3.6. CONFERENCIA PARALELA 6

REGRESIÓN LINEAL BIVARIADA CON RESIDUALES NO VERTICALES

Álvaro de Jesús Villota Viveros, vvvillota@live.com, Universidad Cooperativa de Colombia – Academia Nacional de Medicina.

Resumen. Tradicionalmente, la regresión lineal bivariada recurre a los residuales en la variable dependiente "Y" para definir parámetros, sin embargo las distancias de los puntos a la Recta de Mínimos Cuadrados deben estar medidas en una orientación diferente dependiendo de la relación entre las varianzas de los errores de las dos variables. Esto sucede cuando se practican mediciones en ambas variables. Si las varianzas en los errores son iguales, las distancias entre los puntos y la recta se medirán sobre trayectorias perpendiculares a la misma. Pero si la relación de varianzas tiene otro valor, las orientaciones de las distancias pueden tomar cualquier ángulo en el intervalo (0 - π). En el presente trabajo se indica el proceso para obtener los parámetros bajo cualquier relación de varianzas y se expone un método inédito creación del autor, que generaliza todos los casos y se coteja con la Regresión de Deming que aborda también este tópico.

Palabras claves. Regresión lineal bivariada, relación de varianzas de errores, Recta de Mínimos Cuadrados

4. Presentación del problema.

La relación de varianzas de errores en la Regresión Lineal Bivariada, es un factor que no se aplica con frecuencia en la solución de problemas teórico prácticos, por lo tanto los resultados de los proyectos que emplean la Recta de Mínimos Cuadrados se privan de su importante instrumentación. El estudio que se presenta enfatiza en la importancia del mencionado cociente y suministra las herramientas conceptuales necesarias para su aplicación

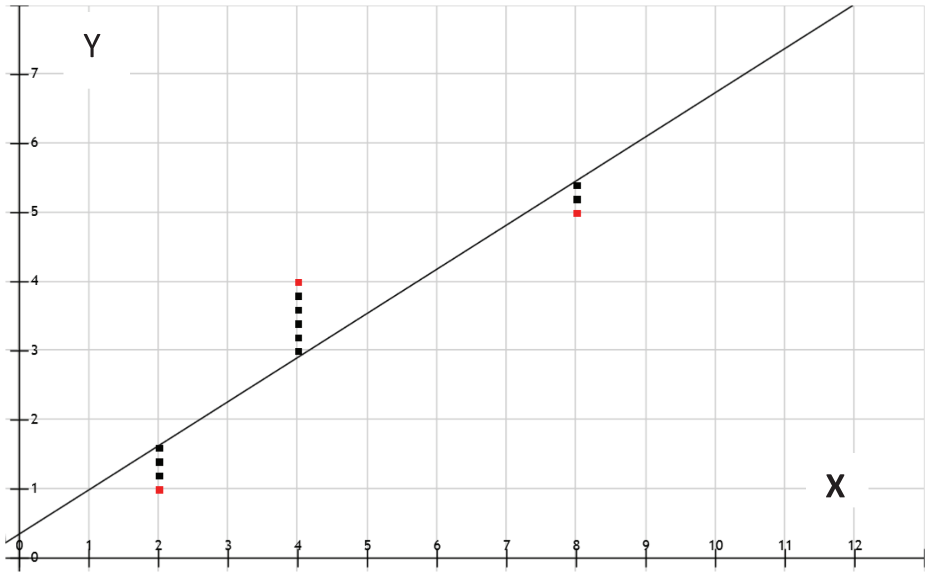
5. Desarrollo de la temática.

1.- RECTA DE MINIMOS CUADRADOS TRADICIONAL (RESIDUALES VERTICALES)

En regresión lineal bivariada tradicionalmente se considera una sola variable susceptible de presentar error, justamente la variable dependiente, en este caso es justificable que los residuales se midan solo en la dirección de dicha variable.

La determinación de los parámetros de corte (a) y de pendiente (b) en la recta:
$y = a + bx$

Se realiza minimizando la suma de cuadrados de distancias verticales a la recta



Gráfica 1: Residuales o distancias verticales que elevadas al cuadrado se sumaran para luego minimizar dicho resultado.

