas y sus modelos ya que nos permite predecir y cuantificar la relación de dependencia entre una variable cuantitativa llamada variable dependiente o respuesta y una o más variables independientes llamadas variables predictoras.

Estas leyes vienen dadas a través de expresiones matemáticas escritas explícitamente en función de las variables dependiente y predictoras, es decir, las magnitudes medidas y un conjunto de parámetros.

Por medio de un proceso de ajuste se obtiene un estimado de estos parámetros desconocidos. Uno de los métodos utilizados para determinarlos es el de los Mínimos Cuadrados que minimiza la suma de los residuos al cuadrado. Una vez hallado los parámetros las tendencias o leyes de comportamiento quedan definidas explícitamente lo cual permite modelar los fenómenos físicos para diferentes aplicaciones.

Los datos que se obtienen de la medición pueden estar afectados por los errores aleatorios y sistemáticos los cuales afectan el valor de los parámetros de ajuste, por este motivo

no basta con determinar el valor del parámetro, también hay que determinar la incertidumbre con la que se han hallado los mismos y comprobar el acuerdo entre los datos y el modelo empleado para el ajuste, la bondad del ajuste.

En el Vocabulario Metroológico Internacional [1] se define como error sistemático de medida (2.17): “la componente del error de medida que, en mediciones repetidas, permanece constante o varía de manera predecible”. Es decir, el error sistemático se puede manifestar de formas diferentes en las mediciones: como un valor constante, como un valor que varía proporcional al valor de la variable predictoras y otras [2]. Las cuales van a estar relacionadas con la distribución de los residuos resultantes de la aplicación de la regresión [3].

La regresión lineal es el ajuste de los valores medidos a una recta basado en cierto modelo o suposición. Sean los valores medidos los N pares ordenados (x_i, y_i) . A partir de ellos se estimarán los parámetros de la regresión lineal, la pendiente y el intercepto. La ecuación de regresión viene expresada como la función $f(x_i)$:

$$\hat{y}_i = f(x_i) = A + Bx_i \quad (1)$$

Donde \hat{y}_i es el valor ajustado por la ecuación de regresión. El residuo $e_i = y_i - \hat{y}_i$ es la desviación entre el valor de la variable de respuesta y_i obtenida en la medición y el valor ajustado para la variable predictora. De esta forma los residuos contienen toda la información sobre la forma que el modelo se ajusta a las propiedades estudiadas.

Según [3] si el modelo es correcto los valores esperados de la variable de respuesta son los valores ajustados por la curva de regresión $\eta_i = E(y_i)$. De donde se puede escribir:

$$y_i - \hat{y}_i = \{y_i - \hat{y}_i - (\eta_i - E(\hat{y}_i))\} + (\eta_i - E(\hat{y}_i)). \quad (2)$$

El primer término de la derecha, de la expresión (2), corresponde con las variaciones aleatorias de los residuos; mientras que el segundo representa una contribución al residuo que depende de lo próximo que están los valores medidos y los valores esperados en el modelo. Es decir, si el modelo es correcto este término se anula pero si no es correcto introduce un sesgo debido a un error sistemático.

Un análisis de los residuos, en las regresiones lineales, permite detectar la influencia de los errores sistemáticos en el ajuste [4], como son:

- 1- La forma en que varía la varianza para los diferentes puntos de observación.
- 2- La expresión matemática seleccionada para representar al modelo es inadecuada, aparece una tendencia de variación lineal en los residuos. Esto ocurre por ejemplo en modelos lineales donde se ha omitido un término constante.
- 3- La expresión matemática seleccionada para representar al modelo es inadecuada, aparece una tendencia de variación parabólica en los residuos. Sinex [2] hace referencia a la aparición de este error sistemático debido a una mínima curvatura en un modelo lineal.

