ESTUDIO COMPARATIVO GEOMÉTRICO Y MATEMÁTICO DE LA RECTA DE REGRESIÓN DE DEMING

ÁLVARO DE JESÚS VILLOTA VIVEROS

UNIVERSIDAD DE NARIÑO
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
MAESTRÍA EN ESTADÍSTICA APLICADA

SAN JUAN DE PASTO, OCTUBRE 31 DE 2023

ESTUDIO COMPARATIVO GEOMÉTRICO Y MATEMÁTICO DE LA RECTA DE REGRESIÓN DE DEMING

ÁLVARO DE JESÚS VILLOTA VIVEROS

Trabajo de Grado como requisito parcial para optar al título

de Magister en Estadística Aplicada

Director de Tesis

ARSENIO HIDALGO TROYA

Mg. en Estadística

UNIVERSIDAD DE NARIÑO
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA
MAESTRÍA EN ESTADÍSTICA APLICADA
SAN JUAN DE PASTO, OCTUBRE 31 DE 2023

NOTA DE RESPONSABILIDAD

Las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva del autor (Artículo 1ro del Acuerdo No. 324 de octubre 11 de 1966 emanado del Honorable Consejo Directivo de la Universidad de Nariño).

Nota de aceptación

Jurado 1

Jurado 2

Jurado 3

San Juan de Pasto, octubre 31 de 2023

AGRADECIMIENTOS

A mi viejo amigo y asesor de trabajo de grado, Mg. Arsenio Hidalgo Troya por su valiosísima orientación, dedicación, apoyo incondicional y confianza.

Al Mg. Hernán Abdón García, Coordinador de maestría, gestor de postgrados de estadística en nuestra región.

DEDICATORIA

A mis amados y siempre recordados padres Guillermo Enrique Villota Ordoñez e Inés Viveros de Villota por su infinita dedicación y ejemplo, en los años lejanos de mi infancia, su preocupación primordial siempre fue la formación académica de sus hijos. Pienso que no los he defraudado.

A todos mis profesores, a mis compañeros de aula y estudiantes en todas las etapas de mi vida.

A Robert Jackson Adcok (1826-1895) y a William Edward Deming (1900-1993), matemáticos y estadísticos, ilustres personajes injustamente olvidados, distantes en el tiempo y el espacio, a quienes trato de conocer en su faceta científica, gracias a la comunicación mágica y trascendente de los libros, compartiendo sus inquietudes. Espero con este modesto aporte, haber interpretado y ampliado, aunque parcial y mínimamente su trabajo.

RESUMEN

Este trabajo considera la regresión de Deming como la familia general de las rectas de regresión simple de mínimos cuadrados, que contempla errores en las dos variables e incluye en su desarrollo la razón entre las varianzas de los errores de la variables dependiente e independiente. Subvalorar esa razón de varianzas y escoger arbitrariamente modalidades de mínimos cuadrados ordinarios o mínimos cuadrados ortogonales pueden causar diferencias notables en cuanto a la estimación de los parámetros, suma de cuadrados de distancias, ángulos, coeficiente de determinación R^2 y métricas en general.

El presente estudio inédito, teórico fundamentado en la estadística, la geometría analítica y el cálculo diferencial, realiza un estudio detallado de la recta de regresión de Deming y define con minuciosidad las diferencias geométricas y matemáticas existentes entre todas las variantes posibles. Se realiza énfasis especial en el coeficiente de determinación y en la parte final se practican estudios teóricos fundamentados en inferencia estadística. Los resultados se expresan mediante ecuaciones generalizadas.

Los desarrollos logrados permitirán al investigador escoger en cada caso particular, con criterios teóricos, la recta de regresión de mínimos cuadrados más adecuada de acuerdo a las condiciones del problema abordado.

Palabras clave: Regresión de Deming, razón de varianza de errores, pendiente de errores, diferencias geométricas, diferencias de parámetros, coeficientes de determinación, intervalos de confianza, métricas.

ABSTRACT

This paper considers the Deming regression as the general family of least-squares simple regression lines that contemplates errors in the two variables and includes in its development the ratio between the variances of the errors of the dependent and independent variables. Underestimating this ratio of variances and arbitrarily choosing modalities of ordinary least squares or orthogonal least squares can cause notable differences in terms of parameter estimation, sum of squares of distances, angles, coefficient of determination R^2 and metrics in general.

This unpublished theoretical study based on statistics, analytical geometry, and differential calculus performs a detailed study of the Deming regression line and meticulously defines the geometric and mathematical differences between all possible variants. Special emphasis is placed on the coefficient of determination and in the final part theoretical studies based on statistical inference are practiced. The results are expressed through generalized equations.

The developments achieved will allow the researcher to choose in each particular case, with theoretical criteria, the most appropriate least squares regression line according to the conditions of the problem addressed.

Keywords: Deming regression, error variance ratio, error slope, geometric differences, parameter differences, coefficients of determination, confidence intervals, metrics.

Tabla de Contenido:

LISTA DE CUADROS

Este trabajo no contiene cuadros.

LISTA DE TABLAS

Este trabajo no contiene tablas.

LISTA DE FIGURAS

LISTA DE ANEXOS

Este trabajo no contiene anexos.

GLOSARIO

n Número de puntos (observaciones).
a Corte en el eje y de la recta de mínimos cuadrados.
b Pendiente recta de mínimos cuadrados.
S Suma de cuadrados de distancias.
m Pendiente de los errores.
b_D Pendiente recta de regresión de Deming.
b_v Pendiente de recta de mínimos cuadrados ordinarios.
b_h Pendiente de recta de mínimos cuadrados horizontales.
b_o Pendiente de recta de mínimos cuadrados ortogonales.
ρ Coeficiente de correlación de Pearson.
R^2 Coeficiente de determinación.
θ Ángulo entre dos rectas de regresión.
p Razón entre pendientes de rectas de regresiones de Deming y Ortogonal.

ε_{iy} Error i de la variable "y".
ε_{ix} Error i de la variable "x".

$\sigma_{\varepsilon y}^2$ Varianza de errores para la variable "y".
$\sigma_{\varepsilon x}^2$ Varianza de errores para la variable "x".

SCT Suma de cuadrados totales
SCM Suma de cuadrados del modelo
SCE Suma de cuadrados de errores

ECM Error cuadrado medio.
$RECM$ Raíz error cuadrado medio.
EAM Error absoluto medio.

$\delta = \frac{\sigma_{\varepsilon y}^2}{\sigma_{\varepsilon x}^2}$ Razón de varianzas de errores.

$$S_{yy} = \sum (y_i - \bar{y})^2$$

$$S_{xx} = \sum (x_i - \bar{x})^2$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$k = \frac{S_{yy}}{S_{yy}}$$

INTRODUCCIÓN

En regresión simple, es frecuente el empleo de la recta de mínimos cuadrados ordinarios, cuyos parámetros de corte con el eje “Y” y pendiente se estiman recurriendo a la mínima suma de los cuadrados de los errores medidos verticalmente.

Adcock en Inglaterra en 1878 planteó la determinación de los parámetros de la recta de mínimos cuadrados recurriendo a errores ortogonales a la recta de regresión, argumentando que las verdaderas distancias de un conjunto de puntos a una recta debían medirse bajo esa dirección considerando los principios elementales de la geometría euclidiana.

Deming en Norteamérica en 1943, propuso la recta de regresión en la que consideró el cociente entre las varianzas de los errores de las dos variables, innovando los criterios que definieron los dos modelos anteriormente referidos.

Este proyecto profundiza en la recta de regresión de Deming y de manera original encuentra las direcciones de los errores que serán oblicuos. De esta manera la regresión de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales solo son casos particulares de la regresión de Deming.

Disponiendo de toda la familia de rectas de regresión de Deming, se propone como problema realizar un estudio comparativo matemático y geométrico, extenso y muy detallado de las diferencias existentes entre ellas respecto a los parámetros, la suma de cuadrados de distancias, diferencias angulares y coeficientes de determinación en función de varianzas, covarianzas y coeficientes de correlación.

Las publicaciones existentes no tratan a profundidad el tema propuesto en el presente trabajo, lo hacen tangencialmente y las diferencias y comparaciones se plantean observando resultados de aplicaciones prácticas, pero no de manera formal, geométrica, matemática y generalizada.

Este trabajo realiza una comparación analítica contemplando todos los escenarios posibles y los resultados serán amplios y teóricos acompañados de gráficas ilustrativas de manera que el estudio sea inédito y útil.

Toda la diferencia se presenta como ecuaciones en función de las variables involucradas y considerando todas las circunstancias posibles. Por la naturaleza teórica del estudio, surgen a lo largo de él interesantes propiedades que se explican y exponen con detalle.

Es interesante advertir que, si no se define adecuadamente la razón de varianzas de errores entre la variable independiente y la variable dependiente, las diferencias angulares pueden ser extremas, incluso llegar al valor $\pi/2$.

Este trabajo incluye el estudio detallado del coeficiente de determinación R^2 de la regresión de Deming y demuestra matemáticamente que el valor es siempre mayor cuando el cociente entre las varianzas de los errores para las dos variables es 1, es decir cuando los mínimos cuadrados son ortogonales.

En el último capítulo, se realiza un análisis de estadística inferencial fundamentado en los intervalos de confianza de la regresión ortogonal tendientes a definir la pertinencia del empleo de la recta de regresión de Deming de acuerdo a las características estadísticas de las variables involucradas.

PRELIMINARES

1.1 Planteamiento del problema

Cuando se elige la regresión lineal simple como modelo de regresión, en general se recurre al método de los mínimos cuadrados ordinarios, asumiendo que la variable “x” no posee errores, bajo ese supuesto se determinan parámetros de corte y de pendiente, desconociendo que si este supuesto es falso los resultados pueden sufrir desfases enormes, especialmente la pendiente.

Se ha propuesto la regresión basada en los mínimos cuadrados ortogonales como solución alterna, en este estudio se demostrará que la suma de cuadrados de los errores medidos en esa dirección es siempre menor a la suma de cuadrados ordinarios (verticales), lo anterior es compatible con los estudios experimentales que concluyen que el método ortogonal ofrece mejor ajuste.

Existe la regresión de Deming que puede considerarse como una regresión lineal de mínimos cuadrados general que incluye entre otras, a las dos modalidades mencionadas, se trata de un método que considera la posible existencia en los errores en las dos variables e incluye en su formulación la razón (δ) entre varianza de error de variable “y” y varianza de variable “x”. La causa por la que no es usual el empleo de la regresión de Deming es la dificultad para estimar el valor (δ).

Es necesario realizar un estudio comparativo en las diferencias de parámetros, suma de cuadrados de distancias, ángulos, coeficientes de determinación y métricas, entre otros aspectos, cuando se usa regresión de Deming, mínimos cuadrados ordinarios, mínimos cuadrados ortogonales u otras variantes posibles, para evidenciar las falencias en las que puede incurrirse cuando se escogen indistintamente una modalidad sin estudio previo.

El coeficiente de determinación R^2 es una medida muy importante de la justificación de la varianza atribuible al modelo planteado. Los tratados existentes no abordan formalmente las variaciones que puede experimentar dicho coeficiente cuando se escogen sin mayor justificación la variante de regresión bivariada de mínimos cuadrados.

A diferencia de los escasos estudios existentes, en los cuales las bondades de los tipos de regresión escogidos y la comparación con otros se establecen al enfrentar problemas específicos y particulares, este trabajo determina cotejos generalizados y globales que consideran todas las circunstancias y se fundamenta en demostraciones matemáticas.

No existen trabajos sobre regresiones lineales que apliquen de manera exhaustiva y profunda la conjunción entre geometría analítica y cálculo y que definan con exactitud las diferencias entre los tipos de regresiones lineales estimados mediante mínimos cuadrados, ameritándose este abordaje.

Es necesario realizar un estudio profundo, detallado, comparativo y matemático sobre los resultados del coeficiente de determinación R^2 cuando se elige regresión lineal.

Siendo la regresión de Deming una solución general que incluye una familia geométrica, es pertinente establecer propiedades que no han sido determinadas con anterioridad.

Finalmente, surge una inquietud que debe plantearse y consiste en la pertinencia de emplear regresión de Deming, por la dificultad que implica la estimación de la razón de varianzas de errores, para ciertos casos específicos en los cuales se puede optar por la regresión ortogonal, sin que existen diferencias significativas.

Este trabajo pretende solucionar los problemas expuestos y constituirse en una herramienta que de manera analítica y general compare las rectas de las regresiones con dos variables determinadas a partir del método de mínimos cuadrados.

1.2 Objetivos

1.2.1 Objetivo general

Realizar un estudio geométrico y matemático comparativo de las variantes de la recta de regresión de Deming.

1.2.2 Objetivos específicos

Encontrar la ecuación general de las rectas de mínimos cuadrados cuando las distancias de los puntos a la recta se miden en una dirección con pendiente m.

Determinar la suma de cuadrados totales, suma de cuadrado de errores y suma de cuadrados del modelo en una recta de mínimos cuadrados con errores medidos en una orientación de pendiente m.

Demostrar que la suma vectorial de errores medidos con una pendiente m en una recta determinada por el método de los mínimos cuadrados es cero.

Definir la pendiente de los errores, para aplicar el método de mínimos cuadrados, a partir de la ecuación de la recta de regresión propuesta por Deming.

Encontrar una expresión general para el coeficiente de determinación de la recta de regresión de Deming y realizar un estudio profundo sobre las diferencias en todas sus variantes.

Realizar una comparación exhaustiva entre el parámetro de pendiente de la recta de regresión de Deming en su expresión general y las rectas de regresión de mínimos cuadrados ordinarios y mínimos cuadrados horizontales.

Realizar un estudio sobre el cambio de pendiente de la recta de regresión de Deming, con el incremento de la razón de varianzas de los errores.

Determinar las diferencias angulares entre la recta de regresión de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios y adelantar un estudio detallado sobre ese aspecto.

Determinar las diferencias angulares entre la recta de regresión de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales y adelantar un estudio detallado sobre ese aspecto.

Desarrollar un estudio completo y exhaustivo que determine las diferencias entre el coeficiente de determinación entre la recta de regresión de mínimos cuadrados ordinarios y la recta de mínimos cuadrados ortogonales.

Realizar estudio detallado sobre la razón entre la pendiente de la recta de regresión de Deming en general y la pendiente de la recta de regresión ortogonal.

Realizar un estudio de pertinencia de empleo de recta de regresión ortogonal sobre recta de regresión de Deming, fundamentado en los valores de la varianza de las variables, la covarianza de las mismas, el coeficiente de correlación de Pearson y la razón de varianzas de errores.

Determinar las rectas de regresión de Deming contenidas en el intervalo de confianza de las pendientes de las rectas de regresión ortogonal.

Determinar los coeficientes mínimos de Pearson para los cuales las pendientes de recta de regresión de Deming con cociente de varianzas de errores $\delta_2 < \delta_o < \delta_1$, están contenidas en el intervalo de confianza $(1-\alpha)100$ de la recta de regresión ortogonal.

Encontrar una manera de chequear que la pendiente de la recta de regresión de Deming, está incluida en un intervalo de confianza de la regresión ortogonal, asumiendo un intervalo de cociente de varianzas de los errores de las dos variables.

Encontrar las expresiones de las métricas de las rectas notables pertenecientes a la regresión de Deming.

1.3. Diseño Metodológico

La modalidad del presente estudio es "Investigación Básica" porque amplía el conocimiento teórico que se dispone sobre las características y propiedades geométricas y matemáticas de la recta de regresión de Deming. Se recurre a la deducción geométrica y matemática para lograr expresiones generales y no a la experimentación ni a la aplicación de casos particulares.

En cada uno de los objetivos específicos se ha diseñado el método para lograr los resultados.

Para obtener la ecuación general de la recta de regresión de mínimos cuadrados cuando las distancias se miden en una dirección con pendiente m, se recurre a los rudimentos de la geometría analítica, específicamente a las ecuaciones de rectas e intersecciones entre ellas, además se emplea el cálculo diferencial y definiciones de momentos, específicamente varianza y covarianza.

Para encontrar suma de cuadrados totales, del modelo y de los errores, medidos en una recta de mínimos cuadrados con una pendiente m, se emplean los parámetros de pendiente y corte obtenidos en la expresión de suma de cuadrados de distancias. La aplicación de la geometría analítica permite encontrar SCE (suma de cuadrado de errores), SCM (suma de cuadrados del modelo) y SCT (suma de cuadrados totales). Con el empleo de algebra de sumatorias y de las definiciones de varianza y covarianza, se demuestra de manera inédita que en estos casos también es válida la expresión SCT=SCM+SCE. Se aplica el álgebra vectorial para demostrar que la suma de los errores para todas las rectas de mínimos cuadrados es cero.

Una vez obtenidos los parámetros de una recta de mínimos cuadrados en los cuales se han medido las distancias con una pendiente m, se logra una relación entre m y el parámetro de pendiente. Si se reemplaza el valor del parámetro de pendiente de la recta de regresión de Deming, en la ecuación respectiva, se logra definir el valor de la pendiente en los errores a partir de las cuales al aplicar el método de los mínimos cuadrados se determina el mencionado parámetro de la recta.

Con la teoría desarrollada, se encuentra la expresión general para obtener el coeficiente de determinación en la recta de regresión de Deming, una vez superada esta etapa, se procede a realizar un estudio detallado sobre las diferencias existentes entre dicho coeficiente en todas las variantes de dicha regresión.

Siendo un propósito del presente estudio la comparación, se considera la pendiente de la recta de regresión de Deming que incluye la recta de regresión ortogonal y se la compara de manera muy detallada con otras rectas notables concretamente la de mínimos cuadrados ordinarios y mínimos cuadrados horizontales. Lo anterior con aplicación de la estadística, la geometría analítica y el cálculo. Aparecen propiedades interesantes las cuales se estudian, demuestran y plantean.

Considerando la expresión de la pendiente de la recta de Regresión de Deming, las varianzas, la covarianza de las variables y el coeficiente de correlación de Pearson y aplicando el álgebra se realiza un estudio sobre el cambio de la pendiente con relación al incremento de la razón de varianzas de los errores de las variables.

Considerando que las rectas de regresión de Deming más notables son las de mínimos cuadrados ortogonales, mínimos cuadrados ordinarios y mínimos cuadrados horizontales, se empleará la geometría analítica, el cálculo y la estadística para determinar las expresiones que permitan calcular las diferencias angulares entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios, también entre mínimos cuadrados horizontales y mínimos cuadrados ortogonales. En los dos casos se determinan esas diferencias en función de la razón entre S_{yy} y S_{xx} que se llamará k y el coeficiente de correlación de Pearson. Se presentan propiedades geométricas y matemáticas interesantes las cuales se demuestran y exponen.

Con el propósito de adelantar una comparación completa entre el coeficiente de determinación de la recta de regresión de mínimos cuadrados ordinarios y la recta de regresión de mínimos ortogonales, se plantean todas las circunstancias posibles de S_{yy} , S_{xx} y coeficientes de correlación de Pearson y se determinan todas las expresiones resultantes que contengan los resultados del coeficiente de determinación, sus diferencias y valores máximos y mínimos, finalmente y de acuerdo a esos resultados se demuestra que el coeficiente estudiado es siempre mayor cuando se utiliza mínimos cuadrados ortogonales.

Para realizar el estudio comparativo de la razón entre la pendiente de la recta de la regresión de Deming y la pendiente de la recta de regresión ortogonal, se emplea el álgebra, el cálculo diferencial y los conceptos básicos de estadística. Se plantea la razón en función de k, el coeficiente de correlación de Pearson y la razón de varianzas de errores, establecida la razón de pendientes, se práctica un estudio detallado, encontrando valores extremos incluidos máximos y mínimos relativos. Luego sobre esos valores se amplía el estudio analizando el efecto del coeficiente de correlación de Pearson.

Para realizar un estudio de pertinencia de empleo de recta de regresión ortogonal sobre la recta de regresión de Deming y como introducción a estudio inferencial, se definen dos números reales (uno de ellos >1 y el otro < 1) que al multiplicarlos por la pendiente resultante de la regresión de mínimos cuadrados ortogonales formen un intervalo de pendientes. Fundamentándose en ese intervalo, se realizan deducciones geométrico matemáticas que definan los valores de razón de varianzas de errores (δ) pertenecientes y por lo tanto las rectas de regresión de Deming contenidas.

Para encontrar las rectas de la regresión de Deming cuyas pendientes están contenidas en el intervalo de confianza 100 $(1-\alpha)$ de la pendiente de la recta de regresión ortogonal, se recurre a las expresiones suministradas por Minitab, una vez logrado el intervalo de confianza, y en consonancia con lo expresado en el párrafo anterior se calculan los dos números reales que serán utilizados para obtener los valores de la razón de varianzas de errores y por lo tanto las rectas de regresión de Deming correspondientes.

En la parte final de inferencia, se dispone previamente del conocimiento de la razón de varianzas, $\boldsymbol{\delta}_2 < \boldsymbol{\delta_o} < \boldsymbol{\delta}_1$ sin precisar su valor (este intervalo se puede establecer mediante una prueba F), luego se establece el intervalo de confianza $100(1-\alpha)$ de la pendiente de la recta de mínimos cuadrados ortogonales, a partir de este último intervalo se definen dos números reales (ya descritos anteriormente), también se realizan razones entre los valores de pendiente de regresión de Deming y regresión ortogonal, usando los valores de cocientes de varianzas de errores extremos, esas razones, determinan otra pareja de números reales que al compararse con la primera pareja definen por medio de chequeo si la pendiente de la recta regresión de Deming, está contenida en el intervalo de confianza de la pendiente de la recta de regresión ortogonal.

Teniendo como fundamento, toda la teoría desarrollada, se deducen expresiones para definir métricas en las rectas de regresión de Deming.

1.4 Marco teórico

Distancia geométrica entre dos puntos.

[1] Lehmann, Ch (1989) en su aparte correspondiente a geometría analítica deduce la expresión para determinar la distancia entre dos puntos y que corresponde a la raíz cuadrada de la suma de los cuadrados de las diferencias de las coordenadas de dichos puntos.

Ecuaciones de líneas rectas en el plano

[2] Wexler, Ch (1977) se determinan las ecuaciones de líneas rectas y sus intersecciones en función de las coordenadas de dos puntos o considerando las coordenadas de un solo punto y su pendiente.

Distancia ortogonal de un punto a una recta

[2] Wexler, Ch (1977) se demuestra que en el plano, la distancia ortogonal de un punto $P_i(x_i, y_i)$ a una recta $Ax + By + C = 0$ es igual al valor absoluto de $Ax_i + By_i + C,$ dividido ente la raíz cuadrada de $A^2 + B^2$.

Conceptos básicos de estadística, geometría analítica y cálculo diferencial

Para el desarrollo del presente estudio en todas sus etapas, se requiere fundamentos sólidos de cálculo diferencial aplicado a la estadística, este sustento lo ofrecen muchos textos de matemática básica. En especial se requiere el manejo adecuado de los criterios de primera y segunda derivada para definir puntos críticos. El sustento de los rudimentos básicos puede encontrarse en [3] Apostol T.M. (2006)

Momentos Estadísticos

La definición de momentos será fundamental al determinar esperanza matemática de cuadrados de distancias ortogonales y varianza de cuadrados de distancias, se recurrirá a bases de la estadística, se consulta a [4] Canavos G.C. que define los momentos de variables aleatorias como los valores esperados de ciertas funciones de dicha variable.

Coeficiente de Correlación de Pearson

En el desarrollo del proyecto es de suma importancia la interpretación estadística, matemática y geométrica del coeficiente de correlación de Pearson, su valor implica información necesaria para definir las diferencias entre las rectas de regresión de Deming: la recta de mínimos cuadrados ordinarios, la recta de los mínimos cuadrados ortogonales y otras. En general el valor del coeficiente de correlación de Pearson es necesario en el estudio de todos los temas incluidos en el proyecto. En [5] Keney, J.F. and Keeping E. S, se encuentra no solo la definición del coeficiente de correlación de Pearson como cociente entre la covarianza de las dos variables y el producto de las desviaciones estándar sino también toda la interpretación geométrica y estadística.

Recta de Mínimos Cuadrados Ordinarios.

En [6] Walpole, Myers, Ye (2007) se deducen y describen los parámetros de la recta de mínimos cuadrados ordinarios es decir aquellos que minimiza la suma de los cuadrados de las distancias verticales a la recta estimada. Siendo la recta y= a+bx, el parámetro b resultante es igual al cociente entre la covarianza de las dos variables sobre la varianza de x. El valor de a, simplemente es igual a la media de los valores de y menos el producto de b por la media de los valores de x.

Regresión de Deming

En [7] Deming W. E. (1943), El estadístico Richard Edwards Deming propuso la regresión lineal con dos variables que lleva su nombre, consideró errores para la variable "x" y la variable "y" e incluyó en la deducción de los parámetros la razón de varianzas de los errores de las variables. Cuando la razón de varianzas entre las variables es 1, la regresión de Deming se transforma en la recta de mínimos cuadrados ortogonales que en el pasado ya había sido abordada por [8] Adcock R. J. (1878) y [9] Coolidge J.L (1913).

Modelo de la Regresión de Deming

El modelo de la regresión de Deming es el siguiente:

$$y_i = \hat{y}_i + \varepsilon_{iy}$$

$$x_i = \hat{x}_i + \varepsilon_{ix}$$

Supuestos:

$$\varepsilon_{iy} \sim N(0, \sigma^2_{\varepsilon y})$$

$$\sum \varepsilon_{iy} = 0$$

Los errores en y son independientes.

$$\varepsilon_{ix} \sim N(0, \sigma^2_{\varepsilon x})$$

$$\sum \varepsilon_{ix} = 0$$

Los errores en x son independientes.

Parámetros en la regresión de Deming.

Definición:

Defínase δ a la razón de varianzas de los errores de la variable "y" y la variable "x",

$$\delta = \frac{\sigma_{\varepsilon y}^2}{\sigma_{\varepsilon x}^2}$$

También se establecen las siguientes definiciones:

S_{yy}: Suma de los cuadrados de las diferencias entre los valores observados de “y” y su media.