El resultado de esa labores el siguiente:

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

$$b = \frac{Cov\,(x, y)}{V\,(x)}$$

Si se llama δ a la razón de varianzas:

$$\delta = \frac{\sigma_y^2}{\sigma_x^2}$$

En este caso:

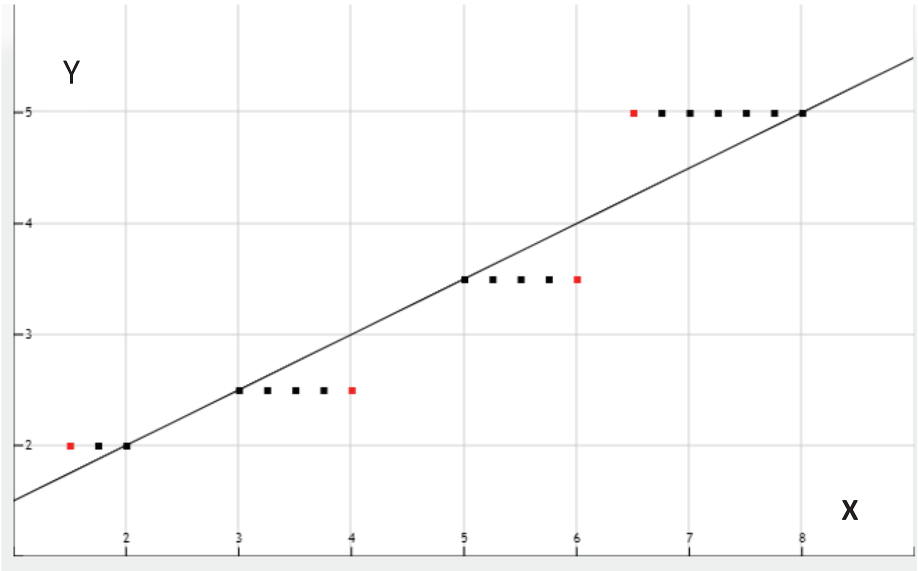
$$\delta = \frac{\sigma_y^2}{0}$$

$$\delta \to \infty$$

2.- RECTA DE MÍNIMOS CUADRADOS HORIZONTALES

Podría suceder que la variable que tiene error es X y no Y.

Los parámetros se determinan minimizando la suma de cuadrados de distancias horizontales.



Grafica 2.- Recta de Mínimos Cuadrados Horizontales

El resultado es el siguiente:

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

$$b = \frac{V(y)}{Cov(x, y)}$$

En este caso:

$$\delta = \frac{0}{\infty}$$

$$\delta = 0$$

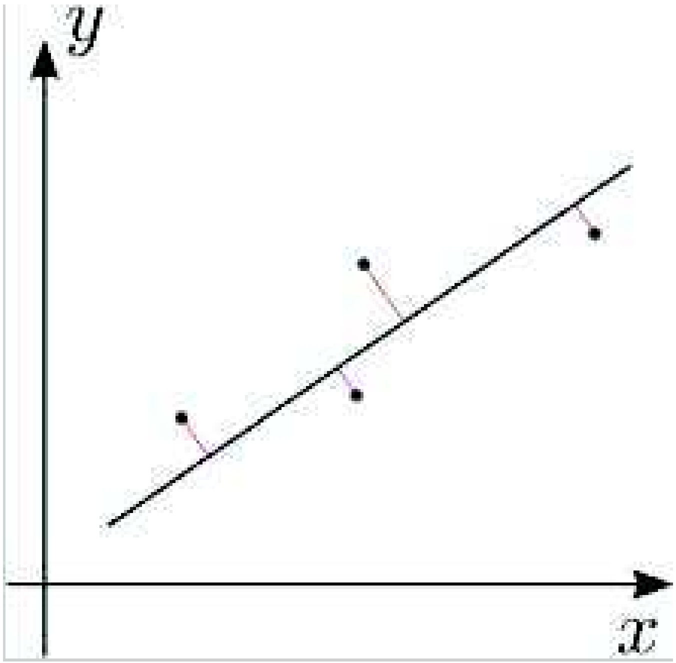
3.- RECTA DE MÍNIMOS CUADRADOS ORTOGONALES

Sin embargo si en algunas circunstancias se admiten errores en las dos variables, y además se considera que las varianzas en esos errores son iguales:

$$\sigma_x^2 = \sigma_y^2$$

$$\delta = 1$$

Las distancias desde cada punto a la recta no pueden ser verticales sino ortogonales.



Gráfica 3. Distancias ortogonales consideradas desde cada punto hasta la recta de regresión.