Es importante tener en cuenta que cuando los residuos varían de forma aleatoria los errores sistemáticos no están presentes o pueden ser despreciados. Las leyes físicas son aproximaciones al modelo estudiado lo cual quiere decir que si el modelo no describe exhaustivamente al fenómeno estudiado la propia expresión matemática introduce un error sistemático. Un adecuado análisis de los residuos permite hacer las correcciones necesarias del modelo seleccionado. En particular, en los ajustes a leyes lineales cuando aparecen desviaciones de la linealidad en los valores medidos estos pueden ser detectados por un análisis de los residuos.

En el presente trabajo se utilizan los valores medidos, en dos prácticas de laboratorio docente, para ajustarlos a las leyes físicas rigen los fenómenos estudiados. En el primer caso se ajusta siguiendo una regresión lineal y con el análisis de residuos se encuentra la pérdida de linealidad en el fenómeno estudiado. En el otro caso demuestra que el análisis de los residuos permite determinar la influencia de un error sistemático constante en un ajuste utilizando una regresión no lineal. Además, en ambos casos se determina la bondad del ajuste utilizando el análisis de los residuos.

II. DESARROLLO

A. Constante elástica de un resorte

Un experimento típico del laboratorio introductorio de física es determinar la constante elástica de un resorte utilizando las variaciones de la elongación del resorte por el incremento de la masa de las cargas. La ley física que se utiliza para analizar los resultados es la de recuperación elástica, una expresión matemática sin intercepto.

$$\Delta l = \frac{g}{km}. \quad (3)$$

En un gráfico de masa-elongación los valores medidos tienen una representación de una recta sin intercepto, es decir, se ajustará a:

$$y = A + Bx. \quad (4)$$

Donde se espera que $A=0$; de la pendiente del ajuste a la recta se determina la constante elástica.

En la Figura.1a se representan los valores resultantes en un experimento donde se varían los valores de las masas que está suspendida de un resorte, ver Anexo 1.

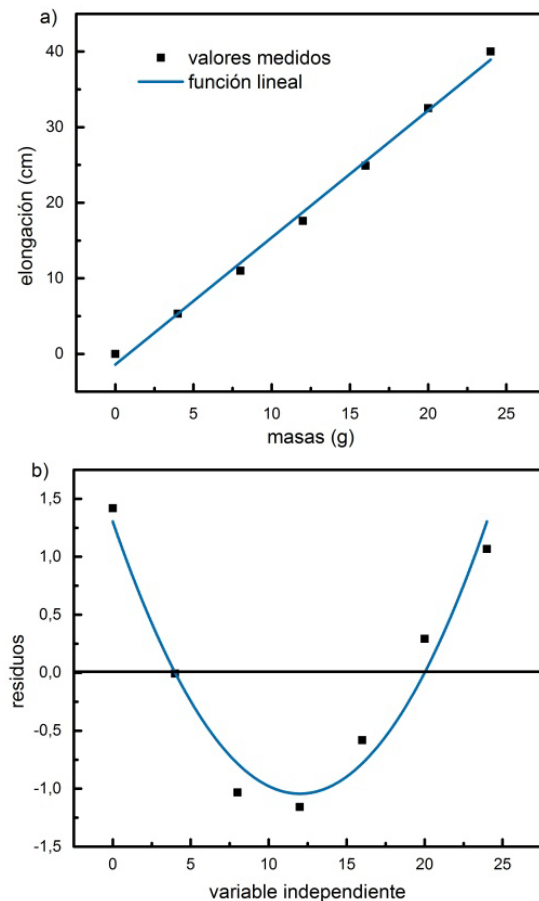


FIGURA.1 Ajuste de los pares ordenados masa - elongación. a) Ajuste lineal y b) Residuos para el ajuste lineal.

TABLA I. Resultado del ajuste para el sistema masa-resorte con una función lineal en la variable independiente.