S_{xx}: Suma de los cuadrados de las diferencias entre los valores observados de “x” y su media

S_{xy}: Suma de los productos de la diferencia entre el valor observado de “x” y su media y la diferencia entre el valor observado de “y” por su media.

$$S_{yy} = \sum (y_i - \bar{y})^2$$

$$S_{xx} = \sum (x_i - \bar{x})^2$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

Los parámetros se obtienen a partir del siguiente procedimiento:

Definición:

Sea S, la suma de los cuadrados de los errores estandarizados:

$$S = \sum \left(\frac{\varepsilon_{iy}^2}{\sigma_{\varepsilon y}^2} + \frac{\varepsilon_{ix}^2}{\sigma_{\varepsilon x}^2} \right)$$

$$S = \frac{1}{\sigma_{\varepsilon y}^2} \sum \left(\varepsilon_{iy}^2 + \delta \varepsilon_{ix}^2 \right)$$

Si la recta de regresión de Deming es:

$$y = a + bx$$

La suma de los cuadrados de los errores estandarizados se puede expresar de la siguiente manera:

$$S = \frac{1}{\sigma_{\varepsilon y}^2} \sum \left((y_i - a - bx)^2 + \delta (x_i - x)^2 \right)$$

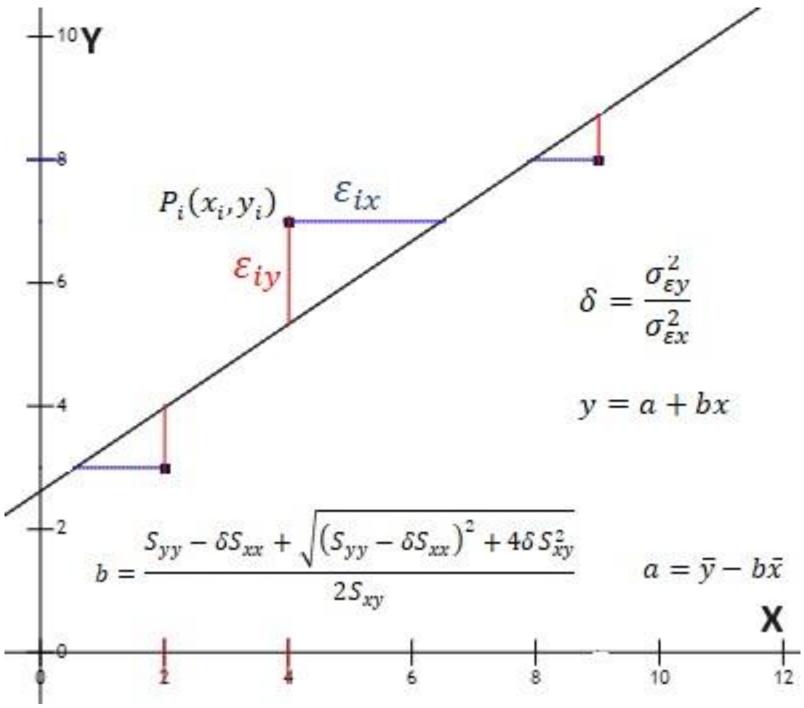
Se deriva con respecto a los dos parámetros y se iguala a cero:

En [10] Jensen A. Ch (2007) se presenta el desarrollo completo de la derivación, obteniéndose los siguientes parámetros:

$$b = \frac{S_{yy} - \delta S_{xx} + \sqrt{\left(S_{yy} - \delta S_{xx}\right)^2 + 4\delta S_{xy}^2}}{2S_{xy}}$$

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

Gráfica ilustrativa de la recta de regresión de Deming.



Recta de Regresión de Deming obtenida a partir de los puntos (2;3), (4;7) y (9,8). Se consideran errores en las dos variables. $\delta = 1/2$ (significa que la varianza de los errores en y, es la mitad de la varianza de los errores en x). Los parámetros respectivos son: a= 2.61726511, b= 0,67654698.

Intervalo de confianza para el parámetro de pendiente de la regresión ortogonal

Varianza de la pendiente de la recta de regresión ortogonal.

La varianza de la pendiente de la recta de la regresión ortogonal b_o es decir cuando $\delta = 1,$ se adecúa de acuerdo a las expresiones que entrega Minitab Statistical Software [11]

Minitab emplea una singular nomenclatura:

$$V(b_o) = \frac{\hat{\sigma}_{xx} S_{vv} + \sigma_u^2 S_{vv} - b_o^2 \sigma_u^4}{(n-1)\hat{\sigma}_{xx}^2}$$

Donde:

$$\hat{\sigma}_{xx} = \frac{S_{xx}\left(\sqrt{(k-1)^2 + 4k\rho^2} - (k-1)\right)}{2(n-1)}$$

$$\sigma_u^2 = \frac{S_{xx}\left(k + 1 - \sqrt{(k-1)^2 + 4k\rho^2}\right)}{2(n-1)}$$

$$S_{vv} = \frac{(n-1)(1+b_o^2)\sigma_u^2}{n-2}$$

Y el intervalo de confianza para la pendiente en regresión ortogonal es:

$$b_o \pm Z_{1-\alpha/2}\sqrt{V(b_0)}$$

Funciones de Densidad de Probabilidad y de Distribución de Probabilidad.

Para realizar inferencia es necesario conocer el significado y aplicación de la Función de Densidad de Probabilidad, [12] Zylberberg A.D. (2006) realiza una amplia explicación del uso de esta función, en especial de la normal estándar.

En el mismo texto Zylberberg A.D. se refiere a la Función de Distribución de Probabilidad.

Los conceptos de función de densidad de probabilidad y de función de distribución de normal estándar se emplean en este estudio en los apartes correspondientes a inferencia estadística en la determinación de intervalos de confianza

Estudios teóricos geométricos y matemáticos comparativos previos relacionados con regresiones bivariadas fundamentadas en mínimos cuadrados.

[13] Villota A, Hidalgo A. (2018) realizaron un estudio previo comparativo fundamentado en las rectas de regresión bivariada de mínimos cuadrados ortogonales, En el presente trabajo desarrolla nuevas y más completas deducciones considerando la familia de las rectas de regresión de Deming.

Métricas en las regresiones lineales bivariadas

Al finalizar el estudio, se deducen y presentan métricas de la regresión de Deming

Las cuales tienen como objeto evaluar el rendimiento medio del modelo. En [14] Willmot, J; Masuura K (2005) se presenta un estudio sobre la conveniencia de cada una de ellas de acuerdo al problema

Estudios aplicados de regresión lineal, considerando errores en las dos variables

La naturaleza de este trabajo es eminentemente teórica, no obstante, es procedente mencionar estudios en los cuales se consideran errores en las dos variables de las regresiones lineales de mínimos cuadrados. La regresión ortogonal es muy empleada en química en reacciones que generan los dos productos con errores en sus cantidades un ejemplo de esta aplicación se encuentra en [15] Ortiz J.V, Pioglani L, Besalú L (2010). En fisiología humana, también se ha aplicado regresión ortogonal para crear modelos entres producción de creatinina y ácido úrico, es destacado en este ámbito el trabajo [16] de Habibollah K, Masoumeh S, Eslam P (2009). Existen estudios sobre la ventaja de la regresión ortogonal sobre la regresión de mínimos cuadrados cuando se aplican estos modelos en laboratorios de medicina [17] Rainer H., Werner W. and Rainer K. (2013). En [18] (2011) Donevska S, Fiserova E, Karel H. (2011), presentan un estudio sobre la conveniencia de la regresión ortogonal en un caso aplicado a las medidas antropológicas en humanos. En el campo de la evaluación en pruebas académicas, [19] Taliha Keles. (2018) concluye luego de emplear dos bases de datos, que la regresión ortogonal presenta en esos casos, menor varianza de los errores con respecto a la regresión de mínimos cuadrados ordinarios.

RECTAS DE MÍNIMOS CUADRADOS CON ERRORES DE PENDIENTE “m”

En este capítulo y con el empleo de la geometría analítica y el cálculo diferencial, se encuentran los parámetros de las rectas de mínimos cuadrados de un conjunto de puntos en el plano, cuando la pendiente de los errores se fija previamente en un valor “m”.

Se determina una propiedad interesante desde el punto de vista matemático y geométrico, consistente en que si una vez obtenida el valor de la pendiente “b” de la recta de mínimos cuadrados, se define como pendiente de errores a “b”, la nueva recta obtenida tiene como pendiente “m”.

Se demuestra vectorialmente que la suma de errores en la recta de regresión lograda es cero.

Se encuentran las expresiones para calcular SCE, SCM y SCT en una recta de regresión de mínimos cuadrados cuando se escoge una pendiente “m” para los errores. Se demuestra que en este tipo de rectas de regresión se cumple que: SCT=SCM+SCE.

Finalmente se encuentra otra propiedad geométrica y matemática consistente en que, si se encuentra una nueva recta de regresión empleando como pendiente de los residuales, la pendiente “b” de la recta de mínimos cuadrados original, la SCM de la segunda recta es igual a la SCE de la primera y la SCE de la segunda es igual a la SCE de la primera.

2.1 Deducción

Recta de mínimos cuadrados

Definición: En el plano, se define como recta de mínimos cuadrados aquella que considerando un conjunto de puntos $P(x, y)$ presenta una suma de cuadrados de distancias euclideanas mínima. Las distancias se miden desde los puntos hasta la recta con una pendiente “m”.

En el inicio del desarrollo de este trabajo, se deduce mediante el empleo de la geometría analítica los parámetros de corte y pendiente de la recta de mínimos cuadrados, bajo la condición anotada que las distancias euclideanas se miden con una pendiente que se denominará “m”

Teorema 2.1 Si $y = a + bx$ es la recta de mínimos cuadrados, en la cual se han tomado los errores del conjunto de puntos con una pendiente “m”, los parámetros tienen como valor:

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

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

Demostración:

Sea la recta de mínimos cuadrados:

$y = a + bx$ **(1)**

Considérese n puntos:

$P_i(x_i, y_i), \quad i = 1, \ldots, n$

Para cada punto $P_i(x_i, y_i)$, la recta con pendiente m que lo incluye tiene la forma:

$y - y_i = m\,(x - x_i)$ **(2)**

Resolviendo el sistema de las dos ecuaciones (1) y (2) se obtiene los puntos de coordenadas (x, y) que corresponden a la intersección de las dos rectas.

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

$$y = \frac{by_i - mbx_i - am}{b - m}$$

Se encuentran las distancias horizontales y verticales entre los puntos $P_i(x_i, y_i)$ y los puntos de intersección con la recta.

$$x - x_i = \frac{y_i - bx_i - a}{b - m} \qquad (\mathbf{3})$$

$$y - y_i = m\left(\frac{y_i - bx_i - a}{b - m}\right) \qquad (\mathbf{4})$$

La suma de cuadrados de distancias S de los puntos P_i a la recta es:

$$S = \sum\left((x - x_i)^2 + ((y - y_i)^2)\right)$$

Reemplazando (3) y (4):

$$S = \frac{m^2 + 1}{(b - m)^2}\sum(y_i - bx_i - a)^2 \qquad (\mathbf{5})$$

Se recurre a la derivación para determinar los parámetros de la recta de mínimos cuadrados:

Se realiza:

$$\frac{dS}{da} = 0$$

$$\frac{dS}{da} = -\frac{2(m^2+1)}{(b-m)^2}\left(\sum y_i - b\sum x_i - na\right) = 0$$

Despejando a:

$a = \bar{y} - b\bar{x}$ **(6)**

Se ha demostrado que toda recta de mínimos cuadrados sin importar la pendiente de los errores, pasa por el centroide del conjunto de puntos considerados.

Reemplazando el valor de a (6) en la expresión de la suma de cuadrados de distancias (5):

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

Se obtiene una expresión de suma de cuadrados de distancias únicamente en función del parámetro b, puesto que la pendiente m de los errores está dada.

Empleando álgebra de sumatorias se llega a:

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

En el marco teórico se definió:

$$S_{xx} = \sum(x_i-\bar{x})^2$$

$$S_{yy} = \sum(y_i-\bar{y})^2$$

$$S_{xy} = \sum(x_i-\bar{x})(y_i-\bar{y})$$

De tal manera que:

$$S = \frac{m^2+1}{(b-m)^2}\left(b^2S_{xx} - 2bS_{xy} + S_{yy}\right) \qquad (\mathbf{7})$$

Ahora se realiza:

$$\frac{dS}{db} = 0$$

$$\frac{dS}{db} = \frac{m^2+1}{(b-m)^4}\Big(\big(2bS_{xx} - 2S_{xy}\big)(b-m)^2 - 2(b-m)\big(b^2S_{xx} - 2bS_{xy} + S_{yy}\big)\Big) = 0$$

$$\frac{dS}{db} = \frac{2(m^2+1)}{(b-m)^3}\Big(\big(bS_{xx} - S_{xy}\big)(b-m) - \big(b^2S_{xx} - 2bS_{xy} + S_{yy}\big)\Big) = 0$$

$$\frac{dS}{db} = \frac{2(m^2+1)}{(b-m)^3}\big(-bmS_{xx} + bS_{xy} + mS_{xy} - S_{yy}\big) = 0$$

$$-bmS_{xx} + bS_{xy} + mS_{xy} - S_{yy} = 0$$

Despejando b:

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}} \qquad (\mathbf{8})$$

La anterior es la pendiente de la recta de los mínimos cuadrados cuando las distancias desde los puntos hasta la recta se miden con una orientación que tiene como pendiente m.

A continuación, se demostrará que el valor de b obtenido corresponde a un mínimo relativo, para ello se recurre al criterio de la segunda derivada.

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^6}\Big(\big(-mS_{xx} + S_{xy}\big)(b-m)^3 - 3(b-m)^2\big(-bmS_{xx} + bS_{xy} + mS_{xy} - S_{yy}\big)\Big)$$

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\Big(\big(-mS_{xx} + S_{xy}\big)(b-m) - 3\big(-bmS_{xx} + bS_{xy} + mS_{xy} - S_{yy}\big)\Big)$$

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\big(-bm\,S_{xx} + m^2S_{xx} + b\,S_{xy} - mS_{xy} + 3bmS_{xx} - 3bS_{xy} - 3mS_{xy} + 3S_{yy}\big)$$

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\big(2bm\,S_{xx} + m^2S_{xx} - 2bS_{xy} - 4mS_{xy} + 3S_{yy}\big)$$

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\big(2b\big(m\,S_{xx} - S_{xy}\big) + m^2S_{xx} - 4mS_{xy} + 3S_{yy}\big)$$

Una simple inspección al siguiente factor mostrara que es positivo. Para $b \neq m$.:

$$\frac{2(m^2+1)}{(b-m)^4} > 0$$

Por lo tanto, solo se reemplaza el valor de b (8) en el factor restante:

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\left(2\left(mS_{xy} - S_{yy}\right) + m^2S_{xx} - 4mS_{xy} + 3S_{yy}\right)$$

$$\frac{d^2S}{db^2} = \frac{2(m^2+1)}{(b-m)^4}\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)$$

Es claro que:

cuando $mS_{xy} \leq 0$

$$m^2S_{xx} - 2mS_{xy} + S_{yy} > 0$$

$$\frac{d^2b}{db^2} > 0$$

Excepto en los inusuales casos en que:

$S_{yy} = 0$ y $m = 0$
$S_{yy} = 0$ y $S_{xx} = 0$

En el caso en que:

$$mS_{xy} > 0$$

Debe considerarse que:

$$m^2S_{xx} - 2mS_{xy} + S_{yy} \geq m^2S_{xx} - 2m\sqrt{S_{xx}S_{yy}} + S_{yy}$$

$$m^2S_{xx} - 2mS_{xy} + S_{yy} \geq \left(m\sqrt{S_{xx}} - \sqrt{S_{yy}}\right)^2 \geq 0$$

Puede notarse que la segunda derivada en este caso es también positiva, excepto en los casos muy inusuales ya mencionados.

Conclusión: Con las salvedades indicadas, el valor de la pendiente y el valor de corte con el eje "y" corresponden a la recta de mínima suma de distancias cuadradas a un conjunto de puntos, cuando dichas distancias se han medido con una pendiente m.

2.1.1 Casos especiales

Es interesante chequear los valores de la pendiente b, en los casos en los cuales la pendiente en la cual se miden los errores tiende a ∞ , o es cero.

Corolario 2.1.1.

Si $m \rightarrow \infty$, el parámetro de pendiente es:

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

Si $m = 0$, el parámetro de pendiente es:

$$b = \frac{S_{yy}}{S_{xy}}$$

Demostración:

Considérese el teorema 2.1:

Cuando:

$$m \rightarrow \infty$$

$$b = \lim_{m\to\infty}\left(\frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}\right)$$

$$b = \lim_{m\to\infty}\left(\frac{S_{xy} - S_{yy}/m}{S_{xx} - S_{xy}/m}\right)$$

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

Que es justamente la pendiente de la recta de mínimos cuadrados ordinarios, en la cual los errores son verticales.

Cuando:

$$m = 0$$

$$b = \frac{(0)S_{xy} - S_{yy}}{(0)S_{xx} - S_{xy}}$$

$$b = \frac{S_{yy}}{S_{xy}}$$

Que es la pendiente de la recta de mínimos cuadrados cuando los errores son horizontales.

Una inquietud que surge desde el punto de vista de la matemática y la geometría analítica, es conocer que pendiente m de orientación en los errores define una pendiente b de recta de mínimos cuadrados igual a cero es decir horizontal.

Corolario 2.1.1.1

Si una recta de mínimos cuadrados, tiene por pendiente $b = 0$, la pendiente con que se midieron los errores es:

$$m = \frac{S_{yy}}{S_{xy}}$$

Si una recta de mínimos cuadrados, tiene por pendiente $b \to \infty$, la pendiente con que se midieron los errores es:

$$m = \frac{S_{xy}}{S_{xx}}$$

Demostración:

Considérese el teorema 2.1:

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

$$0 = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

$$mS_{xy} - S_{yy} = 0$$

$$m = \frac{S_{yy}}{S_{xy}}$$

Que es la pendiente de la recta de mínimos cuadrados cuando se emplean errores horizontales.

También se desea conocer la pendiente m de orientación de distancias que define una pendiente b de recta de mínimos cuadrados igual a infinito, es decir vertical.

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

$$b \to \infty = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

$$mS_{xx} - S_{xy} = 0$$

$$m = \frac{S_{xy}}{S_{xx}}$$

Que es la pendiente de una recta de mínimos cuadrados cuando se emplean errores verticales.

Cuadro resumen de corolarios 2.1.1 y 2.1.1.1

Los anteriores son hallazgos llamativos que se resumen en el siguiente cuadro:

Pendiente del error	Pendiente recta mínimos cuadrados
$\to \infty$	S_{xy}/S_{xx}
0	S_{yy}/S_{xy}
S_{xy}/S_{xx}	$\to \infty$
S_{yy}/S_{xy}	0

Los resultados encontrados conducen una propiedad interesante que a continuación se enuncia y se demuestra:

2.1.2 Propiedad relativa a las pendientes de las rectas y las pendientes de los errores

A continuación y como corolario al teorema 2.1, se expresa la siguiente propiedad relativa a las pendientes de las rectas de mínimos cuadrados y las pendientes de los errores, existe una relación entre ellas que se expresan en el enunciado del siguiente corolario.

Corolario 2.1.2

Si en una recta bidimensional de mínimos cuadrados, se escoge una orientación de errores con pendiente m estableciendo una pendiente b y luego se toman errores para el conjunto de puntos con orientación de pendiente b, se obtiene una recta de mínimos cuadrados de pendiente m.

Demostración:

De la expresión (7) del teorema 2.1 se tiene que:

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

Cuando se eligen errores de pendiente b, se obtiene una recta de mínimos cuadrados b_1:

$$b_1 = \frac{bS_{xy} - S_{yy}}{bS_{xx} - S_{xy}}$$

$$b_1 = \frac{\left(\frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}\right) S_{xy} - S_{yy}}{\left(\frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}\right) S_{xx} - S_{xy}}$$

$$b_1 = \frac{mS_{xy}^2 - mS_{xx}S_{yy}}{S_{xy}^2 - S_{xx}S_{yy}}$$

$$b_1 = \frac{m\left(S_{xy}^2 - S_{xx}S_{yy}\right)}{S_{xy}^2 - S_{xx}S_{yy}}$$

$$b_1 = m$$

La demostración anterior puede resumirse como sigue:

Pendiente del error	Pendiente recta
m	b
b	m

La propiedad que se ha expuesto será importante en desarrollo posterior cuando se estudia el coeficiente de determinación de rectas de mínimos cuadrados con residuales de pendiente m.

Es importante estudiar que sucede cuando m y b son perpendiculares, lo anterior considerando la frecuencia en el empleo de la regresión ortogonal, el siguiente corolario aborda esta particular situación.

Corolario 2.1.2.1

Si la pendiente de la recta de mínimos cuadrados es perpendicular a la pendiente de los errores, se obtiene la recta de mínimos cuadrados ortogonales que justamente corresponde a la recta de regresión de Deming cuando $\delta = 1$.

Demostración:

Para la condición enunciada:

$$mb = -1$$

$$m = -\frac{1}{b}$$

De la expresión (7) del teorema 2.1 se tiene que:

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

Reemplazando el valor de m:

$$b = \frac{-\frac{S_{xy}}{b} - S_{yy}}{-\frac{S_{xx}}{b} - S_{xy}}$$

$$b = \frac{S_{xy} + bS_{yy}}{S_{xx} + bS_{xy}}$$

$$b^2 S_{xy} + b\,(S_{xx} - S_{yy}) - S_{xy} = 0$$

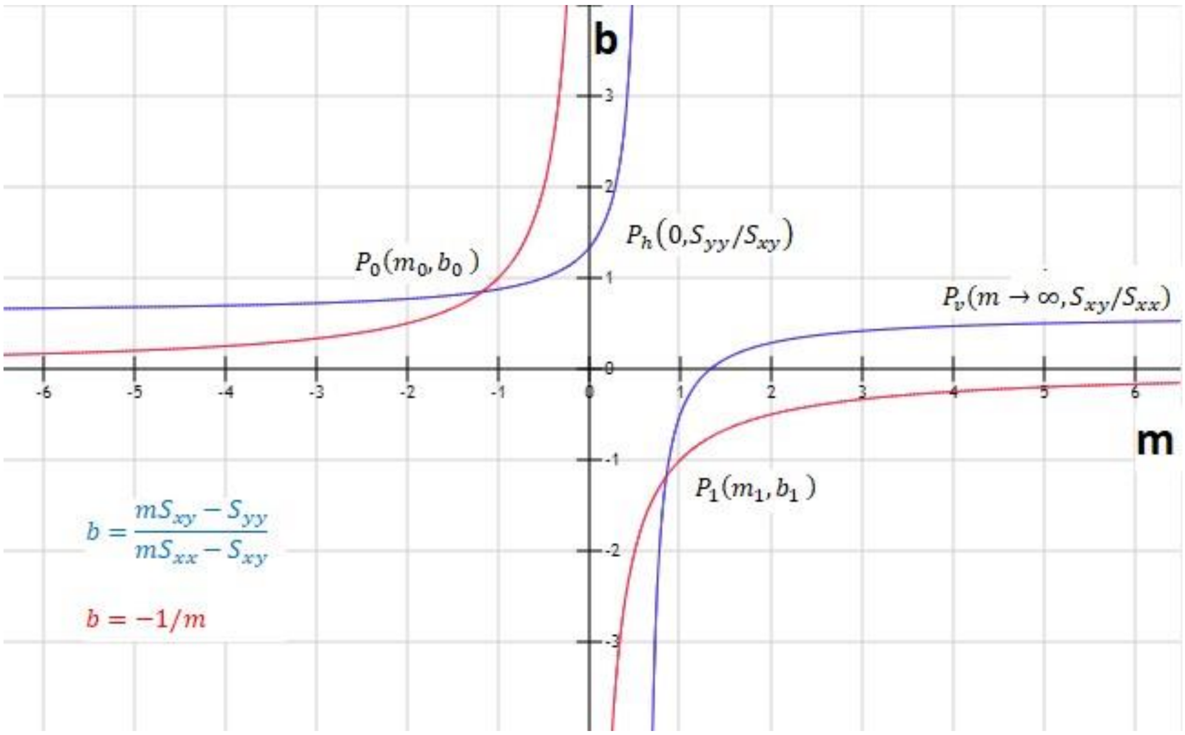
$$b = \frac{S_{yy} - S_{xx} \pm \sqrt{(S_{xx} - S_{yy})^2 + 4{S_{xy}}^2}}{2S_{xy}}$$

Con el signo (+) precediendo el radical se obtiene la pendiente de recta de mínimos cuadrados ortogonales, que corresponde a la recta de regresión de Deming cuando $\delta = 1$.

El signo (-) se reserva para la pendiente de la “recta de máximos cuadrados ortogonales.

Recta de máximos cuadrados ortogonales es la recta cuya suma de distancias cuadradas ortogonales desde un conjunto de puntos es máxima. En [13] Villota A, Hidalgo A. (2018) se deduce esta recta y se demuestra mediante el cálculo diferencial la pertinencia del empleo del signo (-) precediendo el radical.

Grafica 1Pendientes (b) de rectas de mínimos cuadrados considerando errores con pendiente m.



La línea en azul, muestra la pendiente de la recta de mínimos cuadrados en función de la pendiente de los errores. La línea roja indica el valor negativo del inverso de la pendiente. Nótese que el intercepto P_0 tiene por coordenadas la pendiente de los errores y la pendiente de la recta de mínimos cuadrados ortogonales. El intercepto P_1 tiene por coordenadas la pendiente de lo errores y la pendiente de la recta de los máximos cuadrados ortogonales. El punto P_v situado en la pendiente ∞ de los errores (vertical), tiene como ordenada el valor de la pendiente de la recta de mínimos cuadrados ordinarios: S_{xy}/S_{xx}. El punto P_h corresponde a una pendiente de errores cero y corresponde a la pendiente de la recta de mínimos cuadrados horizontales: S_{yy}/S_{xy}. Para construir la gráfica se asumió: $S_{xx} = 10, S_{yy} = 8, S_{xy} = 6$. Por lo tanto: $P_0\left(-6/(\sqrt{37}-1), (\sqrt{37}-1)/6\right)$, $P_1\left(6/(\sqrt{37}+1), -(\sqrt{37}+1)/6\right)$, $P_v(m \to \infty,\ 3/5)$, $P_h(0,\ 4/3)$

2.2 Suma de errores medidos con una pendiente "m" en una recta de mínimos cuadrados.

Existe una propiedad que se aplica a todas las rectas de mínimos cuadrados sin importar la pendiente de las distancias medidas entre los puntos y la recta. Consiste en que la suma vectorial de los errores es cero.

Teorema 2.2

Sea:

$$y = a + bx$$

Una recta de mínimos cuadrados en la cual se han medido los errores con una pendiente m.

Entonces la suma de los errores es cero.

Demostración.

Se considerará que cada error es un vector r:

$$\overrightarrow{r_t}=(x_t - x)i + (y_t - y)j$$

Donde cada punto t de los n totales tiene por coordenadas:

$$P_t(x_t,\ y_t)$$

Y el intercepto del error con la recta de mínimos cuadrados es:

$$P(x, y)$$

De acuerdo con las expresiones (3) y (4):

$$\overrightarrow{r_t} = \left(\frac{y_t - bx_t - a}{b - m}\right) i + m\left(\frac{y_t - bx_t - a}{b - m}\right) j$$

La suma de los n vectores es:

$$\sum \overrightarrow{r_t} = \left(\sum \left(\frac{y_t - bx_t - a}{b - m}\right)\right) i + m\left(\sum \left(\frac{y_t - bx_t - a}{b - m}\right)\right) j$$

$$\sum \overrightarrow{r_t} = \frac{\sum (y_t - bx_t - a)}{b - m}(i\ + mj)$$

$$\sum \overrightarrow{r_t} = \frac{\sum y_t - b\sum x_t - na}{b - m}(i\ + mj)$$

$$\sum \overrightarrow{r_t} = \frac{n(\bar{y} - b\bar{x} - a)}{b - m}(i \ + mj)$$

Anteriormente se había demostrado que todas las rectas de mínimos cuadrados pasan por el centroide:

$C(\bar{x},\ \bar{y})$

Por lo tanto:

$\bar{y} - b\bar{x} - a = 0$

Finalmente:

$$\sum \vec{r} = \frac{n(\bar{y} - b\bar{x} - a)}{b - m}(i \ + mj) = 0$$

Conclusión la suma de errores en todas las rectas de mínimos cuadrados es cero.

2.3 SCT, SCE y SCM en una recta de mínimos cuadrados con errores medidos con una pendiente “m”

En las rectas de regresión existen métricas que son importantes en la evaluación del modelo, especialmente en el cálculo del coeficiente de correlación. En este aparte del estudio, se encuentran las expresiones para el cálculo de la suma de cuadrados totales, suma de cuadrados de los errores y suma de cuadrados del modelo en función de S_{xx}, S_{yy}, S_{xy} y m.

Se emplea la siguiente notación:

SCT = Suma de cuadrados totales.
SCE = Suma de cuadrado de los errores.
SCM= Suma de cuadrados del modelo.