La expresión que define la distancia cuadrada desde un punto cualquiera i a la recta y=a +bx es:

$$d^2 = \frac{(bx_i - y_i + a)^2}{b^2 + 1}$$

Y la suma de distancias cuadradas ortogonales:

$$S = \frac{1}{b_2 + 1} \sum (bx_i - y_i + a)^2$$

Al realizar:

$$\frac{\delta S}{\delta a} = 0$$

Se obtiene:

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

Reemplazando este valor en S:

$$S = \frac{1}{b^2 + 1} \sum \left(b\left(x_i - \bar{x}\right) - \left(y_i - \bar{y}\right)\right)^2$$

$$S = n\left(\frac{b^2 V(x) - 2b.Cov(x, y) + V(y)}{b^2 + 1}\right)$$

Derivando ahora S con respecto a b e igualando a cero:

$$\frac{\delta S}{\delta b} = 0$$

Se obtiene:

$$b^2 Cov\,(x,y) + b\,(V\,(x) - V\,(y)) - Cov(x,y) = 0$$

Resolviendo la ecuación de segundo grado:

$$b = \frac{-\,(V\,(x) - V\,(y)) \pm \sqrt{(V\,(x) - V\,(y))^2 + 4(Cov\,(x,y))^2}}{2.Cov\,(x,y)}$$

Con el criterio de la segunda derivada, puede demostrarse que el valor de la pendiente para la recta de los mínimos cuadrados ortogonales es:

$$b = \frac{-\,(V\,(x) - V\,(y)) + \sqrt{(V\,(x) - V\,(y))^2 + 4(Cov\,(x,y))^2}}{2.Cov\,(x,y)}$$

4. RECTA DE MINIMOS CUADRADOS CON ORIENTACIÓN DE DISTANCIAS DIFERENTES.

Cuando δ toma valores diferentes a los estudiados, es obvio que la pendiente de los segmentos sobre los cuales se mide las distancias a la recta de regresión no es horizontal, vertical ni perpendicular a dicha recta.

En el caso de la regresión ortogonal la pendiente de los segmentos necesariamente es:

$$m_0 = \frac{2Cov\,(x,y)}{V\,(x) - V\,(y) - \sqrt{(V\,(x) - V\,(y))^2 + 4\,(Cov\,(x,y))^2}}$$

Pero para cualquier valor de δ se propone el siguiente valor de pendiente para el segmento:

$$m = \frac{2.\delta.Cov\,(x,y)}{(V\,(x) - V\,(y))^2 - \sqrt{(V\,(x) - V\,(y))^2 + 4\,(Cov\,(x,y))^2}}$$

O también:

$$m = \delta.m_0$$

Nótese que cuando:

$\delta \to \infty \qquad m \to \infty$

$\delta = 0 \qquad m = 0$

$\delta = 1 \qquad m = m_0$

Con la pendiente m de los segmentos entre los puntos y la recta, se procede a calcular la recta de mínimos cuadrados.

La recta de mínimos cuadrados tiene la forma:

$$y = a + bx$$

Y la recta que contiene el segmento sobre el cual se miden las distancias a partir del punto P_i de coordenadas (x_i , y_i) tiene la siguiente forma:

$$y - y_i = m\,(x - x_i)$$

La solución de este sistema entrega la siguiente respuesta para x:

$$x = \frac{-mx_i - a + y_i}{b - m}$$

Pero para todas las rectas de mínimos cuadrados es demostrable que:

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

Reemplazando en el valor de x:

$$x = \frac{-mx_i + y_i - \bar{y} + b\bar{x}}{b - m}$$

$$x - x_i = \frac{(y_i - \bar{y}) - b\,(x_i - \bar{x})}{b - m}$$

$$y - y_i = m\left(\frac{(y_i - \bar{y}) - b\,(x_i - \bar{x})}{b - m}\right)$$

La distancia cuadrada del punto i a la recta de regresión será:

$$d^2 = (x - x_i)^2 + (y - y_i)^2$$

En otros términos:

$$d^2 = \left(1 + m^2\right)\left(\frac{\left((y_i - \bar{y})^2 - b\,(x_i - \bar{x})\right)^2}{(b - m)^2}\right)$$

La Suma de todas las distancias cuadradas tiene la forma:

$$S = \frac{n(1+m^2)}{(b-m)^2}\left(b^2V(x) - 2b.Cov(x,y) + V(y)\right)$$

Si se practica:

$$\frac{\delta S}{\delta b} = 0$$

Se obtiene:

$$b = \frac{m.Cov(x,y) - V(y)}{m.V(x) - Cov(x,y)}$$

$$b = \frac{\delta.m_0.Cov(x,y) - V(y)}{\delta.m_0.V(x) - Cov(x,y)}$$

La aplicación de la segunda derivada a la expresión de suma de cuadrados de distancias, permitirá afirmar dicha pendiente (b) corresponde al valor mínimo.

