

# Cálculo Rápido de las Incertezas de los Parámetros de Regresión Lineal, con una Calculadora de Bolsillo

César F. Medina y Ana C. Gómez Marigliano

Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán, Tucumán, Argentina.

## Resumen

Las calculadoras electrónicas modernas ofrecen rutinas estadísticas para calcular parámetros de regresión lineal, pero no sus incertezas. El cálculo de éstas, mediante las ecuaciones convencionales, implica varias operaciones laboriosas, tornándose tedioso y confuso. En este trabajo se demuestra un algoritmo muy simple e inmediato para calcular estas incertezas.

**Palabras clave:** regresión lineal, incertezas, pendiente, ordenada al origen, calculadora.

## **Quick Computing of Linear Regression Parameter Uncertainties, with a Pocket Calculator**

### **Abstract**

*Modern electronic calculators offer statistical routines for computing linear regression parameters, but not their uncertainties. Computing them through the standard equations involve several laborious operations, and turns out to be annoying and confuse. In this paper, a very simple and immediate algorithm for computing these uncertainties is demonstrated.*

**Keywords:** linear regression, uncertainties, slope, intercept, calculator.

## Introducción

Las aplicaciones estadísticas básicas, tales como cálculo de incertezas de mediciones, correlaciones entre variables, tendencias y ajustes de curvas, constituyen un aspecto muy importante no sólo en ámbitos de enseñanza-aprendizaje e investigación de ciencias exactas y economía, sino de todas las áreas

científicas y técnicas. Para tales fines, la calculadora de bolsillo es especialmente idónea por su versatilidad y portabilidad, y permite la recolección y procesamiento in situ de los datos.

**Hocking** (2003) señala que la disponibilidad de tecnología moderna promueve el análisis de problemas más amplios y complejos, pero al mismo tiempo,

su uso aumenta la probabilidad de equivocaciones.

Esto es particularmente cierto en el caso de las calculadoras que, entre otras prestaciones, ofrecen rutinas estadísticas rápidas y específicas; pero los interesados en tales aplicaciones no siempre tienen la información o la formación necesarias para utilizarlas del modo más eficiente. A esto contribuyen, también, las limitaciones de diseño y la falta de claridad en los manuales del usuario.

En particular, muchas calculadoras son capaces de obtener dos parámetros de regresión lineal<sup>1</sup>, pero no sus incertezas. Para calcularlas es necesario, en el caso más simple de ajustar una recta, operar con sumas sobre términos lineales, cuadráticos y productos mixtos; lo cual es engorroso aun para usuarios avezados. En lo que sigue, se propone un método mucho más simple para calcular las incertezas.

**La disponibilidad de tecnología moderna promueve el análisis de problemas más amplios y complejos, pero al mismo tiempo, su uso aumenta la probabilidad de equivocaciones.**

Metodología

Según nuestra experiencia docente, los cálculos de regresión lineal pueden plantearse en forma clara y ordenada si se calculan las varianzas de cada variable y la covarianza entre las mismas.

Este trabajo demuestra, mediante algunos pasos algebraicos, que incorporando estos parámetros en las fórmulas de las incertezas de los coeficientes de la recta, conjuntamente con la pendiente de la misma y el coeficiente de correlación de Pearson, que son datos que ofrecen las calculadoras, se obtienen expresiones simples e inmediatas para calcular las incertezas.

*Varianzas y Covarianza de una Distribución de Puntos*

Para un dado conjunto de puntos  $P_i(x_i, y_i)$ , ( $i = 1, \dots, n$ ), se puede calcular la varianza muestral de las  $x_i$  ( $S_{xx}$ ), de las  $y_i$  ( $S_{yy}$ ) y la covarianza muestral entre ambas

variables ( $S_{xy}$ ) mediante las siguientes ecuaciones, **Squires** (1985); **Hocking** (2003):

$$S_{xx} = \sum_{i=1}^n x_i^2 / n - \bar{x}^2 \tag{1}$$

$$S_{yy} = \sum_{i=1}^n y_i^2 / n - \bar{y}^2 \tag{2}$$

$$S_{xy} = \sum_{i=1}^n x_i y_i / n - \bar{x} \cdot \bar{y} \tag{3}$$

donde  $\bar{x}$  e  $\bar{y}$  representan los promedios en cada coordenada. Las ecuaciones (1) y (2) dan una medida de la dispersión de los valores en cada variable, y la (3) es proporcional a la correlación entre ambas.

*Coefficientes de una recta e incertezas de los mismos*

Sean  $a$  y  $b$  la ordenada al origen y la pendiente, respectivamente, de la recta que mejor se ajusta a un dado conjunto de puntos. Estos coeficientes y sus incertezas se calculan con las siguientes ecuaciones (**Squires** (1985); **Hocking** (2003):

$$a = \bar{y} - b\bar{x} \tag{4}$$

$$b = \frac{S_{xy}}{S_{xx}} \tag{5}$$

$$(\Delta a)^2 = \frac{1}{n-2} \left( \frac{S_{yy}}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}^2} \right) (S_{xx} + \bar{x}^2) \tag{6}$$

$$(\Delta b)^2 = \frac{1}{n-2} \left( \frac{S_{yy}}{S_{xx}} - \frac{S_{xy}^2}{S_{xx}^2} \right) \tag{7}$$

Por otra parte, el coeficiente de correlación lineal de Pearson,  $r$ , está dado por:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \tag{8}$$