Suma de cuadrados totales.

$$SCT = \sum((x_i - \bar{x})^2 + (y_i - \bar{y})^2)$$

$$SCT = S_{xx} + S_{yy}$$

Suma de cuadrados de los errores:

Teorema 2.3

Sea:

$$y = a + bx$$

Una recta de regresión en la cual los errores se han medido con una pendiente m, entonces la suma de cuadrado de los errores es:

$$SCE = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

Demostración

De acuerdo a (7):

$$SCE = \frac{m^2+1}{(b-m)^2}\left(b^2 S_{xx} - 2bS_{xy} + S_{yy}\right) \qquad (9)$$

Pero también se sabe por (8):

$$b = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

Reemplazando:

$$SCE = \frac{m^2+1}{\left(\frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}} - m\right)^2}\left(\left(\frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}}\right)^2 S_{xx} - 2\left(\frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}}\right)S_{xy} + S_{yy}\right)$$

$$SCE = \frac{m^2+1}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\left(\left(mS_{xy}-S_{yy}\right)^2 S_{xx} - 2\left(mS_{xy}-S_{yy}\right)\left(mS_{xx}-S_{xy}\right)S_{xy} + \left(mS_{xx}-S_{xy}\right)^2 S_{yy}\right)$$

$$SCE = \frac{m^2+1}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\left(m^2S_{xx}^2S_{yy} - 2mS_{xx}S_{yy}S_{xy} + S_{xx}S_{yy}^2 - m^2S_{xx}S_{xy}^2 + 2mS_{xy}^3 - S_{yy}S_{xy}^2\right)$$

$$SCE = \frac{m^2+1}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\left(S_{xx}S_{yy}\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right) - S_{xy}^2\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)\right)$$

$$SCE = \frac{m^2+1}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\left(\left(S_{xx}S_{yy} - S_{xy}^2\right)\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)\right)$$

$$SCE = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2S_{xx} - 2mS_{xy} + S_{yy}} \qquad (\mathbf{10})$$

Suma de cuadrados del modelo:

Teorema 2.3.1

Sea:

$y = a + bx$

Una recta de regresión en la cual los errores se han medido con una pendiente m, entonces la suma de cuadrado del modelo es:

$$SCM = \frac{\left(mS_{xx}-S_{xy}\right)^2 + \left(mS_{xy}-S_{yy}\right)^2}{m^2S_{xx} - 2mS_{xy} + S_{yy}}$$

Demostración

De (3) y (4) del capítulo 1 se obtiene que:

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

$$y = m\left(\frac{y_i - bx_i - a}{b - m}\right) + y_i$$

Donde cada punto $p(x, y)$ corresponde a la intersección del error con la recta de mínimos cuadrados.

Se sabe que:

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

Reemplazando en las ecuaciones anteriores:

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

$$y = \frac{by_i - m\bar{y} - mbx_i + mb\bar{x}}{b - m}$$

Para cada uno de estos puntos $p(x, y)$

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

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

$$y - \bar{y} = \frac{by_i - m\bar{y} - mbx_i + mb\bar{x}}{b - m} - \bar{y}$$

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

Por lo tanto, el cuadrado de la distancia del punto al centroide es:

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

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

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

$$SCM = \frac{b^2+1}{(b-m)^2}\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right) \qquad (\mathbf{11})$$

Por (8) se sabe que:

$$b = \frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}}$$

Reemplazando en SCM:

$$SCM = \frac{\left(\frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}}\right)^2+1}{\left(\frac{mS_{xy}-S_{yy}}{mS_{xx}-S_{xy}}-m\right)^2}\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)$$

$$SCM = \frac{\left(mS_{xx}-S_{xy}\right)^2+\left(mS_{xy}-S_{yy}\right)^2}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)$$

$$SCM = \frac{\left(mS_{xx}-S_{xy}\right)^2+\left(mS_{xy}-S_{yy}\right)^2}{m^2S_{xx} - 2mS_{xy} + S_{yy}}$$

Se demuestra a continuación una propiedad de todas las rectas de mínimos cuadrados que se presenta sin importar el valor de la pendiente de los errores (m).

Dicha propiedad consiste en que la suma de cuadrados totales es igual a la suma de cuadrados del modelo más la suma de cuadrados de los errores.

$$SCT = SCM + SCE$$

Teorema 2.3.2

Sea:

$$y = a + bx$$

Una recta de regresión en la cual los errores se han medido con una pendiente m, entonces la suma de cuadrados totales es igual a la suma del cuadrado del modelo más la suma del cuadrado de los errores.

$$SCT = SCM + SCE$$

Demostración

De los teoremas 2.3 y 2.3.1 se puede establecer lo siguiente:

$$S_{xx} + S_{yy} = \frac{\left(mS_{xx} - S_{xy}\right)^2 + \left(mS_{xy} - S_{yy}\right)^2}{m^2 S_{xx} - 2mS_{xy} + S_{yy}} + \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

$$S_{xx} + S_{yy} = \frac{m^2 S_{xx}^2 - 2mS_{xx}S_{yy} + S_{xx}\ S_{yy} + m^2 S_{xx}\ S_{yy} - 2mS_{yy}S_{xy} + S_{yy}^2}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

$$S_{xx} + S_{yy} = \frac{S_{xx}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right) + S_{yy}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

$$S_{xx} + S_{yy} = \frac{(S_{xx} + S_{yy})\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

$$S_{xx} + S_{yy} = S_{xx} + S_{yy}$$

Por lo tanto:

$$SCT = SCM + SCE$$

Es importante retomar la propiedad que se encontró en 2.1.3, según la cual, si se escoge una orientación de errores con pendiente m y se obtiene una recta de mínimos cuadrados de pendiente b y luego se toman errores para el conjunto de puntos con orientación de pendiente b, se obtiene una recta de mínimos cuadrados de pendiente m.

2.3.1 Propiedad relativa a SCT, SCE y SCM

En este capítulo se encuentra otra propiedad muy llamativa relacionada con la ya mencionada (2.1.3) y consiste en que la suma de cuadrados del modelo, es la suma de los cuadrados de los errores, cuando estos se toman en la orientación de la pendiente b de la recta de mínimos cuadrados y mediante los cuales es factible determinar una recta de mínimos cuadrados de pendiente m.

Corolario 2.3.1

En una recta de regresión de mínimos cuadrados bivariada, la suma de cuadrados del modelo, es la suma de cuadrados de los errores, cuando estos se toman en la orientación de la pendiente b de la recta. El valor de la pendiente b se había obtenido previamente considerando los errores con una pendiente m

Demostración.

Considérese la suma de cuadrados de los errores y del modelo cuando la pendiente de los errores es m. Véase (9) y (11)

$$SCE = \frac{m^2+1}{(b-m)^2}\left(b^2 S_{xx} - 2bS_{xy} + S_{yy}\right)$$

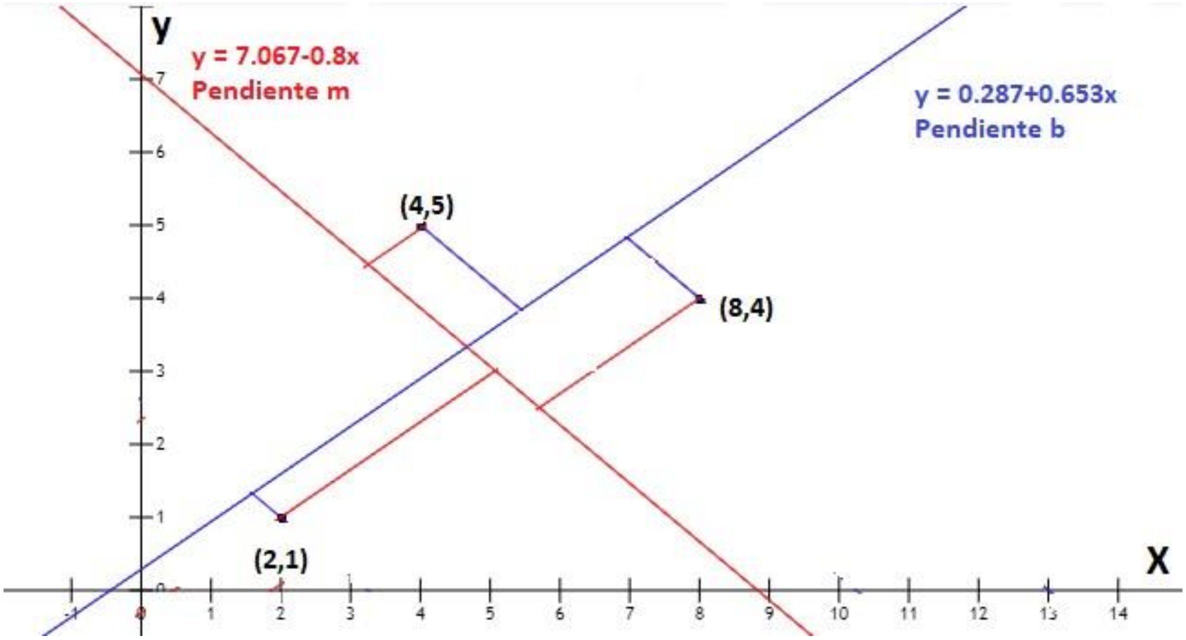
$$SCM = \frac{b^2+1}{(b-m)^2}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)$$

Ahora considérese la suma de cuadrados de los errores y del modelo cuando la pendiente de los errores es b. Bastará en las dos expresiones anteriores reemplazar m por b y b por m. Llámese a las nuevas expresiones SCE_1 y SCM_1

$$SCE_1 = \frac{b^2+1}{(b-m)^2}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)$$

$$SCM_1 = \frac{m^2+1}{(b-m)^2}\left(b^2 S_{xx} - 2bS_{xy} + S_{yy}\right)$$

Grafica 2. Rectas de mínimos cuadrados cuando se intercambia la pendiente m de los errores por la pendiente b calculada.



Se escoge pendiente m =-0.8, obtiene b=0.653 (recta de mínimos cuadrados y errores en azul). Luego se toma m=0.563, se obtiene b=0.8 (recta de mínimos cuadrados y errores en rojo). Se visualiza además que SCE azul = SCM rojo y que SCE rojo = SCM azul.

ESTUDIO DE LA RECTA DE REGRESIÓN DE DEMING: PENDIENTE DE ERRORES, COEFICIENTE DE DETERMINACIÓN Y CAMBIOS EN LA PENDIENTE DE LA RECTA.

En este capítulo, se demuestra que la pendiente de la recta de regresión de Deming, puede obtenerse a partir de una recta de mínimos cuadrados en la cual se utiliza una pendiente en los errores. Se encuentra la expresión para definir la pendiente de errores.

Luego se deduce la expresión para calcular el coeficiente de determinación en la recta de regresión de Deming. Adicionalmente se realiza un estudio detallado de dicho coeficiente y se demuestra que el máximo valor se obtiene cuando se realiza regresión ortogonal. Se calcula el valor particular del coeficiente de determinación para la regresión ortogonal.

Finalmente se realiza un estudio sobre el cambio de la pendiente de la recta de regresión de Deming, en función del cambio de la razón de varianzas de los errores de las variables.

3.1 Pendiente de los errores en la recta de regresión de Deming.

En adelante se llama b_D a la pendiente de la recta de regresión de Deming.

La pendiente de la recta de la regresión de Deming (b_D) es:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}{2S_{xy}}$$

Ver: [10] Jensen A. Ch (2007)

Donde delta (δ)es la razón de varianza de errores en la variable "y" y la varianza de errores en la variable "x"

$$\delta = \frac{\sigma_{\varepsilon y}^2}{\sigma_{\varepsilon x}^2}$$

En el inicio de este capítulo se deduce la pendiente (m) a partir de la cual, empleando el método de mínimos cuadrados, se puede llegar a la pendiente de la recta de regresión de Deming.

Teorema 3.1

La recta de regresión de Deming es una recta de mínimos cuadrados, en la cual los errores se han medido con una pendiente:

$$m = \frac{-\delta}{b_D}$$

Demostración

Por (8) en el capítulo 2:

$$b_D = \frac{mS_{xy} - S_{yy}}{mS_{xx} - S_{xy}}$$

Despejando m:

$$m = \frac{S_{yy} - S_{xy}b_D}{S_{xy} - S_{xx}b_D}$$

Reemplazando el valor de b_D:

$$m = \frac{S_{yy} - S_{xy}\left(\dfrac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}}{2S_{xy}}\right)}{S_{xy} - S_{xx}\left(\dfrac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}}{2S_{xy}}\right)}$$

$$m = \frac{2S_{yy}S_{xy} - S_{xy}\left(S_{yy} - \delta S_{xx} + \sqrt{\left(\delta S_{xx} - S_{yy}\right)^2 + 4\delta S_{xy}^2}\right)}{2S_{xy}^2 - S_{xx}\left(S_{yy} - \delta S_{xx} + \sqrt{\left(\delta S_{xx} - S_{yy}\right)^2 + 4\delta S_{xy}^2}\right)}$$

$$m = \frac{S_{xy}\left(S_{yy} + \delta S_{xx} - \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}\right)}{\left(2S_{xy}^2 - S_{xx}\left(S_{yy} - \delta S_{xx} + \sqrt{\left(\delta S_{xx} - S_{yy}\right)^2 + 4\delta S_{xy}^2}\right)\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}\right)}$$

$$m = \frac{S_{xy}\left(\left(S_{yy} + \delta S_{xx}\right)^2 - \left(S_{yy} - \delta S_{xx}\right)^2 - 4\delta S_{xy}^2\right)}{\left(2\left(S_{xy}^2 - S_{xx}S_{yy}\right) + S_{xx}\left(S_{yy} + \delta S_{xx}\right) - S_{xx}\sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta S_{xy}{}^2}\right)}$$

$$m = \frac{-4\delta S_{xy}\left(S_{xy}^2 - S_{xx}S_{yy}\right)}{2\left(S_{xy}^2 - S_{xx}S_{yy}\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}\right) + S_{xx}\left(\left(S_{yy} + \delta S_{xx}\right)^2 - \left(S_{yy} - \delta S_{xx}\right)^2 - 4\delta {S_{xy}}^2\right)}$$

$$m = \frac{-4\delta S_{xy}\left(S_{xy}^2 - S_{xx}S_{yy}\right)}{2\left(S_{xy}^2 - S_{xx}S_{yy}\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}\right) + S_{xx}\left(4\delta S_{xx}S_{yy} - \delta S_{xx}S_{yy}\right)}$$

$$m = \frac{-4\delta S_{xy}\left(S_{xy}^2 - S_{xx}S_{yy}\right)}{2\left(S_{xy}^2 - S_{xx}S_{yy}\right)\left(S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}\right) - 4\delta S_{xx}\left(S_{xy}^2 - S_{xx}S_{yy}\right)}$$

$$m = \frac{-2\delta S_{xy}}{S_{yy} + \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2} - 2\delta S_{xx}}$$

$$m = \frac{-2\delta S_{xy}}{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}$$

$$m = \frac{-\delta}{\frac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}{2S_{xy}}}$$

Nótese que el denominador de la anterior expresión es la pendiente de la regresión de Deming.

$$m = \frac{-\delta}{b_D}$$

Es importante destacar que, en la recta de regresión de Deming, el producto de la pendiente de los errores por la pendiente de la recta de regresión es $-\delta$ y en el caso particular que $\delta = 1$, el producto de las pendientes mencionadas es -1, valor que indica que cuando $\delta = 1$, la regresión de Deming se convierte en regresión ortogonal. Lo anterior considerando un principio básico demostrable de la geometría analítica que se enuncia de la siguiente manera: el producto de pendientes de dos rectas perpendiculares es -1. Véase [2] Wexler Ch (1977)

3.2 Estudio del coeficiente de determinación en la recta de regresión de Deming

Coeficiente de determinación en la recta de regresión de Deming:

El coeficiente de determinación es una métrica muy importante para evaluar la calidad del modelo, en esta aparte se deduce una expresión matemática para definir dicho coeficiente en función de S_{xx}, S_{yy}, S_{xy} y la pendiente de la regresión de Deming (b_D)

Teorema 3.2

Considerando el valor de suma de cuadrados del modelo (11):

$$SCM = \frac{b^2+1}{(b-m)^2}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)$$

Y la suma de cuadrados totales:

$$SCT = S_{xx} + S_{yy}$$

El valor del coeficiente de determinación del modelo es:

$$R^2 = \frac{(b_D^2+1)\left(S_{xx} + 2\delta b_D S_{xy} + b_D^2 S_{yy}\right)}{(b_D^2+\delta)^2\left(S_{xx}+S_{yy}\right)}$$

Demostración:

Adecuando la suma de cuadrados del modelo a la recta de regresión de Deming:

$$SCM = \frac{b_D^2+1}{(b_D-m)^2}\left(m^2 S_{xx} - 2mS_{xy} + S_{yy}\right)$$

Pero, según lo demostrado en 3.1:

$$m = \frac{-\delta}{b_D}$$

Reemplazando ese valor:

$$SCM = \frac{b_D^2+1}{(b_D^2+\delta)^2}\left(S_{xx} + 2\delta b_D S_{xy} + b_D^2 S_{yy}\right)$$

$$R^2 = \frac{(b_D^2+1)\left(S_{xx} + 2\delta b_D S_{xy} + b_D^2 S_{yy}\right)}{(b_D^2+\delta)^2\left(S_{xx}+S_{yy}\right)} \qquad (\mathbf{12})$$

Donde b_D es la pendiente de la regresión de Deming:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{\left(S_{yy} - \delta S_{xx}\right)^2 + 4\delta S_{xy}^2}}{2S_{xy}}$$

3.3 Recta de regresión de Deming con coeficiente de determinación máximo

Es necesario conocer cuál es la recta de regresión de Deming con mayor coeficiente de determinación, lo anterior desde el punto de vista teórico, pero también considerando aplicaciones prácticas. En este aparte del estudio, se aplica el cálculo diferencial para resolver el problema planteado.

Teorema 3.3

Considerando las rectas de regresión de Deming, la recta con máximo coeficiente de determinación es la recta con razón de varianzas de errores $\delta = 1$, es decir la recta de mínimos cuadrados ortogonales.

Demostración

Retómese la expresión de suma de cuadrados de errores, medidos con una pendiente m (10):

$$SCE = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

El coeficiente de correlación será máximo cuando la suma de cuadrados de los errores sea mínima.

Se recurre a la primera derivada y se iguala a cero:

$$\frac{d(SCE)}{d(m)} = \left(S_{xx}S_{yy} - S_{xy}^2\right)\left(\frac{2m\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right) - \left(2mS_{xx} - 2S_{xy}\right)(m^2+1)}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\right) = 0$$

$$\frac{d(SCE)}{d(m)} = -2\left(S_{xx}S_{yy} - S_{xy}^2\right)\left(\frac{m^2S_{xy} - m\left(S_{yy} - S_{xx}\right) - S_{xy}}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\right) = 0$$

$$m^2S_{xy} - m\left(S_{yy} - S_{xx}\right) - S_{xy} = 0$$

$$m = \frac{S_{yy} - S_{xx} \pm \sqrt{\left(S_{yy} - S_{xx}\right)^2 + 4S_{xy}^2}}{2S_{xy}}$$

Se presentan dos puntos críticos que ameritan estudiarse con el criterio de la segunda derivada:

Al realizar operaciones se obtiene:

$$\frac{d^2(SCE)}{dm^2} = -2\left(S_{xx}S_{yy} - S_{xy}^2\right)\left(\frac{2mS_{xy} - \left(S_{yy} - S_{xx}\right)}{\left(m^2S_{xx} - 2mS_{xy} + S_{yy}\right)^2}\right)$$

Es claro que:

$$\frac{S_{xx}S_{yy}-S_{xy}^2}{\left(m^2S_{xx}-2mS_{xy}+S_{yy}\right)^2}>0$$

Para : $\rho \neq (\pm 1),\ \ m \neq (\pm\infty)$

Para determinar el signo de la segunda derivada, bastará determinar el signo que resulta de reemplazar los valores críticos en:

$$-2\left(2mS_{xy}-\left(S_{yy}-S_{xx}\right)\right)$$

Al reemplazar:

$$m=\frac{S_{yy}-S_{xx}+\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}}{2S_{xy}}$$

Se obtiene:

$$-2\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}<0$$

El valor crítico corresponde a un máximo relativo.

Al reemplazar:

$$m=\frac{S_{yy}-S_{xx}-\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}}{2S_{xy}}$$

Se obtiene:

$$2\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}\ >0$$

El valor crítico corresponde a un mínimo relativo.

De acuerdo a lo anterior:

El valor mínimo de SCE se presenta cuando:

$$m=\frac{S_{yy}-S_{xx}-\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}}{2S_{xy}}$$

Multiplicando numerador y denominador por la conjugada del numerador:

$$m=\frac{\left(S_{yy}-S_{xx}-\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}\right)\left(S_{yy}-S_{xx}+\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}\right)}{2S_{xy}\left(S_{yy}-S_{xx}+\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}\right)}$$

$$m=\frac{-1}{\frac{\left(S_{yy}-S_{xx}+\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}\right)}{2S_{xy}}}=\frac{-1}{b_o}$$

El denominador de la expresión es la pendiente de la recta de regresión de Deming, para una razón de varianzas $\delta=1$. Para ese valor, la pendiente encontrada corresponde a la recta de mínimos cuadrados ortogonales que se llamará b_o.

En otras palabras, la recta de regresión ortogonal es la recta perteneciente a la regresión de Deming con el máximo coeficiente de determinación.

Corolario 3.3.1

El valor de pendiente de errores “m” que produce un máximo relativo para SCE. En este caso:

$$m=\frac{S_{yy}-S_{xx}+\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}}{2S_{xy}}$$

Si se toma está pendiente para los errores, se obtiene la pendiente de recta de máximos cuadrados ortogonales

$$\frac{S_{yy}-S_{xx}-\sqrt{\left(S_{yy}-S_{xx}\right)^2+4S_{xy}^2}}{2S_{xy}}$$

Ver corolario 2.1.2, 2.1.2.1 y teorema 3.1.

3.4 Coeficiente de determinación en recta de mínimos cuadrados ortogonales (recta de regresión de Deming con $\boldsymbol{\delta = 1}$)

Mediante el teorema 3.3 Se conoció que el mayor coeficiente de determinación se obtiene con la recta de mínimos cuadrados ortogonales. Es preciso cuantificar ese valor en función de $S_{xx}, S_{yy} \;\; y \;\; S_{xy}$

Teorema 3.4

El coeficiente de determinación en la recta de mínimos cuadrados ortogonales es:

$$R^2 = \frac{S_{xx} + S_{yy} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xy}^2}}{2\left(S_{xx} + S_{yy}\right)}$$

Se exponen dos demostraciones:

Demostración 1

Para determinar el valor de SCE para la recta de mínimos cuadrados ortogonales, basta con reemplazar el valor de la pendiente de los errores obtenida ($-1/b_o$) en la expresión de SCE que se encuentra en función de m. Sin embargo, se presenta a continuación un artificio a manera de atajo que permite realizar el cálculo.

Llámese SCE_o a la suma de cuadrados de los errores ortogonales. Ver (10).

$$SCE_o = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}$$

Un trabajo algebraico sencillo conduce a:

$$\left(S_{xx}S_{yy} - S_{xy}^2 - SCE_o S_{xx}\right)m^2 + 2S_{xy}SCE_o m + S_{xx}S_{yy} - S_{xy}^2 - SCE_o S_{yy} = 0$$

Se había demostrado anteriormente qué en la recta de mínimos cuadrados ortogonales, está el mínimo SCE (mínimo relativo), por lo tanto, la pendiente m de los errores debe necesariamente ser única.

La última expresión es una ecuación de segundo grado, si se considera como variable m. Por ser única la solución de m, el valor del radical incluido en dicha solución debe ser cero:

$$4S_{xy}^2 SCE_o^2 - 4\left(S_{xx}S_{yy} - S_{xy}^2 - SCE_o S_{xx}\right)\left(S_{xx}S_{yy} - S_{xy}^2 - SCE_o S_{yy}\right) = 0$$

Desarrollando la ecuación anterior se llega a:

$$SCE_o^2 - (S_{xx} + S_{yy})SCE_o + S_{xx}S_{yy} - S_{xy}^2 = 0$$

Resolviendo la anterior ecuación de segundo grado en la cual la variable es SCE_o :

$$SCE_o = \frac{S_{xx} + S_{yy} \pm \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2}$$

El signo menos corresponde necesariamente a la suma de cuadrados de errores a la recta de mínimos cuadrados ortogonales y el signo más corresponde a la suma de cuadrados de errores a la recta de máximos cuadrados ortogonales

Por lo tanto, en la recta de mínimos cuadrados ortogonales:

$$SCE_o = \frac{S_{xx} + S_{yy} - \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2}$$

$$SCM_o = S_{xx} + S_{yy} - SCE_o$$

$$SCM_o = \frac{S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2}$$

$$R^2 = \frac{S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2(S_{xx} + S_{yy})}$$

El anterior es el máximo coeficiente de determinación entre todos los coeficientes de determinación de la regresión de Deming y corresponde a la recta de mínimos cuadrados ortogonales.

Demostración 2

Simplemente se aplica la expresión previamente deducida en (12):

$$R^2 = \frac{(b_D^2 + 1)(S_{xx} + 2\delta b_D S_{xy} + b_D^2 S_{yy})}{(b_D^2 + \delta)^2 (S_{xx} + S_{yy})}$$

Usando:

$$\delta = 1$$

$$b_D = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}}$$

Luego de realizar los reemplazos correspondientes y trabajar la expresión algebraicamente se llega a;

$$R^2 = \frac{S_{xx} + S_{yy} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xy}^2}}{2\left(S_{xx} + S_{yy}\right)}$$

Nota. En el capítulo 5, se presenta una tercera demostración.

3.5 Cambios en la pendiente de la recta de Deming, con el incremento de la razón de varianzas de los errores (δ)

Para un conjunto de puntos dados en el plano, la pendiente de la recta de regresión de Deming, depende del valor de la varianza de errores (δ). En esta parte del estudio se busca definir cómo cambia la pendiente de la recta de regresión de Deming, con el incremento de δ.

Teorema 3.5

Considerando la pendiente de la recta de regresión de Deming:

puede afirmarse que si disminuye δ, el valor absoluto de b_D aumenta y si aumenta $\delta,$ el valor absoluto de b_D disminuye.

Demostración

Se emplea la siguiente notación:

$$k = \frac{S_{yy}}{S_{xx}}$$

Considerar además que:

$$S_{xy} = \rho\sqrt{S_{xx}S_{yy}}$$

Un simple trabajo algebraico en la expresión de la pendiente de la recta de Deming empleando las dos últimas expresiones anotadas transforma:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}{2S_{xy}}$$

En:

$$b_D = \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

Caso 1

Cuando $0 \leq \rho \leq 1$

$$b_D = \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

Considérese el radical:

$\sqrt{(\delta - k)^2 + 4k\delta\rho^2}$

Se nota que el valor mínimo de la expresión se produce cuando $\rho = 0$, en este caso el resultado es:

$|\delta - k|$

También se advierte que el valor máximo del radical se presenta cuando $\rho = \pm 1$, bajo esta circunstancia el resultado es:

$\delta + k$

Por lo tanto, para cada situación particular de ρ el radical puede expresarse de la siguiente manera:

$\sqrt{(\delta - k)^2 + 4k\delta\rho^2} \; = \; \alpha(\delta + k)$

Con:

$$\frac{|\delta - k|}{\delta + k} \leq \alpha \leq 1$$

Por lo tanto:

$$b_D = \frac{k - \delta + \; \alpha(\delta + k)}{2\rho\sqrt{k}}$$

$$b_D = \frac{k(1 + \alpha) - \delta(1 - \alpha)}{2\rho\sqrt{k}}$$

Nótese que:

$1 - \alpha \; \geq 0$

Por lo tanto:

Si disminuye δ aumenta b_D
Si aumenta δ disminuye b_D.

Caso 2.

Cuando $\rho < 0$

El razonamiento es similar al caso 1, pero dado que el signo de la pendiente lo define el signo de ρ:

Si disminuye δ, la pendiente se vuelve más negativa.
Si aumenta δ, la pendiente se vuelve menos negativa.