Un rápido chequeo de la expresión permite concluir que cuando

$\delta = 0$

$$b = \frac{V(y)}{Cov(x,y)}$$

Y cuando

$\delta \to \infty$

$$b = \frac{Cov(x,y)}{V(x)}$$

También puede comprobarse que cuando

$\delta = 1$

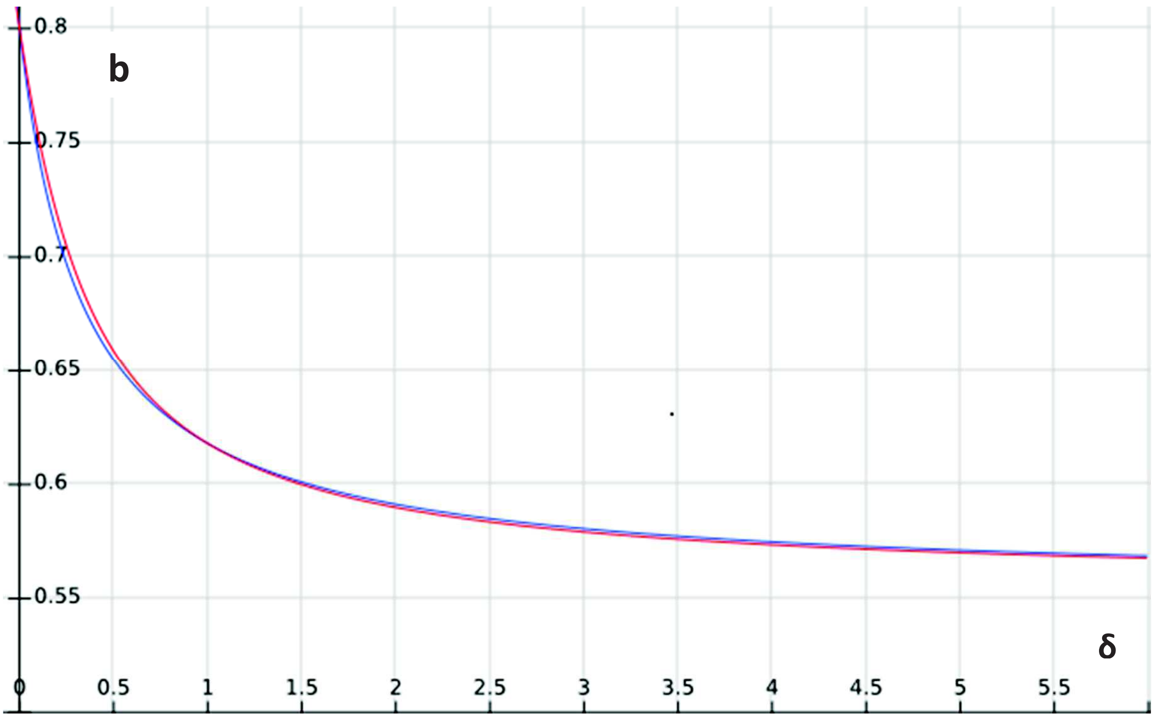
El valor de la pendiente (b) es justamente la correspondiente a la Recta de Mínimos Cuadrados Ortogonales.

WILLIAM EDWARDS DEMING, fue un estadístico norteamericano (1900 -1993) propuso la regresión que lleva su nombre y que considera la relación de varianzas δ.