	Valor del parámetro	Incertidumbre
Intercepto	1.4	0.7
Pendiente	1.68	0.05
R-Square (COD)	0.9953	

Los valores se ajustan utilizando una regresión lineal, en la tabla I se reportan los valores de los parámetros del ajuste. La ley que representa la elongación respecto a la carga es lineal en la variable independiente sin intercepto, ecuación (3). El valor del intercepto diferente de cero está señalando los efectos de un error sistemático constante en la medición. La constante elástica se determina a partir de la pendiente:

$$k = \frac{g}{B} = \frac{980}{1,68} = 583,3 \text{ g cm}^{-2}.$$

La incertidumbre en la constante se determina aplicando la expresión para la incertidumbre estándar combinada:

$$u(k) = gu(B) = 980 * 0,05 = 49 \text{ g cm}^{-2}.$$

El valor de la constante elástica del resorte se reporta como:

$$k = 580 \pm 50 \text{ g cm}^{-2}.$$

En la Figura.1b se muestra que la distribución de los residuos no sigue un comportamiento aleatorio, sino que tiene una marcada dependencia parabólica típica de la existencia de un error sistemático debido a la pérdida de linealidad del modelo utilizado.

La distribución de los residuos que aparece en la Figura.1b se puede ajustar utilizando la expresión no lineal:

$$e_i = A_1 + C(x_i - D)^2. \tag{5}$$

Los valores de los parámetros de ajuste se presentan en la Tabla II.

TABLA II. Valores para el ajuste de la no-linealidad.

	Valor del parámetro	Incertidumbre
A ₁	-1.0	0.2
C	0.016	0.002
D	12	0.4
R-Square (COD)	0.95485	

Es decir, los valores medidos se deben ajustar por una expresión que tenga en cuenta la pérdida de linealidad del modelo expresado en (4). Teniendo en cuenta los resultados de la TABLA II se propone la función de ajuste:

$$y = A_3 + Bx + C(x - 12)^2. \tag{6}$$

En la Figura 2a se muestra cómo queda el ajuste de los valores medidos teniendo en cuenta la expresión (6) que introduce una ligera pérdida de la linealidad en el experimento.

TABLA III. Resultado del ajuste para el sistema masa-resorte con una función no lineal en la variable independiente.

	Parámetro	incertidumbre
A	-2.5	0.2
B	1.68	0.01
C	0.016	0.002
R-Square (COD)	0.99979	