En general puede afirmarse que si disminuye δ, el valor absoluto de b_D aumenta y si aumenta δ, el valor absoluto de b_D disminuye.

DIFERENCIAS DE PENDIENTES Y ÁNGULOS EN LAS RECTAS DE REGRESIÓN DE DEMING

En este capítulo se realiza un estudio sobre las diferencias en las pendientes y los ángulos en las rectas de regresión de Deming. Se emplean para el cotejo, la recta de regresión de Deming en general y rectas notables concretamente la recta de mínimos cuadrados ordinarios, la recta de regresión ortogonal y la recta de mínimos cuadrados horizontales.

Considerando que se ha llamado k a la razón de varianzas entre los valores "y" y los valores de "x", se demuestra que, en valor absoluto, la pendiente de la recta de regresión de Deming en general, es mayor o igual que la pendiente de la recta de regresión de mínimos cuadrados ordinarios. También se demuestra que, en valor absoluto, la pendiente de la recta de regresión de mínimos cuadrados horizontales es mayor o igual que la pendiente de la recta de regresión de Deming en general.

Se realiza un estudio comparativo sobre los ángulos formados por la recta de mínimos cuadrados ortogonales y recta de mínimos cuadrados ordinarios, se determina el valor de k, para el cual ese ángulo es máximo. Se realiza el mismo estudio para los ángulos formados entre la recta de regresión de mínimos cuadrados horizontales y recta de mínimos cuadrados ortogonales, también en este caso se determina el valor de k. En los dos estudios se define ese valor de diferencia angular y se encuentra algo llamativo, el valor de ángulo máximo es el mismo, aunque los valores de k en cada caso son diferentes. Otra propiedad interesante que se encuentra es que para el valor $k = 1$ la diferencia angular es la misma para los dos casos.

4.1. Comparación entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de los mínimos cuadrados ordinarios

Es tradicional el empleo de la recta de regresión de mínimos cuadrados ordinarios, que no considera la existencia de errores de la variable "x". En este ítem se establece cuál de las dos rectas de regresión (mínimos cuadrados ordinarios o recta de regresión de Deming) posee en general mayor valor absoluto en su valor de pendiente.

Llámese b_v a la pendiente de la recta de mínimos cuadrados ordinarios.

$$b_v = \frac{S_{xy}}{S_{xx}}$$

A la anterior expresión se llega aplicando a la pendiente de la recta de regresión de Deming, límite cuando $\delta \to \infty$

Pero:

$S_{xy} = \sqrt{S_{xx}S_{yy}}\ \rho$

Donde ρ es el coeficiente de correlación de Pearson.

Reemplazando este valor en las pendientes de las rectas de regresión de mínimos cuadrados ordinarios y Deming se obtiene:

Pendiente de recta de mínimos cuadrados ordinarios:

$$b_v = \sqrt{\frac{S_{yy}}{S_{xx}}}\ \rho \qquad (\mathbf{12})$$

Pendiente de recta de regresión de Deming:

$$b_D = \frac{S_{yy} - S_{xx}\delta + \sqrt{(S_{yy} - S_{xx}\delta)^2 + 4\delta\rho^2 S_{xx}S_{yy}}}{2\rho\sqrt{S_{xx}S_{yy}}}$$

Caso 1

Cuando $\rho > 0$

Teorema 4.1

Siendo:

$$b_v = \sqrt{\frac{S_{yy}}{S_{xx}}}\ \rho$$

La pendiente de recta de mínimos cuadrados ordinarios:

$$b_D = \frac{S_{yy} - S_{xx}\delta + \sqrt{(S_{yy} - S_{xx}\delta)^2 + 4\delta\rho^2 S_{xx}S_{yy}}}{2\rho\sqrt{S_{xx}S_{yy}}}$$

La pendiente de la recta de Deming

Cuando $\rho > 0$

Se cumple que:

$b_v \leq\ b_D$

Demostración

Sea:

$$k = \frac{S_{yy}}{S_{xx}}$$

$k \in \mathbb{R}, k \geq 0$

Reemplazando el valor de k en (12):

$b_v = \sqrt{k}\,\rho$

$$b_D = \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

Considérese la siguiente expresión:

$4k^2\rho^2(\rho^2 - 1) \leq 0$

$4k\rho^2(k\rho^2 - k) \leq 0$

$4k^2\rho^4 - 4k\rho^2(k - \delta + \delta) \leq 0$

$4k^2\rho^4 - 4k\rho^2(k - \delta) - 4k\rho^2\delta \leq 0$

$4k^2\rho^4 - 4k\rho^2(k - \delta) + (k - \delta)^2 \leq (k - \delta)^2 + 4k\delta\rho^2$

$(2k\rho^2 - (k - \delta))^2 \leq (\delta - k)^2 + 4k\delta\rho^2$

Con respecto a la expresión anterior, el lado izquierdo de la inecuación es ≥ 0, puesto que se trata de una expresión elevada al cuadrado. Por su parte el lado derecho también lo es puesto que $k \geq 0$, $\delta \geq 0$.

Por lo expresado, se cumple que:

$2k\rho^2 - (k - \delta) \leq \sqrt{(\delta - k)^2 + 4k\delta\rho^2}$

$2k\rho^2 \leq (k - \delta) + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}$

$$\sqrt{k}\,\rho \leq \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

$\boldsymbol{b_v \leq b_D}$

(los dos términos de la desigualdad son positivos)

Caso 2

Cuando $\rho < 0$:

Teorema 4.1.1

Siendo:

$$b_v = \sqrt{\frac{S_{yy}}{S_{xx}}}\ \rho$$

La pendiente de recta de mínimos cuadrados ordinarios:

$$b_D = \frac{S_{yy} - S_{xx}\delta + \sqrt{(S_{yy} - S_{xx}\delta)^2 + 4\delta\rho^2 S_{xx}S_{yy}}}{2\rho\sqrt{S_{xx}S_{yy}}}$$

La pendiente de la recta de Deming

Cuando $\rho < 0$

Se cumple que:

$b_v \geq\ b_D$

Demostración:

La demostración es similar al teorema 4.1 hasta llegar a:

$$2k\rho^2\ \leq\ (k - \delta) + \sqrt{(\delta - k)^2 +\ 4k\delta\rho^2}$$

En el paso siguiente cuando se dividen ambos miembros por $2\rho\sqrt{k}$

La cantidad es negativa y por lo tanto el signo de la inecuación se invierte, por lo tanto:

$$\sqrt{k}\,\rho \quad \geq\ \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

Nótese que ambos miembros de la inecuación son negativos.

$\boldsymbol{b_v \geq\ b_D}$

(Los dos términos de la desigualdad son negativos)

Caso especial 1

Se estudia el caso particular en el cual:
$\rho \to 0$

Teorema 4.1.2

Si $\rho \to 0$

Y además:

$k - \delta < 0$

Entonces:

$b_v = b_D = 0$

Demostración:

$$b_{\boldsymbol{v}} = \lim_{\rho \to 0} \sqrt{\frac{S_{yy}}{S_{xx}}}\ \rho$$

$$b_v = \sqrt{\frac{S_{yy}}{S_{xx}}}\ (0)$$

$$b_v = 0$$

En seguida se demuestra que:

$b_D = 0,$ cuando $k - \delta < 0$

$$b_D = \lim_{\rho \to 0} \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

$$b_D = \lim_{\rho \to 0} \frac{\left(k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}\right)\left(k - \delta - \sqrt{(\delta - k)^2 + 4k\delta\rho^2}\right)}{2\rho\sqrt{k}\left(k - \delta - \sqrt{(\delta - k)^2 + 4k\delta\rho^2}\right)}$$

$$b_D = \lim_{\rho \to 0} \frac{-4k\delta\rho}{2\sqrt{k}\left(k - \delta - \sqrt{(\delta - k)^2 + 4k\delta\rho^2}\right)}$$

$$b_D = 0$$

También puede demostrarse que:

$b_D \to -\infty$ en el caso $\rho \to 0^-$, cuando $k - \delta > 0$

$b_D \to \infty$ en el caso $\rho \to 0^+$, cuando $k - \delta > 0$

Caso especial 2

Cuando $\delta \to \infty$

Teorema 4.1.3

Si $\delta \to \infty$

Entonces:

$b_v = b_D$

Demostración:

Cuando $\delta \to \infty$

$$b_D = \lim_{\delta \to \infty} \frac{k - \delta + \sqrt{(\delta - k)^2 + 4k\delta\rho^2}}{2\rho\sqrt{k}}$$

Si se multiplica el numerador y el denominador por la conjugada del numerador:

$$b_D = \lim_{\delta \to \infty} \frac{-4k\delta\rho}{2\sqrt{k}\left(k - \delta - \sqrt{(\delta - k)^2 + 4k\delta\rho^2}\right)}$$

Dividiendo numerador y denominador por δ:

$$b_D = \lim_{\delta \to \infty} \frac{-2k\rho}{\sqrt{k}\left(\frac{k}{\delta} - 1 - \sqrt{\left(1 - \frac{k}{\delta}\right)^2 + \frac{4k\rho^2}{\delta}}\right)}$$

$$b_D = \frac{-2k\rho}{\sqrt{k}\left(0 - 1 - \sqrt{(1 - 0)^2 + 0}\right)}$$

$$b_D = \sqrt{k}\,\rho$$

Que es justamente b_v

Resumen:

En general puede concluirse que si se comparan las pendientes de las rectas de Deming y de mínimos cuadrados ordinarios:

$|\boldsymbol{b_v}| \leq |\boldsymbol{b_D}|$

4.2 Comparación entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de los mínimos cuadrados horizontales

Se va a considerar en esta sección, una recta de regresión de mínimos cuadrados horizontales, este tipo de recta no considera la existencia de varianza en los errores de "y". Se establecerá cuál de las dos rectas de regresión (Deming o mínimos cuadrados horizontales) posee en general mayor valor absoluto en su valor de pendiente.

Llámese b_h a la pendiente de la recta de mínimos cuadrados horizontales, está pendiente se logra cuando se hace $\delta = 0$ en la pendiente de la recta de regresión de Deming.

b_h=$\frac{S_{yy}}{S_{xy}}$

Pero:

$$S_{xy} = \sqrt{S_{xx}S_{yy}}\ \rho$$

Reemplazando este valor:

$$b_h = \frac{1}{\rho}\sqrt{\frac{S_{yy}}{S_{xx}}} \qquad \mathbf{(13)}$$

Haciendo:

$$k = \frac{S_{yy}}{S_{xx}}$$

$$S_{yy} = kS_{xx}$$

Reemplazando este valor en (13):

$$b_h = \frac{\sqrt{k}}{\rho}$$

Caso 1

Cuando $\rho > 0$

Teorema 4.2

Si $\rho > 0$

$\boldsymbol{b_h} \geq \boldsymbol{b_D}$

Demostración

Considérese la siguiente expresión:

$4\delta k(1-\rho^2) \geq 0$

$4\delta k - 4\delta k\rho^2 \ \geq \ \ 0$

$4\delta k \ \geq \ 4\delta k\rho^2$

$2\delta k \ \geq \ -2\delta k + 4\delta k\rho^2$

$k^2 + 2\delta k + \delta^2 \geq \ k^2 - 2\delta k + \delta^2 + 4\delta k\rho^2$

$(k+\delta)^2 \ \geq \ (\delta - k)^2 \ + 4\delta k\rho^2$

En la inecuación anterior, el lado izquierdo es positivo se trata de una suma de dos cantidades positivas al cuadrado, el lado derecho también lo es, puesto que $k \geq 0$, $\delta \geq 0$.

Por lo tanto se cumple que:

$k + \delta \ \geq \ \sqrt{(\delta - k)^2 \ + 4\delta k\rho^2}$

$2k - k + \delta \ \ \geq \ \sqrt{(\delta - k)^2 \ + 4\delta k\rho^2}$

$2k \ \ \geq k - \delta + \ \sqrt{(\delta - k)^2 \ + 4\delta k\rho^2}$

$\frac{\sqrt{k}}{\rho} \ \geq \ \frac{k-\delta+\ \sqrt{(\delta-k)^2 \ +4\delta k\rho^2}}{2\rho\sqrt{k}}$

$\boldsymbol{b_h} \ \geq \ \boldsymbol{b_D}$

(Los dos términos de la desigualdad son positivos)

Caso 2

Cuando $\rho \ < 0$:

Teorema 4.2.1

Si $\rho \ < 0$, entonces

$\boldsymbol{b_h} \leq \ \boldsymbol{b_D}$

Demostración

La demostración es similar a la anterior hasta la línea:

$2k \ \ \geq k - \delta + \ \sqrt{(\delta - k)^2 \ + 4\delta k\rho^2}$

En el paso siguiente cuando se dividen ambos miembros por $2\rho\sqrt{k}$,la cantidad es negativa y por lo tanto el signo de la inecuación se invierte por lo tanto:

$$\frac{\sqrt{k}}{\rho} \leq \frac{k-\delta+\sqrt{(\delta-k)^2+4\delta k\rho^2}}{2\rho\sqrt{k}}$$

$\boldsymbol{b_h} \leq \boldsymbol{b_D}$

(Los dos términos de la desigualdad son negativos)

Caso especial 1

Cuando
$\rho = 0$

Aplicando este valor a las respectivas expresiones de las pendientes, puede deducirse que:

$b_h \rightarrow -\infty$ cuando $\rho \rightarrow 0^-$

$b_h \rightarrow \infty$ cuando $\rho \rightarrow 0^+$

Anteriormente se había demostrado que:

$b_D = 0,$ cuando $k - \delta < 0$

Además:

$b_D \rightarrow -\infty$ en el caso $\rho \rightarrow 0^-$, cuando $k - \delta > 0$
$b_D \rightarrow \infty$ en el caso $\rho \rightarrow 0^+$, cuando $k - \delta > 0$

Caso especial 2

$b_h = b_D$

Cuando $\delta = 0$

Demostración:

$$b_D = \frac{k - 0 + \sqrt{(0-k)^2 + 4k(0)\rho^2}}{2\rho\sqrt{k}}$$

$$b_D = \frac{\sqrt{k}}{\rho}$$

$\boldsymbol{b_h} = \boldsymbol{b_D}$

En general puede concluirse que si se comparan las pendientes de las rectas de Deming y de mínimos cuadrados horizontales:

$$|\boldsymbol{b_h}| \geq |\boldsymbol{b_d}|$$

4.3 Diferencias angulares entre la recta de mínimos cuadrados ortogonales (recta de Deming con $\boldsymbol{\delta} = \mathbf{1}$) y recta de mínimos cuadrados ordinarios (recta de Deming con $\boldsymbol{\delta} \to \infty$)

En esta parte del estudio, se deduce la expresión para calcular la tangente del ángulo que forman la recta de regresión de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios. Nuevamente se recurre a los fundamentos de la geometría analítica. El resultado se expresa en función de k (S_{yy}/S_{xx}) y del coeficiente de correlación de Pearson (ρ).

La tangente del ángulo φ que forman dos rectas con pendientes λ_1 y λ_2 es:

$$tg(\varphi) = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 \lambda_2}$$

La pendiente de la recta de mínimos cuadrados ordinarios es:

$$\sqrt{k}\rho$$

Por su parte la pendiente de la recta de mínimos cuadrados ortogonales es:

$$\frac{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}{2\sqrt{k}\rho}$$

Si θ es el ángulo que forman las dos rectas:

$$tg(\theta) = \frac{\dfrac{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}{2\sqrt{k}\rho} - \sqrt{k}\rho}{1 + \left(\dfrac{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}{2\sqrt{k}\rho}\right)\sqrt{k}\rho} \qquad (\mathbf{14})$$

Tangente del ángulo $\boldsymbol{\theta}$, cuando $\boldsymbol{k} = \mathbf{0}$.

A continuación se determina el valor de la tangente del ángulo que forman la recta de regresión de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios en el caso extremo en que $k = \frac{S_{yy}}{S_{xx}} = 0$.

Teorema 4.3

Sea:

$$\sqrt{k}\rho$$

La pendiente de la recta de los mínimos cuadrados ordinarios y:

$$\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

La pendiente de la recta de mínimos cuadrados ortogonales.

Si $k=0$, entonces la tangente del ángulo formado por las dos rectas es cero.

Demostración:

Es inmediato en este caso (k=0), que la pendiente de la recta de los mínimos cuadrados ordinarios es cero.

Se requiere encontrar el valor de la pendiente de la recta de los mínimos cuadrados ortogonales cuando (k=0):

$$\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

Multiplicando está pendiente por el conjugado del numerador resulta:

$$\frac{\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)\left(k-1-\sqrt{(k-1)^2+4k\rho^2}\right)}{2\sqrt{k}\rho\left(k-1-\sqrt{(k-1)^2+4k\rho^2}\right)}$$

Finalmente se obtiene:

$$\frac{-2\sqrt{k\rho}}{k-1-\sqrt{(k-1)^2+4k\rho^2}}$$

Al hacer k=0, la pendiente es cero.

Por lo tanto:

$$tg(\theta)=\frac{0-0}{1+(0)(0)}=0$$

Conclusión: cuando k=0, la tangente del ángulo formado por las rectas de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales es cero.

Tangente del ángulo θ, cuando $k \to \infty$.

En esta parte del estudio se determina el valor de la tangente del ángulo que forman la recta de regresión de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios en el caso extremo en que $k \to \infty$.

Teorema 4.3.1

Sea:

$\sqrt{k}\rho$

La pendiente de la recta de los mínimos cuadrados ordinarios y:

$$\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

La pendiente de la recta de mínimos cuadrados ortogonales.

Si k → ∞, entonces la tangente del ángulo formado por las dos rectas es cero.

Demostración:

Realizando desarrollo algebraico en la expresión (14) de la tangente del ángulo formado por las dos rectas, se llega fácilmente a:

$$tg(\theta)=\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}-2k\rho^2}{\left(k+1+\sqrt{(k-1)^2+4k\rho^2}\right)\sqrt{k}\rho}$$

Aplicando límite cuando k → ∞:

$$tg(\theta)=\lim_{k\to\infty}\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}-2k\rho^2}{\left(k+1+\sqrt{(k-1)^2+4k\rho^2}\right)\sqrt{k}\rho}$$

Dividiendo numerador y denominador por k:

$$tg(\theta)=\lim_{k\to\infty}\frac{1-1/k+\sqrt{(1-1/k)^2+4\rho^2/k}-2\rho^2}{\left(1+1/k+\sqrt{(1-1/k)^2+4\rho^2/k}\right)\sqrt{k}\rho}$$

$$tg(\theta)=0,\quad \rho\neq 1.$$

Conclusión: cuando k tiende a infinito, la tangente del ángulo θ es cero.

Determinación del valor de k que produce el máximo ángulo entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios.

En esta sección se encuentra un interesante resultado, consistente en que existe un valor de $k = S_{yy}/S_{xx}$ para el cual la diferencia angular entre la recta de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios tiene en valor absoluto un máximo. Se determina ese valor k. Finalmente se establece a partir del valor de dicho ángulo.

Teorema 4.3.2

Siendo:

$$tg(\theta) = \frac{\dfrac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho} - \sqrt{k}\rho}{1+\left(\dfrac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}\right)\sqrt{k}\rho}$$

La tangente del ángulo formado entre la recta de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales, se producen máximos y mínimos relativos en:

$x = 3$

Cuando $\rho > 0$
Máximo relativo.

Cuando $\rho < 0$
Mínimo Relativo

Demostración:

A la pendiente de la recta de los mínimos cuadrados ortogonales se le llama W y a la recta de los mínimos cuadrados ordinarios se le llama Z (Estos reemplazos se realizan para facilitar los cálculos).

$$W = \frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho} \qquad (\mathbf{15})$$

$$Z = \sqrt{k}\rho \qquad (\mathbf{16})$$

$$tg(\theta) = \frac{W-Z}{1+WZ}$$

La tangente del ángulo θ es función de la variable k, entonces:

$$tg(\theta) = F(k)$$

Se encontrará el valor de k para el cual la diferencia angular es máxima, manteniendo el coeficiente de correlación de Pearson (ρ) constante.

Para ello se deriva la función con respecto a la variable k y se iguala a cero:

$$F'(k) = \frac{\left(\frac{dW}{dk} - \frac{dZ}{dk}\right)(1+WZ) - \left(\frac{dW}{dk}Z + \frac{dZ}{dk}W\right)(W-Z)}{(1+WZ)^2}$$

Un trabajo algebraico de la anterior expresión conduce a la siguiente expresión de primera derivada igualada a cero:

$$F'(k) = \frac{\frac{dW}{dk}(1+Z^2) - \frac{dZ}{dk}(1+W^2)}{(1+WZ)^2} = 0$$

$$\frac{dZ}{dk}(1+W^2) \; = \; \frac{dW}{dk}(1+Z^2) \qquad (\mathbf{17})$$

Es inmediato que derivando (15):

$$\frac{dZ}{dk} = \frac{\rho}{2\sqrt{k}} \qquad (\mathbf{18})$$

Es más laborioso encontrar $\frac{dW}{dk}$, sin embargo aplicando procedimientos de derivación a la expresión (16), se llega a:

$$\frac{dW}{dk} = \frac{1}{2\rho k}\left(\left(1 + \frac{k-1+2\rho^2}{\sqrt{(k-1)^2+4k\rho^2}}\right)\sqrt{k} - \frac{1}{2\sqrt{k}}\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)\right) \qquad (\mathbf{19})$$

También por desarrollo, se puede afirmar que:

$$W^2 = \frac{1}{2k\rho^2}\left((k-1)^2 + (k-1)\sqrt{(k-1)^2+4k\rho^2} + 2k\rho^2\right) \qquad (\mathbf{20})$$

Previamente, por (16) se sabía que:

$$Z = \sqrt{k}\rho$$

Reemplazando los valores: (16), (18), (19) y (20) en la ecuación (17):

$$\frac{\rho}{2\sqrt{k}}\left(1 + \frac{1}{2k\rho^2}\left((k-1)^2 + (k-1)\sqrt{(k-1)^2+4k\rho^2} + 2k\rho^2\right)\right) =$$

$$\frac{1}{2\rho k}\left(\left(1+\frac{k-1+2\rho^2}{\sqrt{(k-1)^2+4k\rho^2}}\right)\sqrt{k}-\frac{1}{2\sqrt{k}}\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)\right)(1+k\rho^2)$$

Realizando algunas reducciones:

$$\left(4k\rho^2+(k-1)\right)\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)=(1+k\rho^2)\left(\frac{(k+1)\left(\sqrt{(k-1)^2+4k\rho^2}+k-1\right)}{\sqrt{(k-1)^2+4k\rho^2}}\right)$$

Finalmente se llega a:

$$\left(4k\rho^2+(k-1)^2-(1+k\rho^2)(k+1)\right)\left(\sqrt{(k-1)^2+4k\rho^2}+k-1\right)=0 \qquad \textbf{(21)}$$

Tomando el primer factor de (21):

$$4k\rho^2+(k-1)^2-(1+k\rho^2)(k+1)=0$$

Realizando operaciones:

$$k(3\rho^2+k-3-k\rho^2)=0$$

$$\left(3(\rho^2-1)-k(\rho^2-1)\right)k=0$$

$$(\rho^2-1)(3-k)k=0$$

Se obtienen las soluciones:

$k=3$
$k=0$

Tomando el segundo factor de (21):

$$\sqrt{(k-1)^2+4k\rho^2}+k-1=0$$

Se obtiene la solución:

$k=0$

Criterio de segunda derivada para definir máximos o mínimos relativos.

El desarrollo de la segunda derivada es laborioso y por eso no se incluye en este estudio, pero puede demostrarse que:

Cuando $\rho>0$
$F''(3)<0$
Máximo relativo.

Cuando $\rho < 0$
$F''(3) > 0$
Mínimo Relativo

El valor del ángulo para $k = 0$, ya se analizó al inicio de este capítulo. En este punto no hay continuidad puesto que no se puede establecer limite por la izquierda, teniendo en cuenta, que $k \geq 0$.

Corolario 4.3.3

Considérese el valor crítico encontrado en $k = 3$:

Reemplazando el valor $k = 3$ en la expresión de la tangente del ángulo formado por las dos rectas (14), se obtiene la siguiente expresión de diferencia angular máxima para $\rho > 0$ y mínima para $\rho < 0$.

$$tg(\theta) = \frac{\sqrt{3\rho^2 + 1} - 3\rho^2 + 1}{\sqrt{3}\,\rho\left(\sqrt{3\rho^2 + 1} + 2\right)}$$

$$\theta = arc.tg\left(\frac{\sqrt{3\rho^2 + 1} - 3\rho^2 + 1}{\sqrt{3}\,\rho\left(\sqrt{3\rho^2 + 1} + 2\right)}\right) \qquad (\mathbf{22})$$

El anterior es el máximo valor o mínimo valor del ángulo formado entre la recta de mínimos cuadrados ortogonales y la recta de los mínimos cuadrados ordinarios.

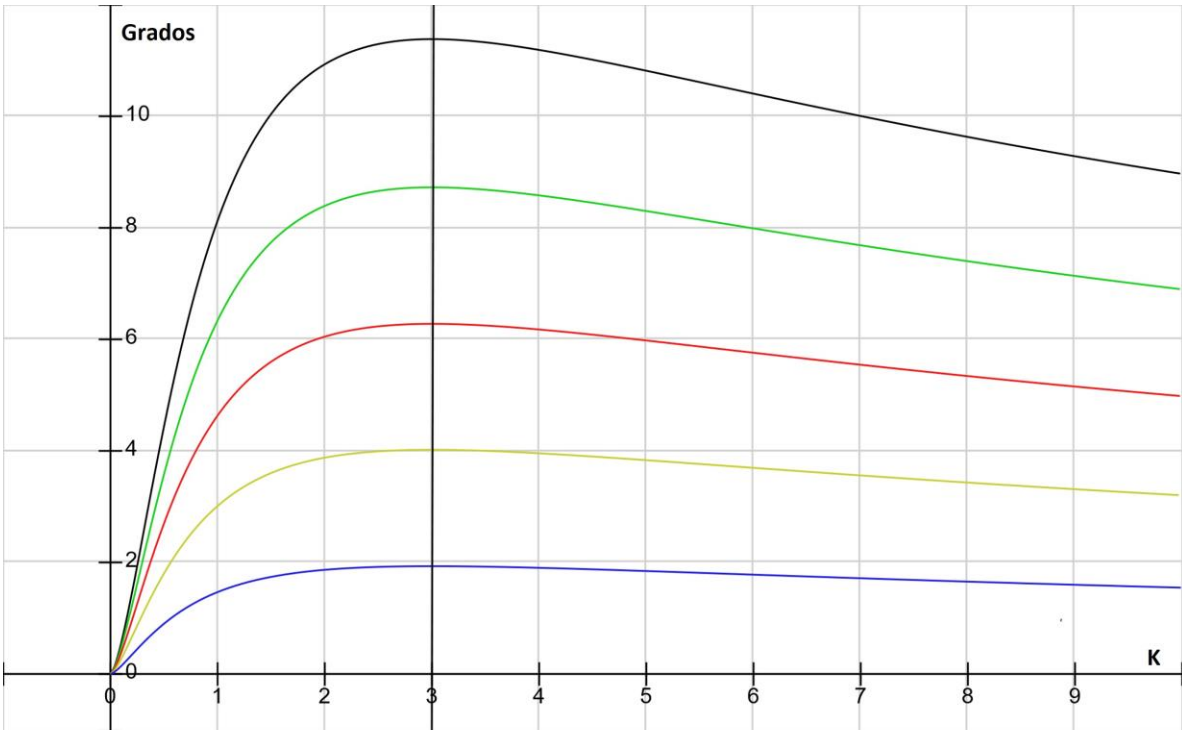
Los valores máximo y mínimo, son iguales en valor absoluto, la anterior afirmación es consecuencia de la simple inspección de la última expresión.