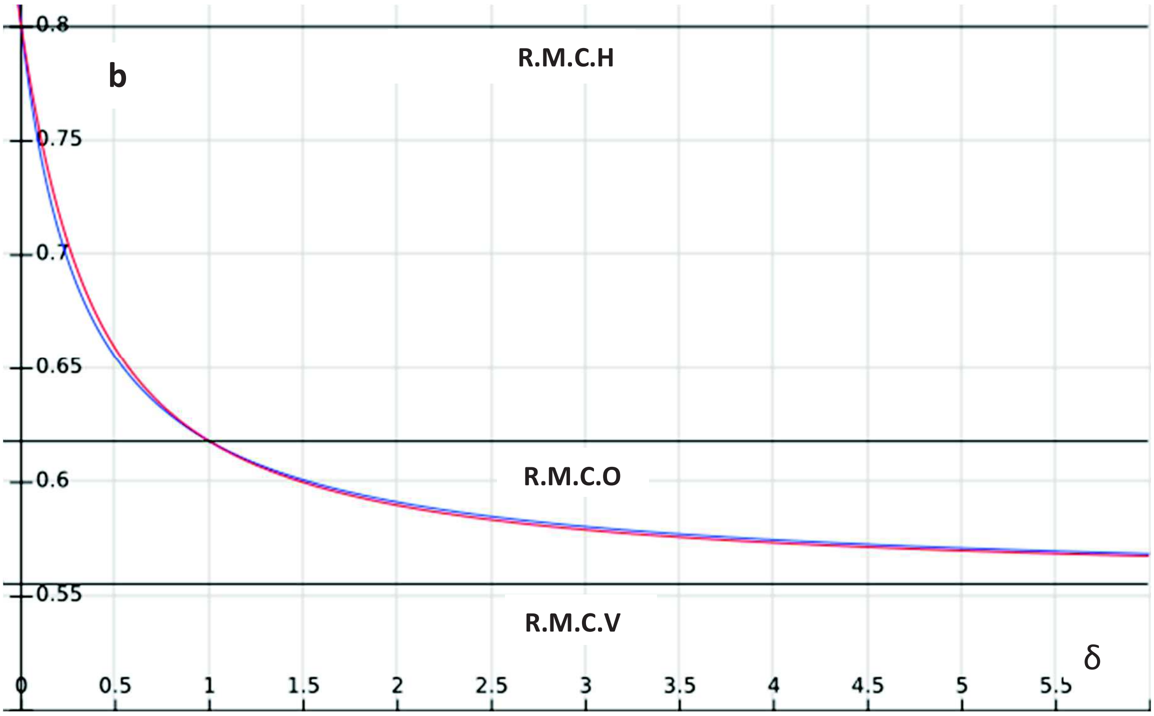
La expresión que   estableció para calcular el parámetro de pendiente fue la siguiente:

$$b = \frac{-\left(\delta.V\left(x\right) - V\left(y\right)\right) + \sqrt{\left(\delta.V\left(x\right) - V\left(y\right)\right)^2 + 4.\delta.\left(Cov\left(x,y\right)\right)^2}}{2.Cov\left(x,y\right)}$$

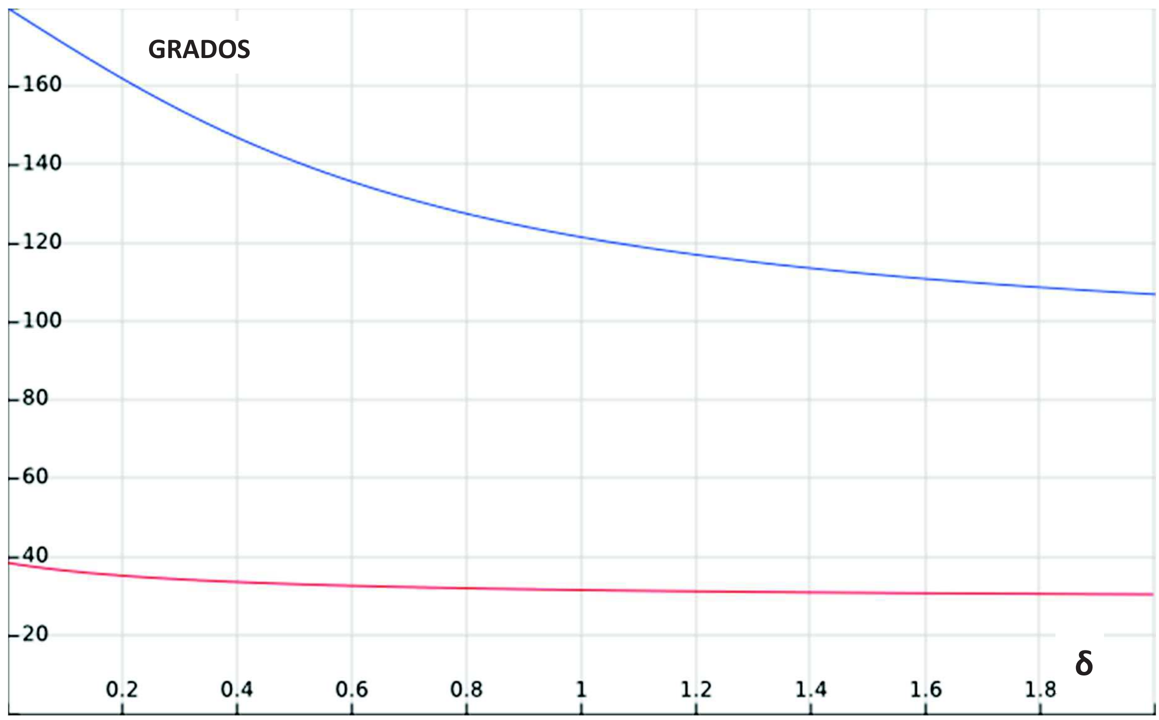
A continuación, y con un ejemplo, se comparan gráficamente los resultados obtenidos por el autor del presente artículo y los logrados aplicando la fórmula de Deming