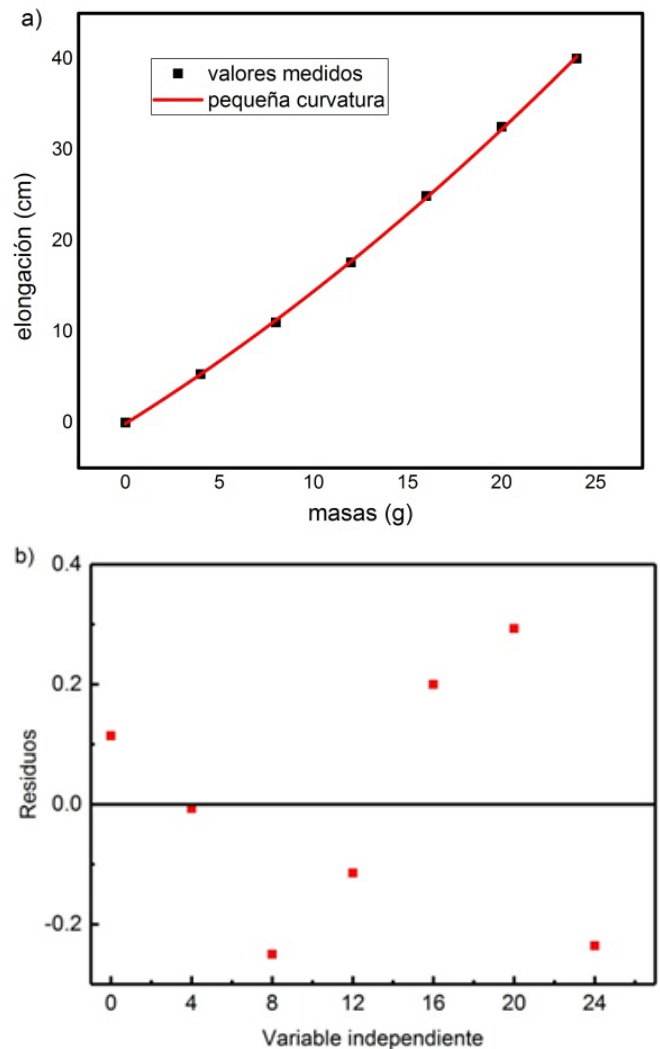


FIGURA.2 Ajuste de los de pares ordenados masa-elongación a una función no lineal. a) Ajuste y b) Residuos para el ajuste.

En la Tabla III se reportan los valores de los parámetros del ajuste para la función (6).

La pendiente para la dependencia lineal no cambia, si disminuye su incertidumbre, quedando el nuevo valor como:

$$k = 580 \pm 20 \text{ g cm}^{-2}.$$

Como se puede observar de la Figura 2b los valores de los residuos se reducen y dejan de tener un patrón definido lo cual es una prueba que el ajuste a la función con una pequeña curvatura es bueno. En esta medición se acumulan los errores sistemáticos constante y de curvatura según la definición utilizada en [2].

B. Ley de Boyle-Mariotte

La comprobación de la ley de Boyle-Mariotte, para una temperatura constante, se cumple que:

$$PV = cte, \tag{7}$$

el producto de la presión absoluta P por el volumen del gas encerrado V , es una constante.

En la práctica V es el volumen de un cilíndrico, se sustituye por el producto del área de la sección transversal del cilindro S por el largo l del mismo quedando entonces la expresión:

$$P = \frac{cte}{Sl}. \tag{8}$$

Esta es otra de las comprobaciones que realizan los estudiantes en los laboratorios docentes. Por lo general la comprobación termina cuando se ajustan los valores medidos a la expresión matemática [5] sin analizar la bondad del ajuste.

Unos valores típicos del experimento se reportan en el Anexo (2). En la Figura.3a se han graficado estos valores, el ajuste se realiza por una función no lineal en los parámetros. En la Figura.3b se presenta el gráfico de los residuos del ajuste.

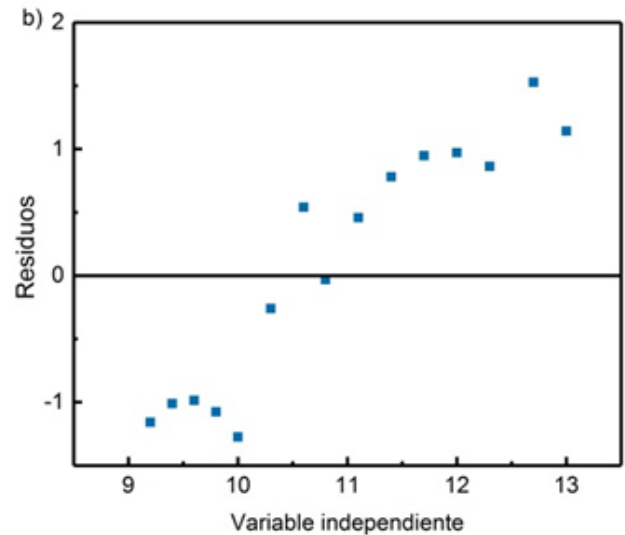


FIGURA 3. Ajuste de los valores experimentales a la expresión de Boyle-Mariotte (ajuste no lineal).

La tendencia de los residuos a aumentar con la variable predictor, esto significa que se necesita una corrección de un valor constante en el volumen del gas comprimido dentro del émbolo.