Si θ es positivo el ángulo se mide desde la recta de mínimos cuadrados ordinarios hasta la recta de mínimos cuadrados ortogonales en sentido anti horario.

Si θ es negativo el ángulo se mide desde la recta de mínimos cuadrados ordinarios hasta la recta de mínimos cuadrados ortogonales en sentido horario.

En general puede afirmarse que el valor del ángulo tomado en valor absoluto es máximo para $k = 3$.

Grafica 3. Ángulos en grados entre la recta de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios para valores de k



Se aprecia lo demostrado, el mayor ángulo entre las rectas de regresión de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales se produce cuando k=3. Se han empleado distintos valores positivos del coeficiente de correlación de Pearson. Las abscisas corresponden al valor de k y las ordenadas al ángulo en grados.

Línea negra: ρ = 0,75, Línea verde: ρ = 0.80, Línea roja: ρ = 0.85, Línea amarilla: ρ = 0.90, Línea azul: ρ = 0.95

Grafica 4. Ángulos entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios cuando k=3, en función del coeficiente de correlación de Pearson

En las abscisas aparecen los valores del coeficiente de correlación de Pearson y en las ordenadas los ángulos en grados

4.4 Diferencias angulares entre la recta de mínimos cuadrados horizontales (recta de Deming con $\delta = 0$) y recta de mínimos cuadrados ortogonales (recta de Deming con $\delta \to \infty$)

En este aparte, se deduce la expresión para calcular la tangente del ángulo que forman la recta de regresión de mínimos cuadrados horizontales y la recta de regresión de Mínimos cuadrados ortogonales. Se aplican conceptos básicos de la geometría analítica. El resultado se expresa en función de k (S_{yy}/S_{xx}) y del coeficiente de correlación de Pearson (ρ).

La pendiente de la recta de mínimos cuadrados horizontales es:

$$\sqrt{k}/\rho$$

Por su parte la pendiente de la recta de mínimos cuadrados ortogonales es:

$$\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

Si θ es el ángulo que forman las dos rectas:

$$tg(\theta) = \frac{\frac{\sqrt{k}}{\rho} - \frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}}{1+\frac{\sqrt{k}}{\rho}\left(\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}\right)} \qquad (\mathbf{23})$$

Tangente del ángulo cuando $k = 0$

Es inmediato , en este caso ($k = 0$), que la pendiente de la recta de los mínimos cuadrados horizontales es cero.

Ya se había demostrado anteriormente que en el caso (k=0), la pendiente de la recta de mínimos cuadrados ortogonales es cero.

Por lo tanto:

$$tg(\theta) = \frac{0-0}{1+(0)(0)} = 0$$

Conclusión: cuando k=0, el ángulo formado por las rectas de mínimos cuadrados horizontales y mínimos cuadrados ortogonales es cero.

Tangente del ángulo θ, cuando k $\to \infty$:

A continuación se determina el valor de la tangente del ángulo que forman la recta de regresión de mínimos cuadrados horizontales y mínimos cuadrados ortogonales en el caso extremo en que $k \to \infty$.

Teorema 4.4

Sea:

$\sqrt{k}/\rho$

La pendiente de la recta de los mínimos cuadrados ordinarios y:

$$\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

La pendiente de la recta de mínimos cuadrados ortogonales.

Si $k \to \infty$, entonces la tangente del ángulo formado por las dos rectas es cero.

Demostración

Realizando trabajo algebraico en la expresión (23), se llega a:

$$tg(\theta) = \frac{\left(k+1-\sqrt{(k-1)^2+4k\rho^2}\right)\rho}{\left(2\rho^2+k-1+\sqrt{(k-1)^2+4k\rho^2}\right)\sqrt{k}}$$

Aplicando límite cuando k → ∞:

$$tg(\theta) = \lim_{k\to\infty} \frac{\left(k+1-\sqrt{(k-1)^2+4k\rho^2}\right)\rho}{\left(2\rho^2+k-1+\sqrt{(k-1)^2+4k\rho^2}\right)\sqrt{k}}$$

Dividiendo numerador y denominador por k:

$$tg(\theta) = \lim_{k\to\infty} \frac{\left(1+1/k-\sqrt{(1-1/k)^2+4\rho^2/k}\right)\rho}{\left(2\rho^2/k+1-1/k+\sqrt{(1-1/k)^2+4\rho^2/k}\right)\sqrt{k}}$$

$$tg(\theta) = 0$$

Por lo tanto, cuando k tiende a infinito, la tangente del ángulo θ es cero.

Determinación del valor de k que produce el máximo ángulo entre la recta de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales.

En este aparte se encuentra un llamativo resultado, consistente en que existe un valor de $k = S_{yy}/S_{xx}$ para el cual la diferencia angular entre la recta de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales tiene en valor absoluto un máximo. Se determina ese valor k. Finalmente se establece a partir del valor de dicho ángulo.

Teorema 4.4.1

Siendo:

$$tg(\theta) = \frac{\frac{\sqrt{k}}{\rho} - \frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}}{1+\frac{\sqrt{k}}{\rho}\left(\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}\right)}$$

La tangente del ángulo formado entre la recta de mínimos cuadrados horizontales y mínimos cuadrados ortogonales, se producen máximos y mínimos relativos en:

$x = 1/3$

Cuando $\rho > 0$
Máximo relativo.

Cuando $\rho < 0$
Mínimo Relativo

Demostración

Previamente se había denominado W a la pendiente de la recta de los mínimos cuadrados ortogonales (15), a la recta de los mínimos cuadrados horizontales se le llama Q:

$$Q = \frac{\sqrt{k}}{\rho} \qquad \mathbf{(24)}$$

$$tg(\theta) = \frac{Q-W}{1+QW}$$

La tangente del ángulo θ es función de la variable k, entonces:

$tg(\theta) = F(k)$

Se encuentra el valor de k para el cual la diferencia angular es máxima, manteniendo ρ constante.

Para ello se deriva con respecto a la variable k y se iguala a cero:

$$F'(k) = \frac{\left(\frac{dQ}{dk} - \frac{dW}{dk}\right)(1 + QW) - \left(\frac{dQ}{dk}W + \frac{dW}{dk}Q\right)(Q - W)}{(1 + QW)^2} = 0$$

Un trabajo algebraico a la anterior expresión conduce a:

$$\frac{dQ}{dk}(1 + W^2) - \frac{dW}{dk}(1 + Q^2) = 0 \qquad (\mathbf{25})$$

Al derivar (24) se obtiene:

$$\frac{dQ}{dk} = \frac{1}{2\sqrt{k}\rho} \qquad (\mathbf{26})$$

También de (24):

$$Q^2 = \frac{k}{\rho^2} \qquad (\mathbf{27})$$

Por (19), se sabe que:

$$\frac{dW}{dk} = \frac{1}{2\rho k}\left(\left(1 + \frac{k - 1 + 2\rho^2}{\sqrt{(k-1)^2 + 4k\rho^2}}\right)\sqrt{k} - \frac{1}{2\sqrt{k}}\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right)\right)$$

También, por (20), se conoce que:

$$W^2 = \frac{1}{2k\rho^2}\left((k-1)^2 + (k-1)\sqrt{(k-1)^2 + 4k\rho^2} + 2k\rho^2\right)$$

Reemplazando: (19), (20), (26) y (27) en (25) y realizando algunas reducciones:

$$\frac{1}{2\sqrt{k}\rho}\left(1 + \frac{1}{2k\rho^2}\left((k-1)^2 + (k-1)\sqrt{(k-1)^2 + 4k\rho^2} + 2k\rho^2\right)\right) =$$

$$\frac{1}{2\rho k}\left(\left(1 + \frac{k - 1 + 2\rho^2}{\sqrt{(k-1)^2 + 4k\rho^2}}\right)\sqrt{k} - \frac{1}{2\sqrt{k}}\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right)\right)\left(1 + \frac{k}{\rho^2}\right)$$

Trabajando algebraicamente, la ecuación, se llega a:

$$\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)(1-3k)=0$$

Las soluciones son:

$k=0$
$k=1/3$

Criterio de segunda derivada para definir máximos o mínimos relativos.

El desarrollo de la segunda derivada es laborioso y por eso no se incluye en este estudio, pero puede demostrarse que:

Cuando $\rho>0$
$F''(1/3)<0$

Máximo relativo.

Cuando $\rho<0$
$F''(1/3)>0$

Mínimo relativo.

El valor del ángulo para $k=0$, ya se analizó al inicio de este capítulo. En este punto no hay continuidad puesto que no se puede establecer límite por la izquierda, considerando que $k\geq 0$.

Corolario 4.4.2

Reemplazando el valor $k=1/3$ en la expresión del ángulo formado por las dos rectas se obtiene la siguiente expresión de desfase angular máximo:

$$\theta=arc.tg\left(\frac{\sqrt{3}\,\rho\left(2-\sqrt{1+3\rho^2}\right)}{3\rho^2-1+\sqrt{1+3\rho^2}}\right)$$

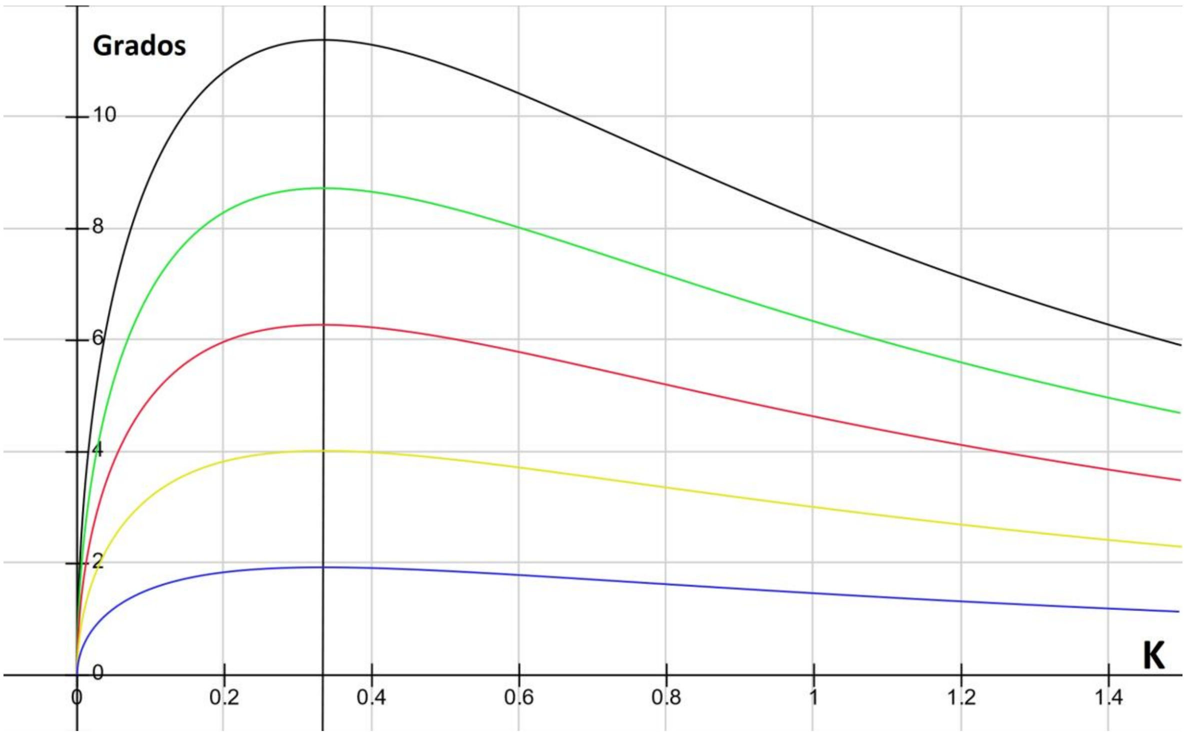
El anterior es el máximo valor del ángulo formado entre la recta de mínimos cuadrados horizontales y la recta de los mínimos ortogonales.

Si θ es positivo el ángulo se mide desde la recta de mínimos cuadrados ortogonales hasta la recta de mínimos cuadrados horizontales en sentido anti horario.

Si θ es negativo el ángulo se mide desde la recta de mínimos cuadrados ortogonales hasta la recta de mínimos cuadrados horizontales en sentido horario.

En general puede afirmarse que el ángulo tomado en valor absoluto es máximo para $k=1/3$.

Grafica 5. Ángulos en grados entre la recta de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales para valores de k.



Se aprecia lo demostrado, el mayor ángulo entre las rectas de regresión de mínimos cuadrados horizontales y mínimos cuadrados ortogonales se produce cuando k=1/3. Se han empleado distintos valores positivos del coeficiente de correlación de Pearson. Las abscisas corresponden al valor de k y las ordenadas al ángulo en grados. Línea negra: ρ = 0,75,Línea verde: ρ = 0.80, Línea roja: ρ = 0.85, Línea amarilla: ρ = 0.90, Línea azul: ρ = 0.95

Propiedad 1. Se obtuvieron las expresiones para obtener los ángulos máximos (desfases) para dos situaciones: a) formado por la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios b) formado por la recta de mínimos cuadrados horizontales y mínimos cuadrados ortogonales. Es interesante notar que cuando se racionalizan los denominadores de las expresiones, se descubre que esas diferencias máximas (en el primer caso cuando k=3 y en el segundo cuando k=1/3) son iguales y se expresan de la siguiente manera:

$$\theta = arc.tg\left(\frac{\sqrt{3}\left(9\rho^2 - 1 - (3\rho^2+1)^{3/2}\right)}{9\rho(\rho^2-1)}\right)$$

Propiedad 2. El ángulo formado entre la recta de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios es igual al ángulo formado entre la recta de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales cuando $k = 1$.

Teorema 4.4.3

El ángulo formado entre la recta de mínimos cuadrados ortogonales y la recta de mínimos cuadrados ordinarios es igual al ángulo formado entre la recta de mínimos cuadrados horizontales y la recta de mínimos cuadrados ortogonales cuando $k = 1$.

Demostración:

Se demuestra planteando la igualdad entre las expresiones (14) y (23)

$$\frac{\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}-\sqrt{k}\rho}{1+\left(\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}\right)\sqrt{k}\rho}=\frac{\frac{\sqrt{k}}{\rho}-\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}}{1+\frac{\sqrt{k}}{\rho}\left(\frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}\right)}$$

El desarrollo algebraico de la ecuación entrega como resultado:

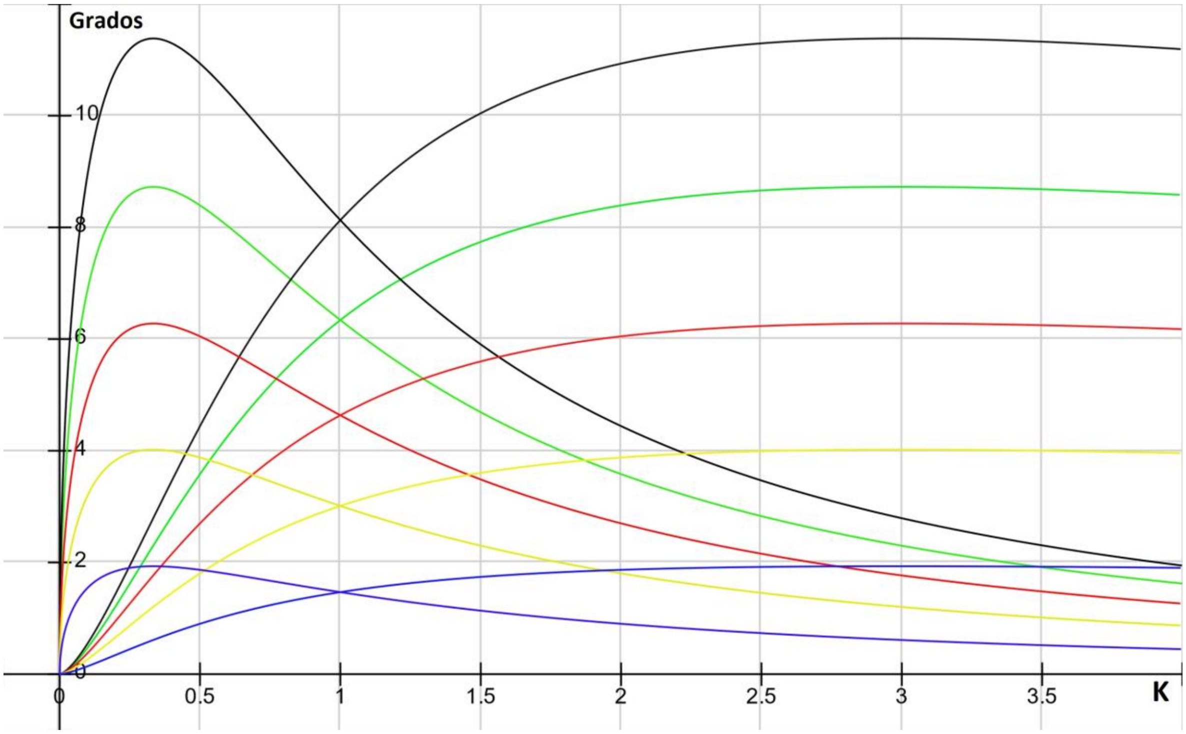
$k = 1$

Y para ese valor se obtiene un ángulo, en los dos casos de:

$arc.tg\left(\frac{1-\rho}{1+\rho}\right)$
Para $\rho \geq 0$

$arc.tg\left(\frac{1+\rho}{\rho-1}\right)$
Para $\rho < 0$

Grafica 6. Resumen de ángulos entre rectas de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios, también de ángulos entre rectas de mínimos cuadrados horizontales y mínimos cuadrados ortogonales.



Se aprecian los ángulos en grados entre rectas de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios, con valores máximos en k=3. También se observan las diferencias entre rectas de mínimos cuadrados horizontales y mínimos cuadrados ortogonales con valores máximos en k=1/3. Nótese que cuando k=1, el ángulo es igual en los dos casos estudiados, se han empleado valores $\rho \geq 0$, de la siguiente manera: Línea negra: ρ = 0,75, Línea verde: ρ = 0.80, Línea roja: ρ = 0.85, Línea amarilla: ρ = 0.90, Línea azul: ρ = 0.95. Las abscisas corresponden a los valores de k y las ordenadas a los ángulos medidos en grados.

COMPARACIÓN DE COEFICIENTES DE DETERMINACIÓN R^2 DE LAS RECTAS DE REGRESIÓN DE MÍNIMOS CUADRADOS ORDINARIOS Y MÍNIMOS CUADRADOS ORTOGONALES

Si se considera que las dos modalidades más empleadas en la familia de regresión de Deming son la recta de mínimos cuadrados ordinarios y la recta de mínimos cuadrados ortogonales, se justifica realizar un estudio comparativo completo y exhaustivo sobre los coeficientes de determinación respectivos.

Se deducirán las expresiones de SCT y SCE para las dos rectas de regresión, con base a lo anterior se encuentran los coeficientes de determinación para cada una de ellas.

Posteriormente se demuestra que la diferencia entre el coeficiente de determinación de la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios es siempre mayor o igual a cero.

Se establece una función $F(\rho)$ para las diferencias de coeficiente de determinación entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios en función del coeficiente de correlación de Pearson, y se determinan con minuciosidad el valor de la función para cada circunstancia de correlación de, varianzas y covarianzas de las variables. Definiéndose máximos y mínimos y presentándose en un cuadro resumen.

5.1 SCE en las rectas de regresión de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales.

La SCE en las rectas de regresión que se consideran en el presente capítulo no son solamente una métrica sino un insumo preliminar para la determinación ulterior del coeficiente de determinación y la comparación de estos valores en los dos casos.

Suma de cuadrados de errores en una recta de regresión de mínimos cuadrados ordinarios

Teorema 5.1

Siendo $P_i(x_i, y_i)$; $i = 1,2,3\ldots,n$ un conjunto de puntos, la suma de cuadrados de errores en una recta de mínimos cuadrados ordinarios es:

$$SCE_v = S_{yy} - \frac{\left(S_{xy}\right)^2}{S_{xx}}$$

Demostración

Considérese nuevamente la recta de regresión:

$$y = a + bx$$

La cual se estableció a partir de un conjunto de n puntos.

La suma de cuadrados de errores verticales (SCE_v) de la Recta de mínimos cuadrados ordinarios se determina de la siguiente manera:

$$SCE_v = \sum (y_i - \hat{y}_i)^2$$

$$SCE_v = \sum (y_i - a - bx_i)^2$$

En todas las regresiones lineales de mínimos cuadrados, la recta pasa por el centroide de los puntos:

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

Por lo tanto:

$$SCE_v = \sum \left((y_i - \overline{y}) - b(x_i - \overline{x})\right)^2$$

$$SCE_v = \sum ((y_i - \overline{y})^2 - 2b(y_i - \overline{y})(x_i - \overline{x}) + b^2(x_i - \overline{x})^2)$$

$$SCE_v = S_{yy} - 2bS_{xy} + b^2 S_{xx}$$

Cuando se usan mínimos cuadrados ordinarios**:**

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

Por lo tanto:

$$SCE_v = S_{yy} - \frac{\left(S_{xy}\right)^2}{S_{xx}} \qquad \mathbf{(28)}$$

Suma de cuadrados de errores en una recta de regresión de mínimos cuadrados ortogonales.

Teorema 5.1.1

Siendo $P_i(x_i, y_i);\ i = 1,2,3 \dots, n$ un conjunto de puntos, la suma de cuadrados de errores en una recta de mínimos cuadrados ortogonales es:

$$SCE_o = \frac{S_{xx} + S_{yy} - \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2}$$

Demostración:

La suma de cuadrados de errores ortogonales (SCE_o) de la Recta de Mínimos Cuadrados Ortogonales se determina de la siguiente manera:

Del teorema 5.1, conocida la expresión de SCE_v como función de b bastará con dividir el resultado por la secante al cuadrado del ángulo de inclinación de la recta de regresión (dicha secante al cuadrado tiene por valor ($b^2 + 1$)):

$$SCE_o = \frac{S_{yy} - 2bS_{xy} + b^2 S_{xx}}{b^2 + 1}$$

Pero en el caso de recta de mínimos cuadrados ortogonales:

$$b = \frac{S_{yy} - S_{xx} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2S_{xy}}$$

Reemplazando este valor y luego de realizar reducciones se llega a:

$$SCE_o = \frac{S_{xx} + S_{yy} - \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2} \qquad (\mathbf{29})$$

5.2 R^2 Recta de mínimos cuadrados ordinarios y recta de mínimos cuadrados ortogonales.

En general, en las rectas de regresión, el coeficiente de determinación se expresa de la siguiente manera:

$$R^2 = 1 - \frac{SCE}{SCT}$$

Donde SCE es la suma de los cuadrados de los errores y SCT es la suma de los cuadrados totales.

Coeficiente de determinación en recta de mínimos cuadrados ordinarios ($\boldsymbol{R_v^2}$):

Teorema 5.2

El coeficiente de determinación R_v^2 en una recta de mínimos cuadrados ordinarios es:

$$R_v^2 = \frac{\left(S_{xy}\right)^2}{S_{xx}S_{yy}}$$

Demostración:

$$R_v^2 = 1 - \frac{SCE_v}{SCT_v} \qquad (\mathbf{30})$$

$$SCT_v = S_{yy} \qquad (\mathbf{31})$$

Reemplazando (28) y (31) en (30):

$$R_v^2 = 1 - \frac{S_{yy} - (S_{xy})^2/S_{xx}}{S_{yy}}$$

$$R_v^2 = \frac{\left(S_{xy}\right)^2}{S_{xx}S_{yy}}$$

Coeficiente de determinación en recta de mínimos cuadrados ortogonales ($\boldsymbol{R_o^2}$)

Teorema 5.2.2

Coeficiente de determinación en recta de mínimos cuadrados ortogonales (R_o^2) es:

$$R_o^2 = \frac{S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4\left(S_{xy}\right)^2}}{2\left(S_{xx} + S_{yy}\right)}$$

Demostración:

$$R_o^2 = 1 - \frac{SCE_o}{SCT_o} \qquad (\mathbf{32})$$

$$SCT_o = S_{xx} + S_{yy} \qquad (\mathbf{33})$$

Reemplazando (29) y (33) en (32):

$$R_o^2 = 1 - \frac{S_{xx} + S_{yy} - \sqrt{(S_{xx} - S_{yy})^2 + 4(S_{xy})^2}}{2\left(S_{xx} + S_{yy}\right)}$$

$$R_o^2 = \frac{S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4\left(S_{xy}\right)^2}}{2\left(S_{xx} + S_{yy}\right)}$$

5.3 Comparación de R^2 entre las rectas de regresión de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios

Se demuestra en este aparte qué para todos los casos posibles, la diferencia entre los coeficientes de determinación entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios es mayor o igual que cero.

Teorema 5.3

Sea R_o^2 el coeficiente de determinación de la recta de mínimos cuadrados ortogonales y R_v^2 el coeficiente de determinación de la recta de mínimos cuadrados ordinarios, entonces:

$$R_o^2 - R_v^2 \geq 0$$

Demostración:

$$R_o^2 - R_v^2 = \frac{S_{xx}+S_{yy}+\sqrt{(S_{xx}-S_{yy})^2+4\left(S_{xy}\right)^2}}{2\left(S_{xx}+S_{yy}\right)} - \frac{\left(S_{xy}\right)^2}{S_{xx}S_{yy}}$$

Si se reemplaza:

$$(S_{xy})^2 = \rho^2 S_{xx}S_{yy}$$

Donde ρ es el coeficiente de correlación de Pearson:

$$R_o^2 - R_v^2 = \frac{S_{xx}+S_{yy}+\sqrt{\left(S_{xx}-S_{yy}\right)^2+4S_{xx}S_{yy}\rho^2}}{2(S_{xx}+S_{yy})} - \rho^2$$

La expresión también puede escribirse así:

$$\frac{1}{2} + \frac{\sqrt{\left(S_{xx}-S_{yy}\right)^2+4S_{xx}S_{yy}\rho^2}}{2(S_{xx}+S_{yy})} - \rho^2$$

Nótese que:

$$\rho^2 < \frac{1}{2}$$

$$R_o^2 - R_v^2 > 0$$

Se estudia a continuación el caso en que:

$$\rho^2 \geq \frac{1}{2}$$

El numerador de la diferencia de coeficientes de determinación es:

$$\sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xx}S_{yy}\rho^2} - \left(S_{xx} + S_{yy}\right)(2\rho^2 - 1)$$

El denominador no se tiene en cuenta porque es positivo

Realizando trabajo algebraico también puede expresarse así:

$$\sqrt{S_{xx}^2 + S_{yy}^2 + 2S_{xx}S_{yy}(2\rho^2 - 1)} - \left(S_{xx} + S_{yy}\right)(2\rho^2 - 1) \qquad (\mathbf{34})$$

Hágase:

$$z = 2\rho^2 - 1$$

Por lo tanto:

$$0 \leq z \leq 1$$

Reemplazando el valor de z en (34) :

$$\sqrt{S_{xx}^2 + S_{yy}^2 + 2zS_{xx}S_{yy}} - \left(S_{xx} + S_{yy}\right)z$$

Si esta cantidad se multiplica y divide por su conjugada y no se escribe el denominador puesto que es positivo:

$$S_{xx}^2 + S_{yy}^2 + 2zS_{xx}S_{yy} - \left(S_{xx} + S_{yy}\right)^2 z^2$$

$$S_{xx}^2 + S_{yy}^2 + 2zS_{xx}S_{yy} - S_{xx}^2 z^2 - 2z^2 S_{xx}S_{yy} - S_{yy}^2 z^2$$

$$S_{xx}^2(1 - z^2) + S_{yy}^2(1 - z^2) + 2zS_{xx}S_{yy}(1 - z) \geq 0$$

Conclusión:

En todos los casos:

$$R_o^2 - R_v^2 \ \geq 0$$

La recta de regresión mínimos cuadrados ortogonales posee mayor o igual coeficiente de determinación que la recta de regresión de mínimos cuadrados ordinarios.