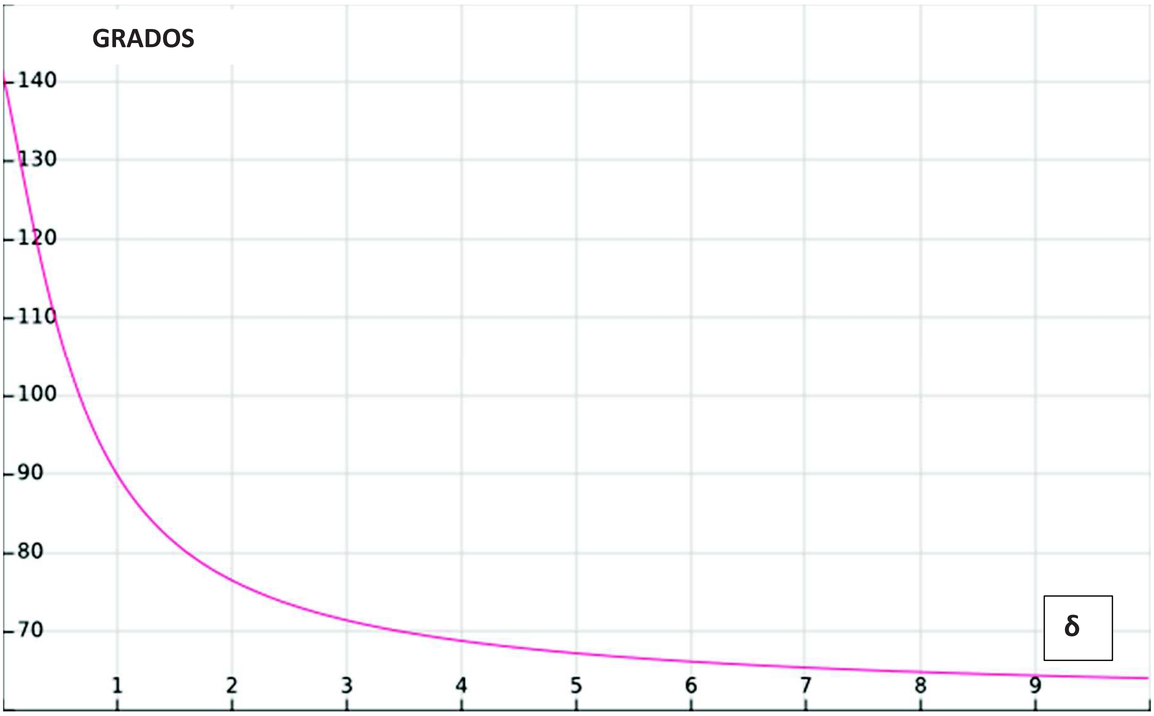
Grafica 4.   Sea V(x)=9, V(y)=4,   Cov(x,y) = 5   Se obtienen las pendientes de la Recta de Regresión de Mínimos Cuadrados: Línea azul: Regresión del autor. Línea Roja: Regresión de Deming.