TABLA IV. Valores del parámetro y el coeficiente de determinación en la ley de Boyle-Mariotte.

Parámetro	Valor	incertidumbre
cte	1274.7	2.7
R-square (COD)	0.99351	

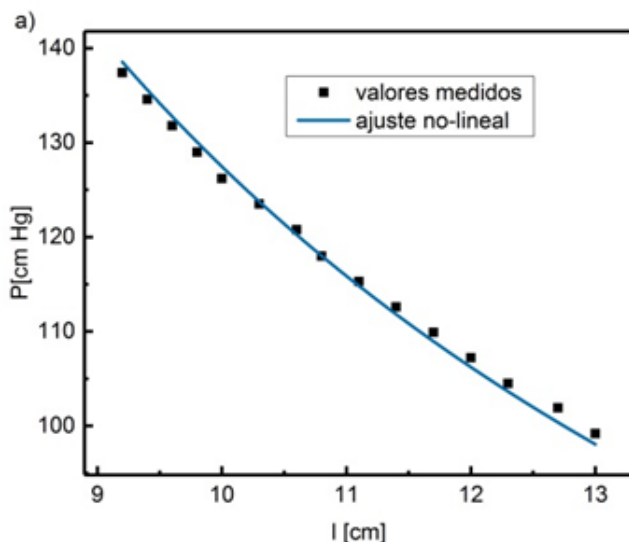
En la TABLA IV se reportan los valores de los parámetros de ajuste para la función y del coeficiente de distribución.

Considerando un error sistemático l_h en la longitud del gas comprimido la ecuación (8) se escribirá como:

$$P = \frac{cte}{S(l+l_h)}. \tag{9}$$

En la Figura.4 se muestra los resultados del ajuste después de la corrección.

Establecer la calidad de un ajuste es fundamental cuando se trabaja con el análisis de regresiones. Como se ha podido ver el uso de los residuos es una manera de comprobar la bondad del ajuste. En la Figura.4a se nota que la curva se ajusta mejor cuando se considera un error sistemático constante en el valor del volumen del gas comprimido. Ese resultado viene respaldado por un coeficiente de determinación mayor y por un patrón aleatorio en la distribución de los residuos.



IV. CONCLUSIONES

En el presente trabajo se ha mostrado como la existencia de un error sistemático en los valores medidos provoca la aparición de un patrón determinado en los residuos del ajuste de los valores. El error sistemático puede estar provocado por diferentes factores que influyen en la medición que provoca la pérdida de representatividad del modelo utilizado para la representación de la ley física.

En particular se comprobó que la pérdida de la linealidad en una ley física provoca un error sistemático derivado de una pequeña curvatura que se debe considerar en la expresión matemática que representa a dicha ley. También se comprobó que un error sistemático constante en una de las variables de una ley no lineal provoca una tendencia en el comportamiento de los residuos.

Los errores sistemáticos no siempre se reflejan en el valor del coeficiente de distribución, en el caso de los errores sistemáticos constantes no se refleja ni en este coeficiente ni en la distribución de los residuos. Solo se pueden estimar si la representación de la ley física no tiene intercepto y al hacer el ajuste aparece un intercepto con valor no despreciable como sería el caso del valor A en la Tabla II.

REFERENCIAS

[1] Centro Español de Metrología, *Vocabulario Internacional de Metrología - Conceptos fundamentales y generales, y términos asociados (VIM)* (3a edición en español), (EDICIÓN DIGITAL), (CEM, Madrid, 2012).
 [2] Sinex Scott, A., *Investigating Types of Errors, Spreadsheets in Education (eJSiE)* 2, 7 (2005).
 [3] Draper, N. R., Smith, H., *Applied regression analysis. (Wiley Series in Probability and Statistics)*, Third Edition, (Wiley-Interscience, USA, 1998).
 [4] Dodge, Y., *The Concise Encyclopedia of Statistics*, (Springer Science + Business Media, New York, 2008), pp. 6-7.
 [5] Rupright, M. E., *Measuring Systematic Error with Curve Fits*, *The Physics Teacher* 49, 54 (2011).