5.4 Diferencias entre R^2 de las rectas de regresión de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios.

La diferencia entre los coeficientes de determinación entre la recta de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios ($R_o^2 - R_v^2$), se plantea ahora como una función del coeficiente de determinación de Pearson:

$$F(\rho) = \frac{S_{xx} + S_{yy} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xx}S_{yy}\rho^2}}{2\left(S_{xx} + S_{yy}\right)} - \rho^2 \qquad (\mathbf{35})$$

En esta sección y mediante el uso del cálculo diferencial, se procede a determinar los puntos críticos para $F(\rho)$. Durante la demostración del teorema que se enuncia a continuación se definen además los valores correspondientes a dichos puntos críticos. El análisis es exhaustivo

Teorema 5.4

Establecida la diferencia de coeficientes de determinación entre la recta de mínimos cuadrados ortogonales como una función del coeficiente de correlación de Pearson (35):

$$F(\rho) = \frac{S_{xx} + S_{yy} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xx}S_{yy}\rho^2}}{2\left(S_{xx} + S_{yy}\right)} - \rho^2$$

Los puntos críticos de dicha diferencia se encuentran en:

$$\rho = 0$$

$$\rho = \pm \frac{1}{2\left(S_{xx} + S_{yy}\right)} \sqrt{\frac{\left(S_{xx}S_{yy}\right)^2 - \left((S_{xx})^2 - \left(S_{yy}\right)^2\right)^2}{S_{xx}S_{yy}}}$$

Demostración:

Se buscan puntos críticos haciendo:

$$F'(\rho) = 0$$

$$F'(\rho) = \frac{2S_{xx}S_{yy}\rho}{\left(S_{xx} + S_{yy}\right)\sqrt{4S_{xx}S_{yy}\rho^2 + \left(S_{xx} - S_{yy}\right)^2}} - 2\rho = 0$$

Trabajando algebraicamente la expresión, se obtienen puntos críticos en:

$\rho = 0$

$$\rho = \pm \frac{1}{2(S_{xx} + S_{yy})} \sqrt{\frac{(S_{xx}S_{yy})^2 - \left((S_{xx})^2 - (S_{yy})^2\right)^2}{S_{xx}S_{yy}}}$$

Se establece la segunda derivada pare definir si en esos puntos críticos existen máximos y/o mínimos relativos:

$$F^{"}(\rho) = \frac{S_{xx}S_{yy}}{(S_{xx} + S_{yy})\sqrt{4S_{xx}S_{yy}\rho^2 + (S_{xx} - S_{yy})^2}} - \frac{4(S_{xx}S_{yy}\rho)^2}{(S_{xx} + S_{yy})\left(4S_{xx}S_{yy}\rho^2 + (S_{xx} - S_{yy})^2\right)^{3/2}} - 1$$

5.4.1 Estudio de puntos críticos en la diferencia de coeficiente de determinación, cuando $\boldsymbol{\rho = 0}$.

Primer caso: Cuando $\boldsymbol{S_{xx} > S_{yy}}$

$$F^{"}(0) = \frac{S_{xx}S_{yy}}{(S_{xx})^2 - (S_{yy})^2} - 1$$

Si:

$S_{xx}S_{yy} > (S_{xx})^2 - (S_{yy})^2$

$F^{"}(0) > 0$

Se está en un mínimo relativo y la diferencia de coeficientes de determinación es:

$$F(\rho) = \frac{S_{xx}}{S_{xx} + S_{yy}}$$

Si, por el contrario:

$S_{xx}S_{yy} < (S_{xx})^2 - (S_{yy})^2$

$F^{"}(0) < 0$

Se está en un máximo relativo y la diferencia de coeficientes de determinación es:

$$F(\rho) = \frac{S_{xx}}{S_{xx} + S_{yy}}$$

Segundo caso: Cuando $\boldsymbol{S_{xx} < S_{yy}}$

$$F^{"}(0) = \frac{S_{xx}S_{yy}}{(S_{yy})^2 - (S_{xx})^2} - 1$$

Si:

$$S_{xx}S_{yy} > \left(S_{yy}\right)^2 - (S_{xx})^2$$

$$F^{"}(0) > 0$$

Se está en un mínimo relativo y su valor es:

$$F(0) = \frac{S_{yy}}{S_{xx} + S_{yy}}$$

Si:

$$S_{xx}S_{yy} < \left(S_{yy}\right)^2 - (S_{xx})^2$$

$$F^{"}(0) < 0$$

Se está un máximo relativo y su valor es:

$$F(0) = \frac{S_{yy}}{S_{xx} + S_{yy}}$$

Estudio de puntos críticos en la diferencia del coeficiente de determinación cuando:

$$\rho = \pm \frac{1}{2\left(S_{xx} + S_{yy}\right)} \sqrt{\frac{\left(S_{xx}S_{yy}\right)^2 - \left((S_{xx})^2 - \left(S_{yy}\right)^2\right)^2}{S_{xx}S_{yy}}} \qquad \mathbf{(36)}$$

Dichos valores son reales bajo las siguientes condiciones:

$a)\ \ S_{xx} > S_{yy} \qquad S_{xx}S_{yy} > (S_{xx})^2 - \left(S_{yy}\right)^2$

$b)\ \ S_{xx} < S_{yy} \qquad S_{xx}S_{yy} > \left(S_{yy}\right)^2 - (S_{xx})^2$

En estos casos puede demostrarse que la segunda derivada es menor que cero, en consecuencia, en estos valores de coeficiente de correlación, existen máximos relativos.

Reemplazando (36) en (35) se obtiene:

$$F(\rho) = \frac{1}{4}\left(\frac{(S_{xx})^2 + \left(S_{yy}\right)^2}{S_{xx}S_{yy}} + \frac{S_{xx}S_{yy}}{\left(S_{xx} + S_{yy}\right)^2}\right)$$

Un caso que amerita tratarse de manera especial se presenta cuando:

Caso especial:

$S_{xx} = S_{yy}$

Bajo esta circunstancia:

$$F(\rho) = \frac{1 + \rho - 2\rho^2}{2}$$

Para $\rho \geq 0$ y:

$$F(\rho) = \frac{1 - \rho - 2\rho^2}{2}$$

Para $\rho < 0$

Los criterios de primera y segunda derivada permiten demostrar que en:

$$\rho = \pm\frac{1}{4}$$

Existen máximos relativos cuyo valor es:

$$F\left(\pm\frac{1}{4}\right) = \frac{9}{16}$$

Para todos los casos

Para todos los casos posibles:

$F(\pm 1) = 0$

Resumen

En el siguiente resumen, se sintetiza el resultado del estudio para la diferencia de los coeficientes de determinación entre la recta de regresión de mínimos cuadrados ortogonales y mínimos cuadrados ordinarios, para todas las circunstancias posibles:

Resumen

Condiciones	Mínimo	Máximos	$F(\pm 1)$
$S_{xx} > S_{yy}$ $S_{xx}S_{yy} > (S_{xx})^2 - (S_{yy})^2$	$\rho = 0$ $F(\rho) = \frac{S_{xx}}{S_{xx}+S_{yy}}$ Relativo	$\rho = \pm\frac{1}{2(S_{xx}+S_{yy})}\sqrt{\frac{(S_{xx}S_{yy})^2 - ((S_{xx})^2 - (S_{yy})^2)^2}{S_{xx}S_{yy}}}$ $F(\rho) = \frac{1}{4}\left(\frac{(S_{xx})^2 + (S_{yy})^2}{S_{xx}S_{yy}} + \frac{S_{xx}S_{yy}}{(S_{xx}+S_{yy})^2}\right)$ Relativos	$F(\pm 1) = 0$
$S_{xx} < S_{yy}$ $S_{xx}S_{yy} > (S_{yy})^2 - (S_{xx})^2$	$\rho = 0$ $F(0) = \frac{S_{yy}}{S_{xx}+S_{yy}}$ Relativo	$\rho = \pm\frac{1}{2(S_{xx}+S_{yy})}\sqrt{\frac{(S_{xx}S_{yy})^2 - ((S_{xx})^2 - (S_{yy})^2)^2}{S_{xx}S_{yy}}}$ $F(\rho) = \frac{1}{4}\left(\frac{(S_{xx})^2 + (S_{yy})^2}{S_{xx}S_{yy}} + \frac{S_{xx}S_{yy}}{(S_{xx}+S_{yy})^2}\right)$ Relativos	$F(\pm 1) = 0$
$S_{xx} > S_{yy}$ $S_{xx}S_{yy} < (S_{xx})^2 - (S_{yy})^2$		$\rho = 0$ $F(0) = \frac{S_{xx}}{S_{xx}+S_{yy}}$ Relativo	$F(\pm 1) = 0$
$S_{xx} < S_{yy}$ $S_{xx}S_{yy} < (S_{yy})^2 - (S_{xx})^2$		$\rho = 0$ $F(0) = \frac{S_{yy}}{S_{xx}+S_{yy}}$ Relativo	$F(\pm 1) = 0$
$S_{xx} = S_{yy}$	$\rho = 0$ $F(0) = \frac{1}{2}$	$\rho = \pm\frac{1}{4}$ $F\left(\pm\frac{1}{4}\right) = \frac{9}{16}$ Relativos	$F(\pm 1) = 0$

Nótese que en el intervalo de coeficiente de correlación:

$-1 \leq \rho \leq 1$

Para todos los casos posibles:

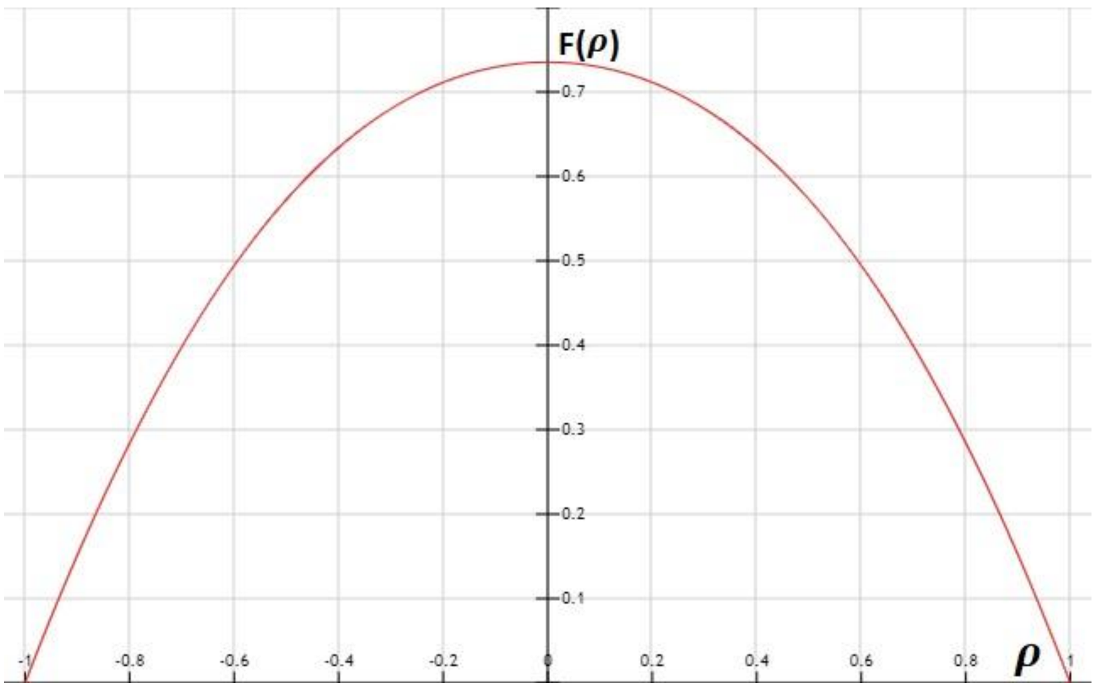
$F(\rho) \geq 0$

Se muestra nuevamente que:

$R_o^2 - R_v^2 \geq 0$

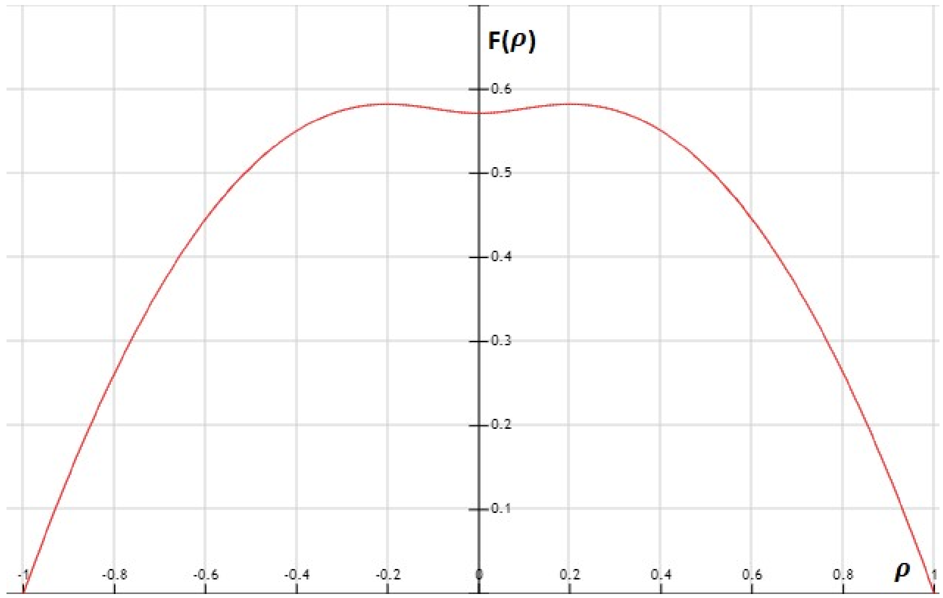
El resultado de este aparte, es compatible con lo demostrado en el capítulo 3, respecto a qué para todas las rectas de regresión de Deming, el mayor coeficiente de determinación se obtiene cuando se emplea la regresión ortogonal.

Grafica 7. Diferencia de coeficiente de determinación F(ρ) para las condiciones: S_{xx}>S_{yy}. $S_{xx}S_{yy}$ < $(S_{xx})^2$-$(S_{yy})^2$



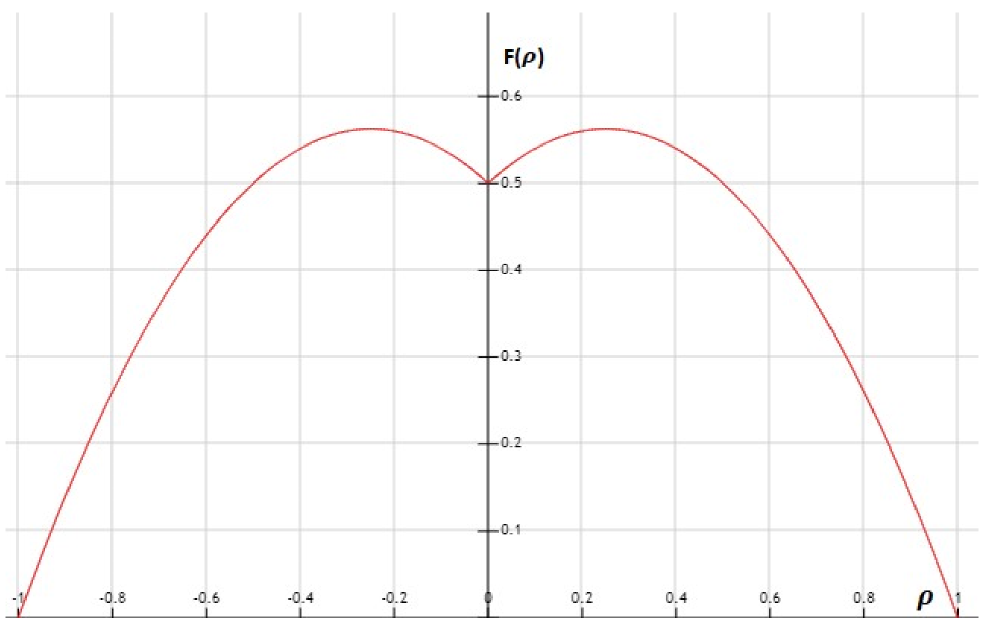
En las abscisas el coeficiente de correlación de Pearson (ρ), en las ordenadas la diferencia en los coeficientes de determinación $F(\rho)$. En este caso se ha empleado:$S_{xx} = 25$, $S_{yy} = 9$
Existe un máximo relativo en: $\rho = 0$, $F(0) = S_{xx}/(S_{xx} + S_{yy}) = 25/34$

Grafica 8. Diferencia de coeficiente de determinación F(ρ) para las condiciones: $S_{xx}>S_{yy}$, $S_{xx}S_{yy} > (S_{xx})^2 - (S_{yy})^2$



En las abscisas el coeficiente de correlación de Pearson (ρ), en las ordenadas la diferencia en los coeficientes de determinación ($F(\rho)$). En este caso se ha empleado: $S_{xx} = 8, S_{yy} = 6$. Existe un mínimo relativo en: $\rho = 0$, $F(0) = S_{xx}/(S_{xx} + S_{yy}) = 4/7$. Máximos relativos en : $\rho = \pm\frac{1}{28}\sqrt{95/3}$, $F(\rho) = 1369/2352$

Grafica 9. Diferencia de coeficiente de determinación F(ρ) para la condición: $S_{xx}=S_{yy}$.



En las abscisas el coeficiente de correlación de Pearson (ρ), en las ordenadas la diferencia en los coeficientes de determinación ($F(\rho)$). Se ha empleado $S_{xx} = S_{yy}$=5. Mínimo en $\rho = 0$, $F(0) = 1/2$. Máximos relativos en $\rho = \pm 1/4$, $F(\pm 1/4) = 9/16$

RAZONES DE PENDIENTES ENTRE RECTAS DE REGRESIÓN DE DEMING Y ORTOGONAL. RECTAS DE REGRESIÓN DE DEMING INCLUIDAS EN EL INTERVALO DE CONFIANZA DE LA RECTA DE REGRESIÓN ORTOGONAL

En este capítulo, se logra amalgamar el estudio geométrico y matemático de la recta de regresión de Deming con inferencia estadística, concretamente con los intervalos de confianza.

En esta oportunidad, también se hace:

$k = S_{yy}/S_{xx}$

Se encuentra la razón entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal, se realiza un estudio completo sobre ese cociente, incluyendo valores máximos y mínimos, encontrándose que cuando $k = \sqrt{\delta}$ el valor de la razón es mínimo cuando $\delta > 1$ y máximo cuando $\delta < 1$. A esas razones (máximas o mínimas) se les llama extremas y sobre estas razones extremas se adelanta un estudio en función del coeficiente de correlación.

Realizado el estudio de razones entre pendientes y como una parte introductoria a los intervalos de confianza, se fija un desfase (intervalo) de la pendiente de la recta de regresión ortogonal y se encuentra las razones de varianzas de errores (δ)que quedan contenidas en ese intervalo, esto significa las rectas de regresión de Deming incluidas en el desfase.

Desarrollada la teoría expuesta en el anterior párrafo, se determina un intervalo de confianza $(1-\alpha)100$ para la pendiente de regresión ortogonal y se encuentra las razones de varianza de errores (δ) incluidas en el intervalo, es decir las rectas de regresión de Deming incluidas en el intervalo de confianza de la recta de regresión ortogonal.

Finalmente, se plantea un chequeo fundamentado en un conocimiento previo que pueda tenerse sobre un intervalo de los posibles valores para la razón de varianzas de errores (δ). El procedimiento desarrollado permite concluir si la recta de regresión de Deming correspondiente, queda incluida en el intervalo de confianza de la recta de regresión ortogonal.

6.1 Razón entre la pendiente de la recta de regresión de Deming y la pendiente de la regresión ortogonal

Se establece la expresión general para la razón entre la recta de regresión de Deming y la pendiente de la regresión ortogonal. En función de k =S_{yy}/S_{xx} y el coeficiente de correlación de Pearson ρ.

Haciendo:

$$k = \frac{S_{yy}}{S_{xx}}$$

La relación entre las dos pendientes se expresa de la siguiente manera:

$$\frac{(k - \delta + \sqrt{(k-\delta)^2 + 4k\delta\rho^2})/2\sqrt{k}\rho}{(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2})/2\sqrt{k}\rho}$$

Simplificando:

$$\frac{k - \delta + \sqrt{(k-\delta)^2 + 4k\delta\rho^2}}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}} \quad (\mathbf{37})$$

6.1.1 Valor de la razón cuando $\boldsymbol{k \to 0}$:

Se estudia el valor de la razón cuando $\boldsymbol{k \to 0}$

Teorema 6.1.1

Siendo:

$$\frac{k - \delta + \sqrt{(k-\delta)^2 + 4k\delta\rho^2}}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}$$

La razón entre la pendiente de la regresión de Deming y la pendiente de la regresión ortogonal, si $k \to 0$, el valor de dicha razón es 1.

Demostración:

$$\lim_{k \to 0} \frac{k - \delta + \sqrt{(k-\delta)^2 + 4k\delta\rho^2}}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}$$

Multiplicando numerador y denominador por la conjugada del numerador y multiplicando numerador y denominador por la conjugada del denominador, se obtiene:

$$\lim_{k\to 0} \frac{-4k\delta\rho^2(k-1-\sqrt{(k-1)^2+4k\rho^2})}{-4k\rho^2(k-\delta-\sqrt{(k-\delta)^2+4k\delta\rho^2}}$$

$$\lim_{k\to 0} \frac{\delta(k-1-\sqrt{(k-1)^2+4k\rho^2})}{(k-\delta-\sqrt{(k-\delta)^2+4k\delta\rho^2}}$$

$$\frac{\delta(0-1-\sqrt{(0-1)^2+4(0)\rho^2})}{(0-\delta-\sqrt{(0-\delta)^2+4(0)\delta\rho^2}}$$

$$\frac{\delta(-2)}{-2\delta}=1$$

6.1.2 Valor de la razón cuando $\boldsymbol{k} \to \infty$

Se estudia el valor de la razón cuando $\boldsymbol{k} \to \infty$:

Teorema 6.1.2

Siendo:

$$\frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

La razón entre la pendiente de la regresión de Deming y la pendiente de la regresión ortogonal, si $k \to \infty$, el valor de dicha razón es 1.

Demostración

$$\lim_{k\to\infty} \frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

$$\lim_{k\to\infty} \frac{1-\delta/k+\sqrt{(1-\delta/k)^2+4\delta\rho^2/k}}{1-1/k+\sqrt{(1-1/k)^2+4\rho^2/k}}$$

Aplicando el límite:

$$\frac{1-0+\sqrt{(1-0)^2+0}}{1-0\ +\sqrt{(1-0)^2+0}}=1$$

6.1.3 Máximos o mínimos relativos en la razón

En el estudio de la razón entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal, se determina el valor de k para el cual existen máximos o mínimos relativos en dicha razón.

Teorema 6.1.3

Siendo la razón de pendientes de la recta de regresión de Deming y la recta de regresión de mínimos cuadrados ortogonales (37):

$$\frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

Existe un mínimo relativo de la razón en $k=\sqrt{\delta}$, cuando $\delta>1$.
Existe un máximo relativo de la razón en $k=\sqrt{\delta}$, cuando $\delta<1$.

Demostración

Se deriva la expresión con respecto a k y se iguala a cero, obteniéndose finalmente:

$$\left(1+\frac{k-\delta+2\delta\rho^2}{\sqrt{(k-\delta)^2+4k\delta\rho^2}}\right)\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)=$$

$$\left(1+\frac{k-1+2\rho^2}{\sqrt{(k-1)^2+4k\rho^2}}\right)\left(k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}\right)$$

El desarrollo de la ecuación para obtener las soluciones es largo y no se incluye, pero puede demostrarse que dichas soluciones son:

$k=0$
$k=\sqrt{\delta}$

Se estudió previamente, la solución $k=0$, en este punto al no hay continuidad para la función.

En $k=\sqrt{\delta}$ hay un punto crítico. No es necesario recurrir a la segunda derivada puesto que:

Cuando $\delta>1$, se obtiene un mínimo relativo en la razón, lo cual se puede inferir fácilmente si se considera que, para estos valores de cociente de varianzas de errores, la pendiente que se obtiene en la regresión de Deming es menor en valor absoluto que la pendiente lograda en la recta de mínimos cuadrados ortogonales

Cuando $\delta < 1$, se obtiene un máximo relativo en la razón, esto puede inferirse si se tiene en cuenta que, para estos valores de cociente de varianzas de errores, la pendiente que se obtiene en la regresión de Deming es mayor en valor absoluto que la pendiente lograda en la recta de mínimos cuadrados ortogonales. (Véase sección 3.5)

Para mayor claridad de lo expuesto en los dos párrafos anteriores, véase las dos gráficas de razones entre pendientes de las rectas de regresión de Deming y de regresión ortogonal que se presentan más adelante en este capítulo. (Véase sección 3.5)

6.1.4 Valores máximos y mínimos de la razón entre las pendientes de las rectas de regresión de Deming y mínimos cuadrados ortogonales.

Con los resultados obtenidos con la demostración del teorema anterior, se calculan los valores máximos y mínimos en la razón de pendientes entre la recta de regresión de Deming y la recta de mínimos cuadrados ortogonales.

Corolario 6.1.4

Aplicando el resultado del teorema 6.1.3, según el cual:

Existe un mínimo relativo de la razón en $k = \sqrt{\delta}$, cuando $\delta > 1$.
Existe un máximo relativo de la razón en $k = \sqrt{\delta}$, cuando $\delta < 1$.