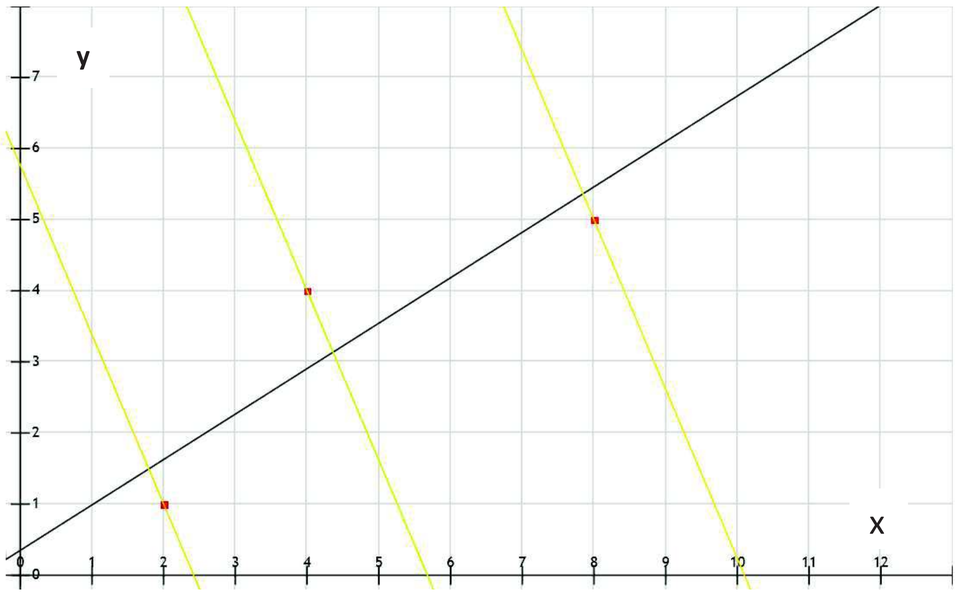
Gráfica 5: Muestra la Regresión del Autor y la Regresión de Deming, marcándose con líneas horizontales las pendientes para la Recta de Mínimos Cuadrados Horizontales (superior), Recta de Mínimos Cuadrados Ortogonales (línea media) y Recta de Mínimos Cuadrados Ortogonales, para los datos: V(x)=9, V(y)=4, Cov(x,y) =5



Gráfica 6. Para los datos: V(x)=9, V(y)=4, Cov(x,y) =5, se aprecia con línea roja los ángulos formados por l la recta de regresión de mínimos cuadrados con la horizontal para los valores de δ, en línea azul los ángulos que forman con la horizontal, los segmentos de recta que se dirigen desde los puntos hasta a recta.



Gráfica 7: Para los datos V(x)=9, V(y)=4, Cov(x,y) =5 se observa la diferencia de ángulo entre la Recta de Regresión de Mínimos Cuadrados y los segmentos dirigidos desde el conjunto de puntos a dicha recta para distintos valores de δ.



Gráfica 8. Para el conjunto de puntos: (2;1), (4;4), (8,5) se grafica la Recta de Mínimos Cuadrados y la dirección de los segmentos de que se dirigen desde los puntos, para la condición δ=2 . Ecuación de la recta: y= 0,6391267 x + 0,35074206. Pendiente de los segmentos m = -2,38384429.

6. CONCLUSIONES

1.- Se expuso la importancia de la relación de varianzas de errores en la Regresión Lineal Bivariada.
2.- Se indicaron los métodos para determinar parámetros usando rectas de mínimos cuadrados
3.- Se entregaron los resultados analíticos que definen los parámetros bajo todas las circunstancias
4.- Se puso a consideración un método novedoso general que se puede aplicar a todos los casos para determinar parámetros.
5.- Se cotejaron los resultados del método inédito con los entregados por la Regresión de Deming.

7. Referencias bibliográficas.

Canavos G.C. Probabilidad y Estadística. Editorial Mc Graw Hill. Mexico.

Adcock R.J. A Problem in Least Squares. Annals of Mathematics. Vol 5 No 2 . 1878.

Abdus Sattar Gazdar. Distance from a point to a plane. The College Mathematics Journal. Vol 23 No.5

J. Vicente de Julián Ortiz. Lionello Pioglani. Emili Besalú. Two-Variable Linear Regression: Modeling with Orthogonal Least-Squares Analysis. Journal of Chemical Education. (ACS Publications 2010)

Deming W. E. *Statistical adjustment of data*. Wiley, NY (Dover Publications edition, 1985).

Walpole, Myers, Myers, Ye, Probabilidad y Estadística, Octava Edición, Pearson, Pretnice Hall, 2007.

Lehmann, Geometría Analítica. Edit. Limusa, 1989