Anexo 1

Valores medidos masa-elongación para determinar la alteración en el comportamiento lineal de la ecuación de los esfuerzos.

Masas (g)	Elongación (cm)
0	0
4	5.3
8	11
12	17.6
16	24.9
20	32.5
24	40

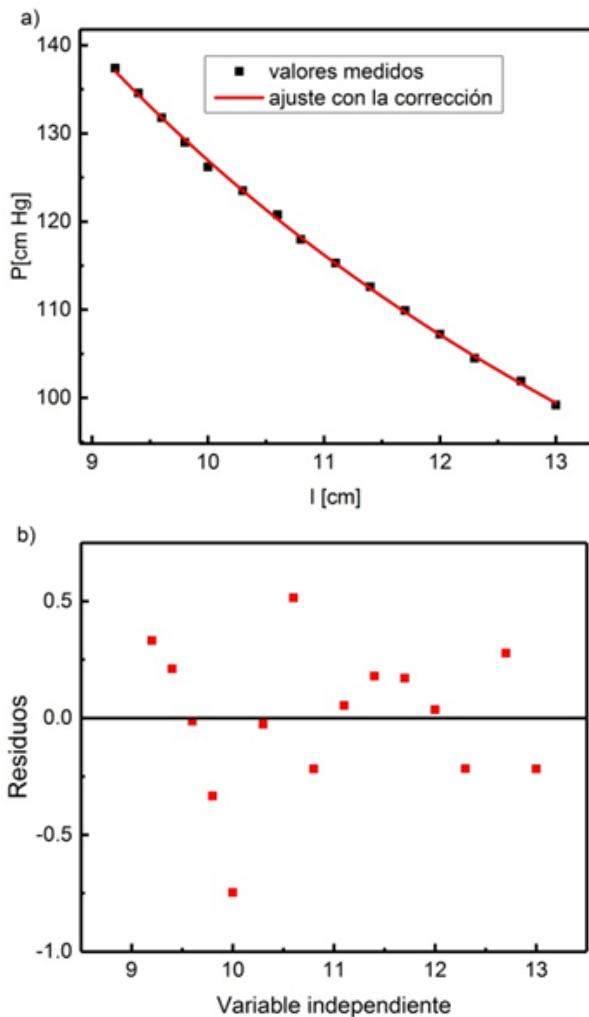


FIGURA 4. Ajuste de los valores experimentales con la corrección del error sistemático (ajuste no lineal).

TABLA V. Valores del parámetro, el error sistemático y el coeficiente de regresión para el ajuste de la ley de Boyle-Mariotte con una función no lineal.

Parámetro	Valor	incertidumbre
cte	1375.3	10.1
l_h	0.83	0.08
R-Square (COD)	0.99929	

III. DISCUSIÓN DE LOS RESULTADOS

Como se ha podido apreciar el error sistemático no tiene que ser un valor constante, puede aparecer de formas diferentes siempre que el valor esperado del modelo no coincida con los experimentales, expresión (2). Esto puede suceder si el modelo es una función lineal y las condiciones de linealidad no son mantenidas en el experimento.

El uso del análisis de los residuos y el coeficiente de distribución permiten evaluar la bondad del ajuste. En la medida que los errores sistemáticos disminuyen la distribución de los residuos son más aleatorios.

Anexo 2

Valores medidos longitud-presión del émbolo para determinar el error sistemático en las mediciones.

Longitud (cm)	Presión (cm Hg)
13	99.2
12.7	101.9
12.3	104.5
12	107.2
11.7	109.9
11.4	112.6
11.1	115.3
10.8	118
10.6	120.8
10.3	123.5
10	126.2
9.8	129
9.6	131.8
9.4	134.6
9.2	137.4