Siendo la razón de pendientes entre la recta de regresión de Deming y la recta de regresión ortogonal:

$$\frac{k - \delta + \sqrt{(k-\delta)^2 + 4k\delta\rho^2}}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}$$

Demostrar que esa razón tiene por valor:

$$\left(\frac{1 - \sqrt{\delta} + \sqrt{\left(\sqrt{\delta} - 1\right)^2 + 4\sqrt{\delta}\rho^2}}{2\rho}\right)^2$$

Demostración:

Se obtiene reemplazando el valor de $k = \sqrt{\delta}$ en la expresión de relación (37):

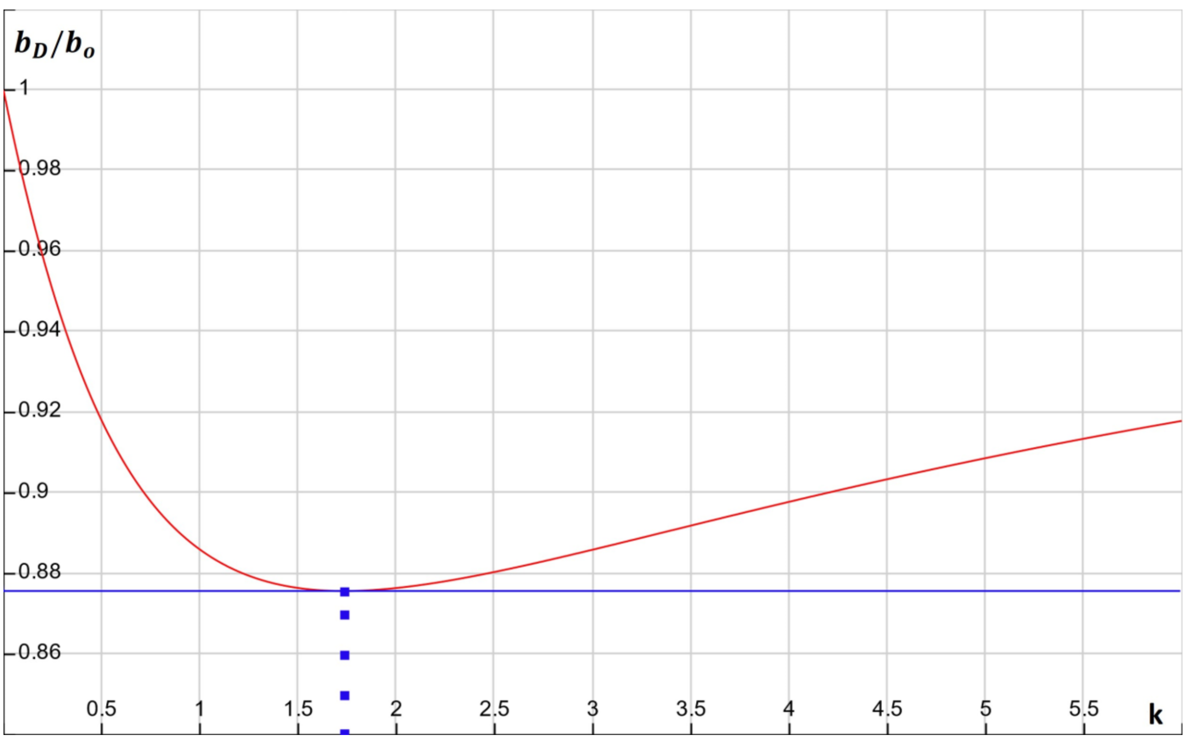
$$\frac{\sqrt{\delta}-\delta+\sqrt{\left(\sqrt{\delta}-\delta\right)^{2}+4\delta\sqrt{\delta}\rho^{2}}}{\sqrt{\delta}-1+\sqrt{\left(\sqrt{\delta}-1\right)^{2}+4\sqrt{\delta}\rho^{2}}}$$

$$\sqrt{\delta}\left(\frac{1-\sqrt{\delta}+\sqrt{\left(\sqrt{\delta}-1\right)^{2}+4\sqrt{\delta}\rho^{2}}}{-\left(1-\sqrt{\delta}\right)+\sqrt{\left(\sqrt{\delta}-1\right)^{2}+4\sqrt{\delta}\rho^{2}}}\right)$$

Racionalizando denominador:

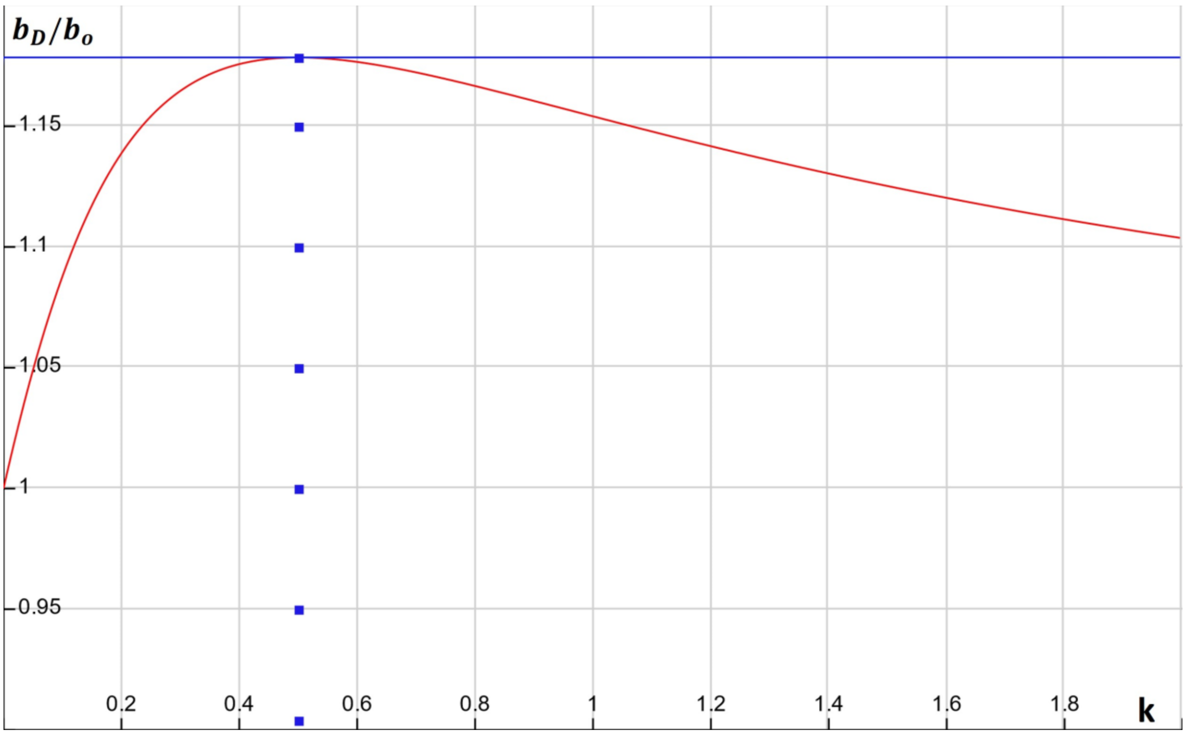
$$\left(\frac{1-\sqrt{\delta}+\sqrt{\left(\sqrt{\delta}-1\right)^{2}+4\sqrt{\delta}\rho^{2}}}{2\rho}\right)^{2} \qquad (\mathbf{38})$$

Grafica 10. Razón entre la pendiente de la recta de regresión de Deming y la pendiente de la regresión ortogonal cuando δ>1



En las abscisas los valores de k y en ordenadas la relación entre las pendientes de la recta de Deming y la pendiente de la recta de regresión ortogonal. En el ejemplo se ha empleado $\delta = 3$ y $\rho = 0.8$, nótese que el mínimo relativo se encuentra en $k = \sqrt{3}$ y su valor es 0.87573019 justamente el que corresponde a la expresión planteada.

Grafica 11. Razón entre la pendiente de la recta de regresión de Deming y la pendiente de la regresión ortogonal cuando δ<1



En las abscisas los valores de k y en ordenadas la relación entre las pendientes de la recta de Deming y la pendiente de la recta de regresión ortogonal. En el ejemplo se ha empleado $\delta = 0.25$ y $\rho = 0.8$, nótese que el mínimo relativo se encuentra en $k = 0.5$ y su valor es 1,17848894 justamente el que corresponde a la expresión planteada.

6.1.5 Razones extremas (máximas o mínimas) entre pendientes

En el anterior aparte, se encontró la expresión para calcular las razones extremas (máximas o mínimas) entre la pendiente de la recta de Deming y la pendiente de la recta de regresión ortogonal en función del coeficiente de correlación de Pearson se demostró que esto sucede cuando $k = \sqrt{\delta}$ .Dicha expresión contenida en (38) es:

$$\left(\frac{1-\sqrt{\delta}+\sqrt{\left(\sqrt{\delta}-1\right)^{2}+4\sqrt{\delta}\rho^{2}}}{2\rho}\right)^{2}$$

Inicialmente se estudiara el valor de esa razón extrema cuando $\delta = 0$

El reemplazo respectivo entrega el valor:

$$\frac{1}{\rho^2}$$

Se estudiara que sucede con esas razones extremas cuando $\delta \to \infty$

$$\lim_{\delta\to\infty}\left(\frac{1-\sqrt{\delta}+\sqrt{\left(\sqrt{\delta}-1\right)^2+4\sqrt{\delta}\rho^2}}{2\rho}\right)^2$$

Racionalizando numerador:

$$\lim_{\delta\to\infty}\left(\frac{2\sqrt{\delta}\rho^2}{\rho\left(\sqrt{\left(\sqrt{\delta}-1\right)^2+4\sqrt{\delta}\rho^2}-\left(1-\sqrt{\delta}\right)\right)}\right)^2$$

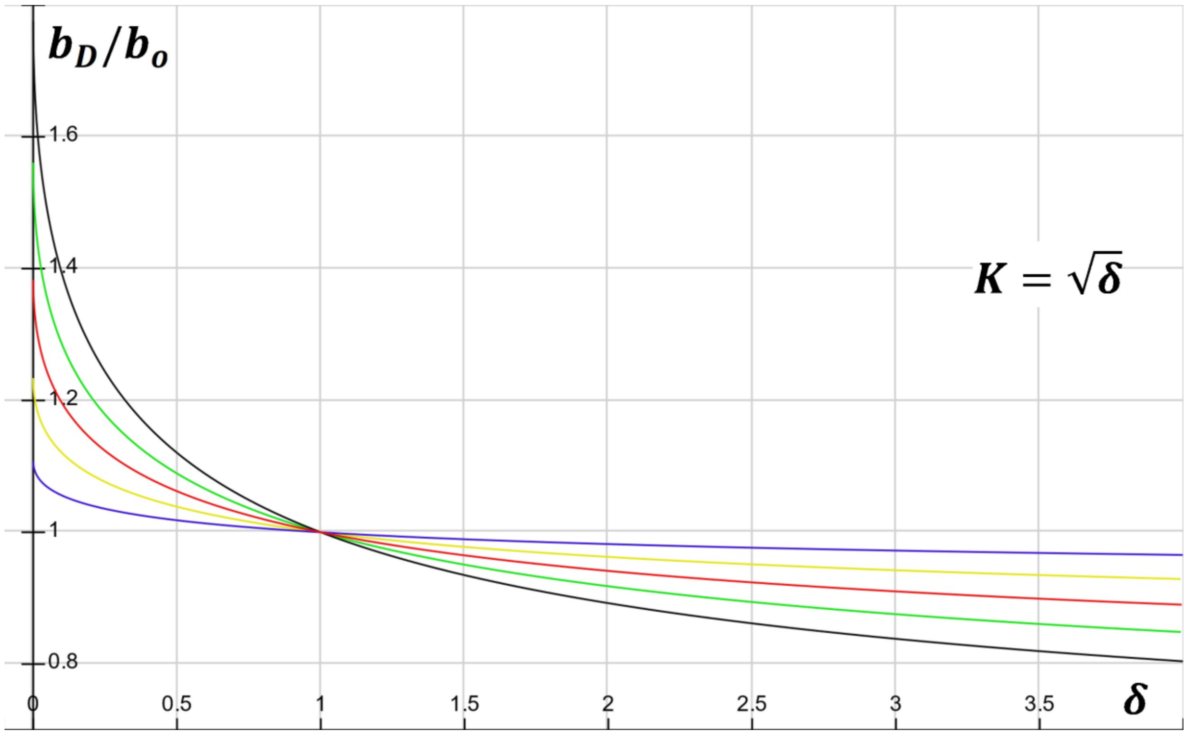
Dividiendo numerador y denominador por $\sqrt{\delta}$:

$$\lim_{\delta\to\infty}\left(\frac{2\rho^2}{\rho\left(\sqrt{\left(1-1/\sqrt{\delta}\right)^2+4\rho^2/\sqrt{\delta}}-\left(1/\sqrt{\delta}-1\right)\right)}\right)^2$$

$$\left(\frac{2\rho}{\sqrt{(1-0)^2+0}-(0-1)}\right)^2=\rho^2$$

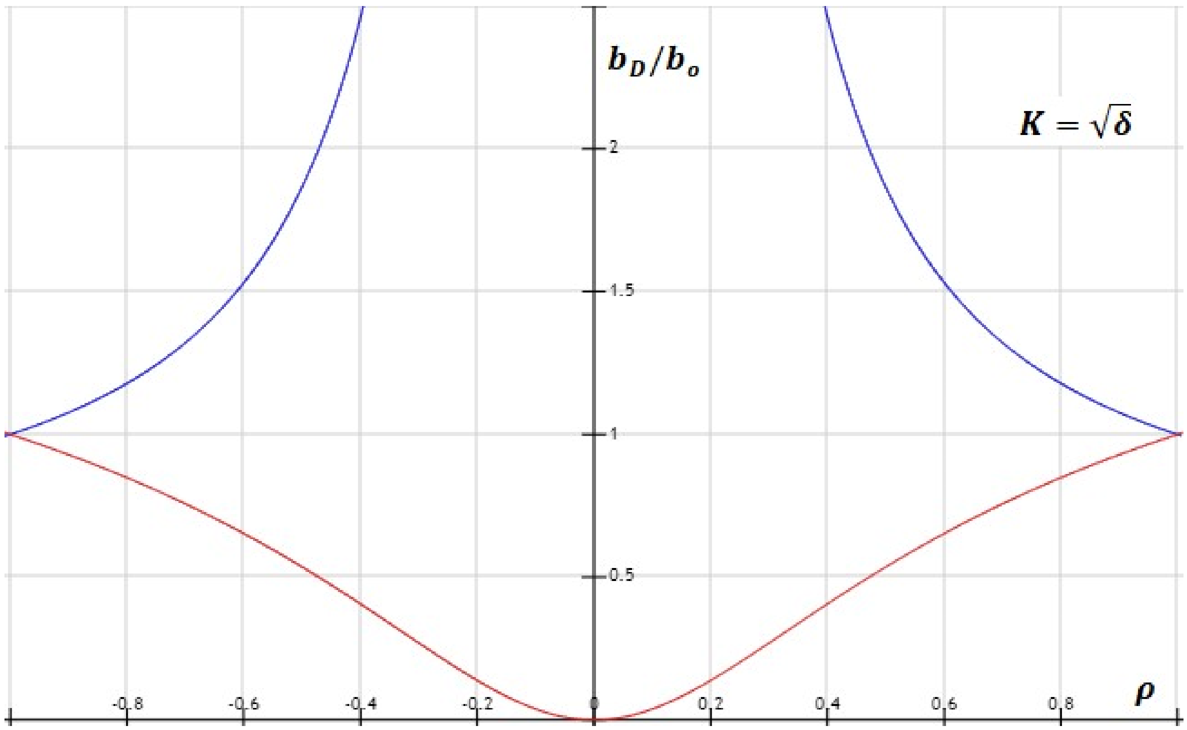
También es inmediato que cuando $\delta = 1$, la razón es 1 sin importar el valor del coeficiente de correlación de Pearson.

Grafica 12. Razones extremas entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal en función de la razón (δ) entre las varianzas de los errores de la variable dependiente y de la variable dependiente, para distintos valores del coeficiente de correlación de Pearson.



Las abscisas corresponden al valor δ y las ordenadas a las razones extremas entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal. Línea negra: ρ = 0,75, Línea verde: ρ = 0.80, Línea roja: ρ = 0.85, Línea amarilla: ρ = 0.90, Línea azul: ρ = 0.95. Se aprecia que a medida que aumenta en valor absoluto del valor del coeficiente de correlación de Pearson, la razón se aproxima más a 1. Nótese además que cuando $\delta < 1$, la razón es mayor que 1 y cuando $\delta > 1$ la razón es menor que 1.

Grafica 13. Razones extremas entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión de mínimos cuadrados ortogonales, en función del coeficiente de correlación de Pearson.



En las abscisas se registra el coeficiente de correlación de Pearson, en las ordenadas la razón de pendientes. Se ha empleado $k = \sqrt{\delta}$, con el propósito de encontrar razones extremas entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal. Con línea azul aparece el valor de razón para todos los valores posibles de coeficiente de correlación de Pearson, cuando $\delta < 1$ (En este caso se empleó $\delta = 0.25$) obsérvese que en estos casos la razón es mayor o igual a 1. Con línea roja aparece el valor de razón para todos los valores posibles de coeficiente de correlación de Pearson, cuando $\delta > 1$ (En este caso se empleó $\delta = 4$) obsérvese que en estos casos la razón es menor o igual a 1. Para todos los casos, en la medida que el valor absoluto del coeficiente de correlación de Pearson se acerca a 1, la razón también se aproxima a ese valor.

6.2 Rectas de regresión de Deming incluidas en un desfase definido de la pendiente de la recta de regresión ortogonal

Se considera un desfase en ambos sentidos en la pendiente de la recta de regresión ortogonal, se encuentra las rectas de regresión de Deming que quedan contenidas en dicho desfase.

Teorema 6.2

Sea p, un valor para la razón entre la pendiente de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal.

$$\frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}=p \qquad (\mathbf{39})$$

El valor δ que satisface esta condición es:

$$\delta=F(p,k,\rho)=\frac{p\left((k-pk+p)\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)-2kp\rho^2\right)}{p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)-2k\rho^2}$$

Demostración:

A continuación, se despeja el valor de δ que satisface la igualdad:

$$k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}=p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)$$

$$\sqrt{(k-\delta)^2+4k\delta\rho^2}=p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)-(k-\delta)$$

Elevando ambos miembros al cuadrado:

$$(k-\delta)^2+4k\delta\rho^2=$$
$$p^2\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)^2-2(k-\delta)p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)+(k-\delta)^2$$

Trabajando algebraicamente se llega a:

$$\delta=\frac{p\left(2k\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)-p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)^2\right)}{2\left(p\left(k-1+\sqrt{(k-1)^2+4k\rho^2}\right)-2k\rho^2\right)}$$

Simplificando se obtiene:

$$\delta = \frac{p\left((k - pk + p)\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2kp\rho^2\right)}{p\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2k\rho^2}$$

Considerando que δ es función de p, k, ρ, se puede expresar que:

$$\delta = F(p, k, \rho)$$

Por lo tanto:

$$\delta = F(p, k, \rho) = \frac{p\left((k - pk + p)\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2kp\rho^2\right)}{p\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2k\rho^2} \qquad (\mathbf{40})$$

Desfase de la pendiente de la recta de regresión ortogonal

Escójase dos valores reales positivos: p_1, p_2 ,tales que:

$$\frac{p_1 + p_2}{2} = 1$$

Y, además:

$$p_1 < p_2$$

El porcentaje de desfase es:

$$100(p_2 - 1) = 100(1 - p_1)$$

Sea b_o la pendiente de la regresión ortogonal. los valores de δ que garantizan que se obtendrá un pendiente de recta de regresión en el intervalo:

$$[p_1 b_o \,;\, p_2 b_o]$$

Pertenecen al intervalo de δ:

$$[F(p_2, k, \rho) \,;\, F(p_1, k, \rho)]$$

Lo anterior solo es válido para:

$$F(p_2, k, \rho) \geq 0$$

Debe considerarse que el cálculo de δ puede entregar valores negativos, esto sucede cuando:

$$F(p_2, k, \rho) < 0$$

Situación que no es posible por la definición de la razón de varianzas de errores.

Para solucionar este inconveniente, debe encontrarse el valor de k, para el cual $\delta = 0$ y además que $p = p_2$

Teorema 6.2.1

El valor de k para el cual $\delta = 0$ y además $p = p_2$ es:

$$k = \frac{p_2(1 - p_2\rho^2)}{p_2 - 1}$$

Demostración:

La razón de pendientes (39), tomará la forma:

$$\frac{2k}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}} = p_2$$

$$2k = p_2(k-1) + p_2\sqrt{(k-1)^2 + 4k\rho^2}$$

$$p_2^2((k-1)^2 + 4k\rho^2) = 4k^2 - 4kp_2(k-1) + p_2^2(k-1)^2$$

$$4k\rho^2 p_2^2 = 4k^2 - 4kp_2(k-1)$$

$$\rho^2 p_2^2 = k - p_2(k-1)$$

$$\rho^2 p_2^2 = k - p_2 k + p_2$$

$$\rho^2 p_2^2 - p_2 = k(1 - p_2)$$

$$k = \frac{p_2(1 - p_2\rho^2)}{p_2 - 1}$$

Consecuencias del teorema 6.2.1

A partir de ese valor de k:

$$k = \frac{p_2(1 - p_2\rho^2)}{p_2 - 1}$$

El valor inferior del intervalo de δ es cero.

El límite superior del intervalo de pendiente, se altera y considerando (39) su valor está definido por:

$$\frac{2kb_o}{k-1+\sqrt{(k-1)^2+4k\rho^2}} \qquad \mathbf{(41)}$$

Debe tenerse en cuenta que:

$$b_o = \frac{k-1+\sqrt{(k-1)^2+4k\rho^2}}{2\sqrt{k}\rho}$$

Por lo tanto:

$$k-1+\sqrt{(k-1)^2+4k\rho^2} = 2\sqrt{k}\rho b_o \qquad \mathbf{(42)}$$

Reemplazando este valor en el denominador de (41):

$$\frac{2kb_o}{2\sqrt{k}\rho b_o} = \frac{\sqrt{k}}{\rho}$$

Límite superior del intervalo:

$$\frac{\sqrt{k}}{\rho}$$

Resumen:

Cuando:

$$k \leq \frac{p_2(1-p_2\rho^2)}{p_2-1}$$

El intervalo de la razón de varianzas de errores (δ) es:

$$[F(p_2,k,\rho); \quad F(p_1,k,\rho)]$$

Que corresponden al intervalo de pendientes:

$$[p_1 b_o \,;\, p_2 b_o]$$

Pero cuando:

$$k > \frac{p_2(1-p_2\rho^2)}{p_2-1}$$

El intervalo de la razón de varianzas de errores (δ) es:

$$[0\,; \quad F(p_1,k,\rho)]$$

Que corresponde al intervalo de pendientes:

$\left[p_1 b_o\,;\ \sqrt{k} \div \rho\right]$

Donde :

$\sqrt{k} \div \rho < p_2 b_o$

Si se desea encontrar el valor p que está amplificando a la pendiente de regresión ortogonal en la parte derecha del último intervalo, basta con hacer el siguiente cálculo:

$$p = \frac{\sqrt{k}}{\rho b_o}$$

Finalmente, es también importante averiguar el valor de k, para el cual se logra dentro del intervalo $[p_1;\ p_2]$ el valor de $\delta\ \rightarrow \infty$.

Teorema 6.2.2

Sea $[p_1;\ p_2]$ el intervalo de razones entre la pendiente de la regresión de Deming y la pendiente de la regresión ortogonal, el valor de k para el cual $\delta \rightarrow \infty$ es:

$$k = \frac{p_1(1 - p_1)}{(p_1 - \rho^2)}$$

Demostración:

$\delta \rightarrow \infty$, se presenta cuando el denominador de la siguiente expresión es cero:

$$F(p_1, k, \rho) = \frac{p_1\left((k - p_1 k + p_1)\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2kp_1\rho^2\right)}{p_1\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2k\rho^2}$$

$$p_1\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2k\rho^2 = 0$$

$$p_1\sqrt{(k-1)^2 + 4k\rho^2} = \ 2k\rho^2 - p_1(k-1)$$

$$p_1^2(k-1)^2 + 4kp_1^2\rho^2 = 4k^2\rho^4 - 4p_1 k\rho^2(k-1) + p_1^2(k-1)^2$$

$$p_1^2 = k\rho^2 - p_1(k-1)$$

$$p_1^2 = k\rho^2 - p_1 k + p_1$$

$$k(p_1 - \rho^2) = p_1(1 - p_1)$$

$$k = \frac{p_1(1-p_1)}{(p_1-\rho^2)}$$

Resumen:

Para valores de:

$$k < \frac{p_1(1-p_1)}{(p_1-\rho^2)}$$

El límite superior del intervalo de δ es ∞.

Y el intervalo de δ tendrá la forma:

$$[F(p_2, k, \rho); \quad \infty]$$

En estos casos el intervalo de razones de pendientes también se cambia, de acuerdo a lo siguiente:

La razón entre la pendiente de la regresión de Deming y la pendiente de la regresión ortogonal, cuando $\delta \to \infty$, considerando (37), se calcula así:

$$\lim_{\delta \to \infty} \frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}} = p$$

Multiplicando por la conjugada del numerador:

$$\lim_{\delta \to \infty} \frac{-4k\delta\rho^2}{(k-1+\sqrt{(k-1)^2+4k\rho^2})\left(k-\delta-\sqrt{(k-\delta)^2+4k\delta\rho^2}\right)}$$

Dividiendo numerador y denominador por δ :

$$\lim_{\delta \to \infty} \frac{-4k\rho^2}{(k-1+\sqrt{(k-1)^2+4k\rho^2})\left(k/\delta-1-\sqrt{(k/\delta-1)^2+4k\rho^2/\delta}\right)}$$

Aplicando el límite:

$$\frac{2k\rho^2}{k-1+\sqrt{(k-1)^2+4k\rho^2}} \qquad \mathbf{(43)}$$

Pero de acuerdo a (41):

$$k-1+\sqrt{(k-1)^2+4k\rho^2} = 2\sqrt{k}\rho b_o$$

Reemplazando en (43) se obtiene la razón inferior de pendientes:

$$\frac{k\rho^2}{\sqrt{k}\rho b_o}$$

$$\frac{\sqrt{k}\rho}{b_o}$$

Pero la pendiente inferior es:

$$\frac{\sqrt{k}\rho}{b_o}(b_o)$$

Y el valor inferior del intervalo de pendientes es:

$$\sqrt{k}\rho$$

El intervalo de pendiente se expresará de la siguiente manera:

$$[\sqrt{k}\rho;\ p_2 b_o]$$

Donde:

$$\sqrt{k}\rho > p_1 b_0$$

Si se desea encontrar el valor p que está modificando la pendiente de regresión ortogonal en la parte izquierda del último intervalo, basta con hacer el siguiente cálculo:

$$p = \frac{\sqrt{k}\rho}{b_0}$$

6.2.1 Ejemplo de aplicación

En el ejemplo que sigue y acompañado de la respectiva gráfica se explica el desarrollo teórico planteado:

Se han escogido los siguientes valores:

$p_1 = 0.95$
$p_2 = 1.05$

Se encuentra el intervalo de valores de δ para los cuales el rango de las pendientes de la recta de regresión de Deming resultantes estén 5% por encima o por debajo de la pendiente de la recta de regresión ortogonal.

Se ha escogido coeficiente de correlación de Pearson de: 0.90, aunque también están graficadas las situaciones para coeficientes de correlación de Pearson de 0.85, 0.8 y 0.7

Con línea verde se ha graficado:

$[F(1.05, k, 0.9); \quad F(0.95, k, 0.9)]$

Elíjase un valor $k = 2$

$$\frac{p\left((k - pk + p)\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2kp\rho^2\right)}{p\left(k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}\right) - 2k\rho^2}$$

$$F(1.05, 2, 0.9) = \frac{1.05\left((2 - 1.05(2) + 1.05)\left(2 - 1 + \sqrt{(2-1)^2 + 4(2)0.9^2}\right) - 2(2)(1.05)0.9^2\right)}{1.05\left(2 - 1 + \sqrt{(2-1)^2 + 4(2)0.9^2}\right) - 2(2)0.9^2}$$

$F(1.05, 2, 0.9) =$0.22520161

$$F(0.95, 2, 0.9) = \frac{0.95\left((2 - 0.95(2) + 0.95)\left(2 - 1 + \sqrt{(2-1)^2 + 4(2)0.9^2}\right) - 2(2)(0.95)0.9^2\right)}{0.95\left(2 - 1 + \sqrt{(2-1)^2 + 4(2)0.9^2}\right) - 2(2)0.9^2}$$

$F(0.95, 2, 0.9) =$2.60056153

Lo anterior significa que se podría utilizar cualquiera de los valores del intervalo de δ:

$[0.22520161 \,; \; 2.60056153]$

Sin que la pendiente de la recta de la regresión de Deming se aleje de $\pm 5\%$ de la pendiente de la recta de la regresión ortogonal.

La pendiente de la recta de regresión ortogonal se calcula así:

$$b_o = \frac{2 - 1 + \sqrt{(2-1)^2 + 4(2)0.9^2}}{2\sqrt{2}(0.9)}$$

$b_o =$1.46723041

Y el intervalo de pendientes resultantes será:

$[0.95(1.46723041) \,; 1.05(1.46723041)]$

$[1.39386889\ ;\ 1.54059193]$

Es importante probar que sucede cuando $\delta = 0$

Esto según lo demostrado, ocurre, cuando:

$$k = \frac{1.05(1 - 1.05(0.9^2)}{1.05 - 1}$$

k =3.1395

Tal como se aprecia en la gráfica.