1. Cabe destacar que aunque los manuales para el usuario llaman regresión lineal a la rutina que modela una recta, y usan otros nombres como regresión cuadrática, logarítmica, etc., a las rutinas que ajustan otros tipos de curva; la expresión "regresión lineal" no indica que la función ajustada sea lineal, es decir, una recta; sino que

el sistema de ecuaciones que se plantea, a partir de establecer una condición de mínimo para el error, es lineal en sus coeficientes. Sin embargo, el ajuste de una recta es el más simple algebraicamente, y permite modelar, mediante cambios de variables, cualquier función de dos parámetros.

o bien, usando la ecuación (5),

$$r = b \sqrt{\frac{S_{xx}}{S_{yy}}} \quad (9)$$

con la cual, la ecuación (7) puede escribirse

$$(\Delta b)^2 = \frac{b^2}{n-2} \left( \frac{1}{r^2} - 1 \right) \quad (10)$$

La ecuación (10) permite calcular la incerteza de la pendiente a partir del número de puntos  $n$ , y de dos parámetros dados directamente por la calculadora:  $b$  y  $r$ .

Nótese que, dado que  $r$  varía entre  $-1$  y  $1$ , el paréntesis del segundo miembro varía entre  $0$  e  $\infty$ , como debe ser. La incerteza es nula cuando todos los puntos están perfectamente alineados, y es infinita cuando están completamente dispersos y sin correlación, dado que en tal caso la pendiente varía entre  $-\infty$  e  $\infty$ .

Por su parte, el cálculo de la incerteza de la ordenada al origen puede simplificarse reemplazando la ecuación (7) en la (6) y reordenando según la (1):

$$(\Delta a)^2 = (\Delta b)^2 \frac{\sum_{i=1}^n x_i^2}{n} \quad (11)$$

donde el primer factor del segundo miembro puede calcularse según la ecuación (10), y al segundo factor lo da directamente la calculadora.

Nótese que la incerteza de la ordenada al origen depende de la incerteza de la pendiente, como debe ser.

#### Rutinas estadísticas - Recomendaciones

Esta sección no pretende reemplazar al manual de la calculadora, que conviene leer. Antes bien, es una mera guía sinóptica de los estándares más comunes en diversas marcas.

Las teclas estadísticas suelen estar marcadas con corchetes azules y/o con el nombre de la rutina en

cuestión (Stat, SD, LR, etc.). En los modelos modernos, para ingresar a ellas se debe apretar la tecla "mode" y seguir el menú correspondiente (una sola variable o regresiones de dos variables). Si el ingreso de los datos no es tabular, uno debe asegurarse de borrar las memorias estadísticas antes de comenzar. Para cualquier cálculo anexo o acceso a memorias o variables, debe salirse del submenú de ingreso, pues de lo contrario se hará un ingreso erróneo. Las principales funciones estadísticas y los parámetros suelen presentarse en el teclado numérico como segunda función (*shift+tecla numérica*), no se debe confundir los parámetros de la recta con las unidades de memoria  $A$  y  $B$ .

#### Conclusiones

- Las incertezas de ambos parámetros de la recta modelada pueden obtenerse inmediatamente de las ecuaciones (10) y (11).

- Estas ecuaciones pueden contribuir a una mayor eficiencia en la obtención de resultados y una mayor confiabilidad en las predicciones científicas y técnicas, sobre todo, teniendo en cuenta que las calculadoras de bolsillo son herramientas insustituibles en investigaciones de campo, jornadas de estudio, etc., y un gran sector de usuarios no tiene formación específica en matemática.



## Referencias Bibliográficas

**Hocking, R. R.** (2003) *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 2<sup>nd</sup> ed., John Wiley & Sons Inc., Hoboken, Nueva Jersey, EEUU.

**Squires, G. L.** (1985) *Practical Physics*, 3<sup>rd</sup> ed., Cambridge University Press, Nueva York, EEUU.

Este artículo se escribió en el mes de setiembre de 2013 en la Facultad de Ciencias Exactas y Tecnología de la Universidad Nacional de Tucumán.

**César Francisco Medina**

Doctor en Física, Facultad de Ciencias Exactas y Tecnología, UNT. Se desempeña como Profesor Adjunto en la Cátedra de Laboratorio de Física I y II (FACET - UNT) y ha dictado cursos de grado y posgrado en física del plasma y metrología. Revista con categoría II en el Programa de Incentivos. Ha dirigido y codirigido diversos proyectos de investigación de física de la alta atmósfera y ha publicado trabajos en revistas y congresos internacionales sobre temas de aeronomía, mecánica de fluidos, física experimental y docencia. Se ha desempeñado como árbitro de la revista *Earth, Planets and Space (EPS)*, y ha realizado diversas tareas de gestión, organización, divulgación, servicios especiales y asistencia técnica a nivel internacional.

**Ana Clelia Gómez Marigliano**

Licenciada en Física de la Facultad de Ciencias Exactas y Tecnología (FACET) de la UNT y Doctora en Ciencias Químicas (Facultad de Bioquímica, Química y Farmacia (FBQyF), UNT). Se desempeña como Profesor Adjunto en la Cátedra de Laboratorio de Física I y II (FACET - UNT) y revista con categoría III en el Programa de Incentivos. Dirige el Proyecto de Investigación "Termodinámica de Soluciones de Sistemas Líquidos Binarios no Electrolíticos. Análisis Espectroscópico" del CIUNT, que actualmente consta de dos líneas de trabajo, una en Físicoquímica y otra en Física Médica. Ha publicado numerosos trabajos de investigación en revistas internacionales y dictado seminarios. Es revisor de revistas Internacionales como el *J. Chem. Eng Data* y ha desarrollado Estadías de Investigación en Centros Nacionales como el Cequinor e Internacionales como el Dpto. de Ingeniería Química de la Universidad de Santiago de Compostela, España.