También es interesante analizar lo que sucede cuando:

k >3.1395

Sea:

$k = 4$

Se sabe que el límite inferior para el intervalo de δ es cero.

Entonces solo se encuentra:

$$F(0.95, 4, 0.9) = \frac{0.95\left((4 - 0.95(4) + 0.95)\left(4 - 1 + \sqrt{(4-1)^2 + 4(4)0.9^2}\right) - 2(4)(0.95)0.9^2\right)}{0.95\left(4 - 1 + \sqrt{(4-1)^2 + 4(4)0.9^2}\right) - 2(4)0.9^2}$$

$F(0.95, 4, 0.9)$ =3,10146924

Lo anterior significa que se podría utilizar cualquiera de los valores del intervalo de δ:

$[0;\ 3{,}10146924]$

Sin que la pendiente de la recta de la regresión de Deming se aleje de $\pm$5% de la pendiente de la recta de la regresión ortogonal.

El valor de la pendiente de la recta de regresión ortogonal será:

$$b_o = \frac{4 - 1 + \sqrt{(4-1)^2 + 4(4)0.9^2}}{2\sqrt{4}(0.9)}$$

b_o =2.13504161

El intervalo de pendientes correspondientes será:

$$\left[0.95(2.13504161)\ ;\ \sqrt{4} \div 0.9\right]$$

$[2.02828953\ ;\ 2.22222222]$

Para estos valores de k es muy adecuado calcular el valor p para el límite superior:

$$p = \frac{\sqrt{4}}{0.9(2.13504161)}$$

p =1.04083321

Es decir que ese amplio rango de valores de δ ya indicado, producen un intervalo de variación de la pendiente de regresión ortogonal 5% por debajo y solo 4% por encima de la misma.

Analícese que sucede cuando:

$$k < \frac{p_1(1-p_1)}{(p_1-\rho^2)}$$

$$k < \frac{0.95(1-0.95)}{(0.95-0.9^2)}$$

k <0.33928571

Sea:

$k = 0.25$

Se sabe que el límite superior para el intervalo de δ es infinito.

Entonces solo se encuentra:

$F(1.05, 0.25, 0.9) =$

$$\frac{1.05\left((0.25-1.05(0.25)+1.05)\left(0.25-1+\sqrt{(0.25-1)^2+4(0.25)0.9^2}\right)-2(0.25)(1.05)0.9^2\right)}{1.05\left(0.25-1+\sqrt{(0.25-1)^2+4(0.25)0.9^2}\right)-2(0.25)0.9^2}$$

$F(1.05, 0.25, 0.9)$ =0,33763364

El intervalo de δ es:

[0,33763364; ∞]

Para el intervalo de pendientes se requiere calcular b_o :

$$b_o = \frac{0.25-1+\sqrt{(0.25-1)^2+4(0.25)0.9^2}}{2\sqrt{0.25}(0,9)}$$

b_o = 0.46837495

El intervalo de pendiente:

$[\sqrt{0.25}(0.9)\ ; 1.05(0{,}46837495)]$

$[0.45\ ; 0.49179369\ ]$

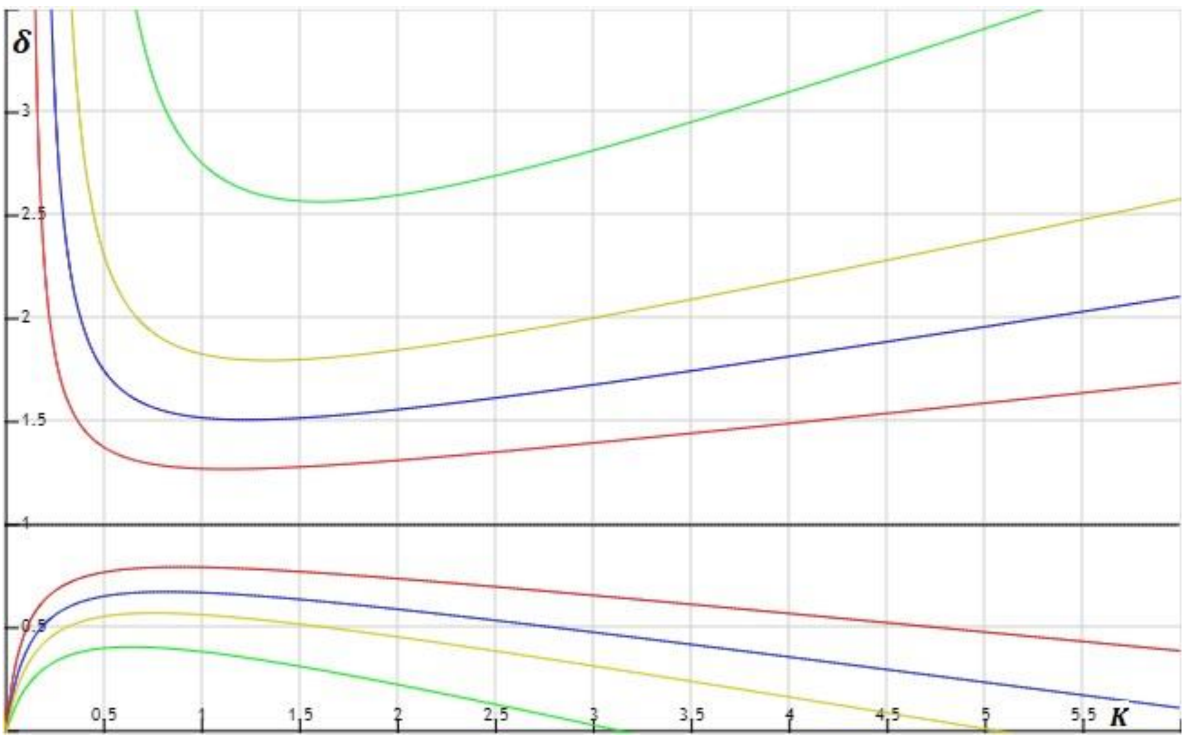
Para estos valores de k es muy adecuado calcular el valor p para el límite inferior:

$$p = \frac{\sqrt{0.25}(0.9)}{0.46837495}$$

$$p = 0.96076873$$

Es decir que ese amplio rango de valores de δ ya indicado, producen un intervalo de variación de la pendiente de regresión ortogonal menor del 4 % por debajo y 5% por encima de la misma.

Grafica 14. Intervalos de δ que generan pendientes de la recta de regresión de Deming, comprendidas entre 0,95 y 1.05 veces la recta de regresión ortogonal en función de k y ρ.



En las abscisas se lee el valor de k y en las ordenadas el valor de δ, se han empleado los siguientes colores para el coeficiente de correlación de Pearson: verde = 0.90, amarillo = 0.85, azul =0.8, rojo=0.7. A partir del valor k, en el cual la curva intercepta el eje, el límite superior del intervalo de pendientes estipuladas disminuye, esto constituye una ventaja, puesto que esas pendientes se alejan menos, superiormente de la pendiente de la recta de regresión ortogonal.

6.3 Rectas de la regresión de Deming incluidas en un intervalo de confianza de la pendiente de la regresión ortogonal.

En esta sección se tendrá en cuenta la teoría desarrollada en la sección 6.2. En este nuevo aparte, el desfase corresponde a un intervalo de confianza de la recta de regresión ortogonal y se define las rectas de regresión de Deming que quedan contenidas al interior de dicho intervalo de confianza.

El intervalo de confianza al (1-$\alpha) * 100$ para la pendiente b_o de la recta de regresión ortogonal es:

$$b_o \pm Z_{(1-\alpha/2)}\sqrt{V(b_0)}$$

Donde la varianza de la pendiente de la recta de regresión ortogonal es:

$$V(b_0)$$

De acuerdo a lo expuesto en el aparte anterior, si se hace:

$$p_1 = 1 - \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} \qquad (\mathbf{44})$$

$$p_2 = 1 + \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} \qquad (\mathbf{45})$$

Se puede determinar las rectas de regresión de Deming que quedan comprendidas en el intervalo:

$$[p_1 b_0\,; p_2 b_0]$$

$$\left[\left(1 - \frac{Z_{1-\propto/2}\sqrt{V(b_0)}}{b_0}\right) b_0\,;\; \left(1 + \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right) b_0\right]$$

El intervalo de valores de la razón de varianzas de errores (δ) incluidos en el intervalo de confianza de la pendiente estará definido de acuerdo a (40) por:

$$[F(p_2, k, \rho)\,;\; F(p_1, k, \rho)]$$

Que en este caso también puede expresarse como:

$$\left[F\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0},k,\rho\right)\ ;\ \ F\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0},k,\rho\right)\right]$$

Caso especial:

Cuando

$$k>\frac{p_2(1-p_2\rho^2)}{p_2-1}$$

Que en este caso es:

$$k>\frac{\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)\left(1-\rho^2\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)\right)}{\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}}$$

El intervalo de la razón de varianzas de errores (δ) es:

$$[0\ ;\quad F(p_1,k,\rho)]$$

Que en este caso es:

$$\left[0\ ;\quad F\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0},k,\rho\right)\ \right]$$

Que corresponde al intervalo de pendientes:

$$\left[p_1 b_o\ ;\ \ \sqrt{k}\div\rho\right]$$

Es decir:

$$\left[\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)b_o\ ;\ \ \sqrt{k}\div\rho\right]$$

Donde :

$$\sqrt{k}\div\rho<p_2 b_o$$

Es decir:

$$\sqrt{k}\div\rho<\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)b_o$$

Si se desea encontrar el valor p que está amplificando a la pendiente de regresión ortogonal en la parte derecha del último intervalo, basta con hacer el siguiente cálculo:

$$p = \frac{\sqrt{k}}{\rho b_o}$$

Caso especial:

Cuando

$$k < \frac{p_1(1-p_1)}{(p_1-\rho^2)}$$

Que es este caso es:

$$k < \frac{\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)\left(\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)}{\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}-\rho^2\right)}$$

El límite superior del intervalo de δ es ∞.

Y el intervalo de δ tendrá la forma:

$$[F(p_2,k,\rho);\quad \infty]$$

$$\left[F\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0},k,\rho\right);\quad \infty\right]$$

En estos casos el intervalo de pendientes también se cambia, de acuerdo a lo siguiente:

La razón entre la pendiente de la regresión de Deming y la pendiente de la regresión ortogonal, cuando $\delta \rightarrow \infty$, luego de realizar trabajo algebraico, toma la forma:

$$\frac{k\rho^2}{\sqrt{k}\rho b_o}$$

$$\frac{\sqrt{k}\rho}{b_o}$$

Y el valor inferior del intervalo de pendiente es:

$\sqrt{k}\rho$

El intervalo de pendiente se expresará de la siguiente manera:

$[\sqrt{k}\rho;\ p_2 b_o]$

O lo que es lo mismo:

$$\left[\sqrt{k}\rho;\ \left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)b_o\right]$$

Donde:

$\sqrt{k}\rho > p_1 b_0$

$$\sqrt{k}\rho > \left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)b_0$$

Si se desea encontrar el valor p que está modificando la pendiente de regresión ortogonal en la parte izquierda del último intervalo, basta con hacer el siguiente cálculo:

$$p=\frac{\sqrt{k}\rho}{b_0}$$

6.3.1 Varianza de la pendiente de la recta de regresión ortogonal.

La varianza de la pendiente de la recta de la regresión ortogonal se adecúa de acuerdo a las expresiones que entrega Minitab Statistical Software.

Minitab emplea una singular nomenclatura:

$$V(b_o)=\frac{\hat{\sigma}_{xx}S_{vv}+\sigma_u^2 S_{vv}-b_o^2\sigma_u^4}{(n-1)\hat{\sigma}_{xx}^2}$$

Donde:

$$\hat{\sigma}_{xx}=\frac{S_{xx}\left(\sqrt{(k-1)^2+4k\rho^2}-(k-1)\right)}{2(n-1)}$$

$$\sigma_u^2 = \frac{S_{xx}\left(k+1-\sqrt{(k-1)^2+4k\rho^2}\right)}{2(n-1)}$$

$$S_{vv} = \frac{(n-1)(1+b_o^2)\sigma_u^2}{n-2}$$

6.3.2 Ejercicio de Aplicación

Sean doscientas observaciones (x, y) con las siguientes características:

$S_{xx} = 16$

$S_{yy} = 25$

$\rho = 0.85$

$S_{xy} = 17$

$n = 200$

Al realizar los cálculos con las expresiones respectivas se obtiene:

k =1.5625

b_o = 1.29914738

$V(b_o)$ = 0.00327611

Intervalo de confianza del 95% para la pendiente de la recta de regresión ortogonal:

$[1.18696223;\ 1.41133252]$

$$1 - \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} = 0.9136471$$

$$1 + \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} = 1.0863529$$

El intervalo de valores de la razón de varianzas de errores (δ) incluidos en el intervalo de confianza de la pendiente estará definido por:

$$\left[F\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}, k, \rho\right);\ F\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}, k, \rho\right)\right] =$$

$[F(1.0863529,\ \ 1.5625,\ \ 0.85);\ F(0.9136471,\ \ 1.5625,\ \ 0.85)\ ]$

$F(1.0863529,\ \ 1.5625,\ \ 0.85) =$0.254725

$F(0.9136471,\ \ 1.5625,\ \ 0.85) =$2.87391749

$[0.254725;\ 2.87391749]$

Lo anterior significa que si se emplea regresión de Deming con un valor de razón de varianzas de errores δ, perteneciente al intervalo establecido, la pendiente estimada, cae en el intervalo de confianza del 95% de la regresión ortogonal.

Se debe establecer que el valor de k, cumpla las siguientes dos condiciones:

Primera condición:

$$k > \frac{\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)\left(1-\rho^2\left(1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)\right)}{\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}}$$

$$k \leq \frac{1.0863529(1-0.85^2(1.0863529))}{0.0863529}$$

$k \leq$ 2.4910573

1.5625$\leq$ 2.4910573

Condición cumplida.

Segunda condición:

$$k \geq \frac{\left(1-\frac{Z_{1-\frac{\alpha}{2}}\sqrt{V(b_0)}}{b_0}\right)\left(\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}\right)}{\left(1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}-\rho^2\right)}$$

$$k \geq \frac{(0.9136471)(0{,}0863529)}{(0.9136471-0.85^2)}$$

$k \geq$0.41275059

1.5625 $\geq$0.41275059

Condición cumplida.

En este caso se cumplen las dos condiciones y no hace falta realizar los ajustes propuestos.

Finalmente, y a manera de comprobación de la teoría expuesta en este aparte, debe cumplirse que al calcular las pendientes de las rectas de regresión de Deming con los valores extremos del intervalo de δ deben coincidir con las pendientes extremas del intervalo de confianza de la pendiente de regresión ortogonal.

Se emplea la expresión original de la regresión de Deming:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}{2S_{xy}}$$

En efecto para:

$\delta = 2.87391749$

$$b_D = \frac{25 - 16(2.87391749) + \sqrt{\left(25 - 16(2.87391749)\right)^2 + 4(2.87391749)(17^2)}}{2(17)} = 1.18696223$$

Para:

$\delta = 0.254725$

$$b_D = \frac{25 - 16(0.254725) + \sqrt{\left(25 - 16(0.254725)\right)^2 + 4(0.254725)(17^2)}}{2(17)} = 1.41133252$$

6.4 Chequeo de inclusión de pendiente de la regresión de Deming en el intervalo de confianza de la pendiente de la regresión ortogonal, asumiendo valor extremo en el cociente de varianzas de errores

Es posible que el investigador encuentre dificultades para definir la razón de varianzas de errores δ, pero que fundamentado en su conocimiento piense que su valor se encuentra entre $1 < \delta \leq \delta_2$. Ante tal incertidumbre, puede optar por realizar regresión ortogonal siempre y cuando la recta pendiente de la regresión de Deming quede incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal. También podría ocurrir que el investigador desconozca δ pero que de acuerdo a su conocimiento piense que $\delta_1 \leq \delta < 1$ y desea saber si la pendiente de recta de regresión de Deming queda incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal.

Caso 1

$1 < \delta \leq \delta_2$.

Teorema 6.4

Bajo la condición:

$1 < \delta \leq \delta_2$

Considerando la razón de pendientes entre la recta de regresión de Deming y la recta de regresión ortogonal.

Sean:

$$p_1^* = \frac{k - \delta_2 + \sqrt{(k-\delta_2)^2 + 4k\delta_2\rho^2}}{k - 1 + \sqrt{(k-1)^2 + 4k\rho^2}}$$

$$p_1 = 1 - \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}$$

Si:

$p_1^* \geq p_1$

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal.

Demostración:

De acuerdo a (37) la razón entre pendientes de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal es:

$$\frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

Si a esa relación aplicada a la razón extrema de cocientes de variables se le llama p_1^*:

$$p_1^*=\frac{k-\delta_2+\sqrt{(k-\delta_2)^2+4k\delta_2\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

Por (44) se sabe que:

$$p_1 = 1-\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}$$

Donde p_1 es el factor que, multiplicado por la pendiente de la recta de regresión de mínimos cuadrados ortogonales, define el valor mínimo del intervalo de confianza al $(1-\alpha)100$ de dicha pendiente.

De lo anterior se puede concluir que si:

$$p_1^* \geq p_1$$

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal.

Caso 2

$$\delta_1 \leq \delta < 1$$

Teorema 6.4.2

Bajo la condición:

$$\delta_1 \leq \delta < 1$$

Considerando la razón de pendientes entre la recta de regresión de Deming y la recta de regresión ortogonal.

Sean:

$$p_2^*=\frac{k-\delta_1+\sqrt{(k-\delta_1)^2+4k\delta_1\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

$$p_2 = 1+\frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}$$

Sí:

$p_2^* \leq p_2$

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal.

Demostración:

De acuerdo a (37) la razón entre pendientes de la recta de regresión de Deming y la pendiente de la recta de regresión ortogonal es:

$$\frac{k-\delta+\sqrt{(k-\delta)^2+4k\delta\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

Si a esa relación aplicada a la razón extrema de cocientes de variables se le llama p_2^*:

$$p_2^* = \frac{k-\delta_1+\sqrt{(k-\delta_1)^2+4k\delta_1\rho^2}}{k-1+\sqrt{(k-1)^2+4k\rho^2}}$$

Por (45) se sabe que:

$$p_2 = 1 + \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0}$$

Donde p_2 es el factor que, multiplicado por la pendiente de la recta de regresión de mínimos cuadrados ortogonales, define el valor máximo del intervalo de confianza al $(1-\alpha)100$ de dicha pendiente.

De lo anterior se puede concluir que si:

$p_2^* \leq p_2$

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza $(1-\alpha)100$ de la pendiente de la regresión ortogonal.

6.4.1 Ejercicio de aplicación.

Retómese el ejemplo propuesto en 6.3.2

Sean doscientas observaciones (x, y) con las siguientes características:

$S_{xx} = 16$

$S_{yy} = 25$

$\rho = 0.85$

$S_{xy} = 17$

$n = 200$

Al realizar los cálculos con las expresiones respectivas se obtiene:

k =1.5625

b_o = 1.29914738

$V(b_o)$ = 0.00327611

Intervalo de confianza del 95% para la pendiente de la recta de regresión ortogonal:

$[1.18696223;\ 1.41133252]$

$$p_1 = 1 - \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} = 0.9136471$$

$$p_2 = 1 + \frac{Z_{1-\alpha/2}\sqrt{V(b_0)}}{b_0} = 1.0863529$$

Caso 1:

Supóngase que se sabe que el cociente de varianzas de errores (δ) es:

$1 < \delta \leq 2$.

$$p_1^* = \frac{1.5625 - 2 + \sqrt{(1.5625-2)^2 + 4(1.5625)(2)(0,85)^2}}{1.5625 - 1 + \sqrt{(1.5625-1)^2 + 4(1.5625)(0,85)^2}}$$

p_1^* =0.94157054

$p_1^* > p_1$

Conclusión:

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza del 95% de la pendiente de la regresión ortogonal.

Caso 2

Supóngase que se sabe que el cociente de varianzas de errores (δ) es:

$0.5 \leq \delta < 1.$

$$p_2^* = \frac{1.5625 - 0.5 + \sqrt{(1.5625 - 0.5)^2 + 4(1.5625)(0.5)(0{,}85)^2}}{1.5625 - 1 + \sqrt{(1.5625 - 1)^2 + 4(1.5625)(0{,}85)^2}}$$

$p_2^* =$ 1.0514784

$p_2^* < p_2$

Conclusión:

La pendiente de la recta de regresión de Deming queda incluida en el intervalo de confianza del 95% de la pendiente de la regresión ortogonal.

MÉTRICAS EN LA REGRESIÓN DE DEMING

En general las métricas se emplean para evaluar las calidades de los modelos, en este capítulo final se deducen las expresiones para calcular: Suma de cuadrados de los errores, error cuadrado medio, raíz del error cuadrado medio, coeficiente de determinación, y error absoluto medio para las siguientes modalidades de regresión bivariada de mínimos cuadrados: Deming en general, mínimos cuadrados ortogonales, mínimos cuadrados ordinarios y mínimos cuadrados horizontales.

7.1 Suma de cuadrados de los errores (SCE)

De acuerdo a (10):

$$SCE = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2S_{xx} - 2mS_{xy} + S_{yy}}$$

7.1.1 SCE recta de regresión de Deming en general

En 3.1 se demostró que en la recta de regresión de Deming:

$$m = \frac{-\delta}{b_D}$$

$$SCE = \frac{(m^2+1)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{m^2S_{xx} - 2mS_{xy} + S_{yy}}$$

Luego de reemplazar el valor de m se obtiene:

$$SCE = \frac{(\delta^2 + b_D^2)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{\delta^2 S_{xx} + 2b_D\delta S_{xy} + b_D^2 S_{yy}}$$

Donde:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{(\delta S_{xx} - S_{yy})^2 + 4\delta {S_{xy}}^2}}{2S_{xy}}$$

Error cuadrado medio (ECM)

$$ECM = \frac{(\delta^2 + b_D^2)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{n(\delta^2 S_{xx} + 2b_D\delta S_{xy} + b_D^2 S_{yy})}$$

Raíz del cuadrado medio (RECM)

$$RECM = \sqrt{\frac{(\delta^2 + b_D^2)\left(S_{xx}S_{yy} - S_{xy}^2\right)}{n(\delta^2 S_{xx} + 2b_D \delta S_{xy} + \ b_D^2 S_{yy})}}$$

7.1.2 SCE recta de regresión ortogonal

En (29) se demostró que:

$$SCE_o = \frac{S_{xx} + S_{yy} - \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2}$$

Error cuadrado medio (ECM)

$$ECM_o = \frac{S_{xx} + S_{yy} - \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2n}$$

Raíz del cuadrado medio (RECM)

$$RECM_o = \sqrt{\frac{S_{xx} + S_{yy} - \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4\left(S_{xy}\right)^2}}{2n}}$$

7.1.3 SCE recta de regresión de mínimos cuadrados ordinarios

De acuerdo a la deducción realizada en (28), puede afirmarse que:

$$SCE = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}} \qquad (\mathbf{46})$$

Error cuadrado medio (ECM)

$$ECM = \frac{S_{xx}S_{yy} - S_{xy}^2}{nS_{xx}}$$

Raíz del error cuadrado medio (RECM):

$$RECM = \sqrt{\frac{S_{xx}S_{yy} - S_{xy}^2}{nS_{xx}}}$$

7.1.4 SCE Recta de regresión de mínimos cuadrados horizontales

Haciendo $m = 0$, en (10), se obtiene:

$$SCE = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{yy}} \qquad (\mathbf{47})$$

Error cuadrado medio (ECM):

$$ECM = \frac{S_{xx}S_{yy} - S_{xy}^2}{nS_{yy}}$$

Raíz del error cuadrado medio (RECM):

$$RECM = \sqrt{\frac{S_{xx}S_{yy} - S_{xy}^2}{nS_{yy}}}$$

7.2 Coeficiente de determinación (R^2)

7.2.1 R^2 para la recta regresión de Deming en general:

Se demostró en (12) que:

$$R^2 = \frac{(b_D^2 + 1)\left(S_{xx} + 2\delta b_D S_{xy} + b_D^2 S_{yy}\right)}{(b_D^2 + \delta)^2\left(S_{xx} + S_{yy}\right)}$$

Donde b_D es la pendiente de la regresión de Deming:

$$b_D = \frac{S_{yy} - \delta S_{xx} + \sqrt{\left(S_{yy} - \delta S_{xx}\right)^2 + 4\delta S_{xy}^2}}{2S_{xy}}$$

δ es la relación de varianzas de los errores, tal como se definió anteriormente.

7.2.2 R^2 para la recta de regresión ortogonal:

Se demostró en 3.4:

$$R^2 = \frac{S_{xx} + S_{yy} + \sqrt{\left(S_{xx} - S_{yy}\right)^2 + 4S_{xy}^2}}{2\left(S_{xx} + S_{yy}\right)}$$

7.2.3 R^2 para la recta de regresión de mínimos cuadrados ordinarios

De acuerdo a (46):

$$SCE = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}$$

Cuando se trabaja con mínimos cuadrados ordinarios, la varianza de los errores solo se considera en "y"

$$R^2 = 1 - \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}S_{yy}}$$

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

$$R^2 = \rho^2$$

7.2.4 R^2 para la recta de regresión de mínimos cuadrados horizontales

De acuerdo a (47):

$$SCE = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{yy}}$$

Cuando se trabaja con mínimos cuadrados ordinarios, la varianza de los errores solo se considera en "x"

$$R^2 = 1 - \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}S_{yy}}$$

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

$$R^2 = \rho^2$$

7.3 Error absoluto medio

7.3.1 Error absoluto medio para la recta de regresión de Deming.

La suma de cuadrados de distancias de n puntos a una recta $y = a + bx$, cuando los errores se miden con pendiente m, de acuerdo a (5) es:

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

Pero cuando se trata de una recta de regresión de Deming

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

Es por ello que:

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

También se sabe que:

$$m = -\frac{\delta}{b_D}$$

Reemplazando:

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

Pero cuando se desea la suma de los errores absolutos (s)

$$s = \frac{\sqrt{b_D^2 + \delta^2}}{b_D^2 + \delta} \sum |y_i - \bar{y} - b_D(x_i - \bar{x})|$$

Y el error absoluto medio (EAM) será:

$$EAM = \frac{\sqrt{b_D^2 + \delta^2}}{n(b_D^2 + \delta)} \sum |y_i - \bar{y} - b_D(x_i - \bar{x})| \qquad (\mathbf{48})$$

7.3.2 Error absoluto medio para la recta de regresión ortogonal

Se retoma (48):

En este caso la pendiente es b_o y $\delta = 1$, realizando esos reemplazos:

$$EAM = \frac{1}{n\sqrt{b_o^2 + 1}} \sum \left| y_i - \overline{y} - b_o (x_i - \overline{x}) \right|$$

7.3.3 Error absoluto medio para la recta de regresión de mínimos cuadrados ordinarios

En la recta de mínimos cuadrados ordinarios la razón de varianzas de errores es infinito. Aplicando límite en (48):

$$EAM = \lim_{\delta \to \infty} \left(\frac{\sqrt{b_D^2 + \delta^2}}{n(b_D^2 + \delta)} \sum |y_i - \bar{y} - b_D (x_i - \bar{x})| \right)$$

$$EAM = \frac{1}{n} \sum |y_i - \bar{y} - b_D (x_i - \bar{x})|$$

Pero la pendiente en la recta de mínimos cuadrados ordinarios es:

$$b_D = \frac{S_{xy}}{S_{xx}}$$

Reemplazando:

$$EAM = \frac{1}{n} \sum \left| y_i - \bar{y} - \frac{S_{xy}}{S_{xx}} (x_i - \bar{x}) \right|$$

7.3.4 Para la recta de regresión de mínimos cuadrados horizontales

En la recta de regresión de mínimos cuadrados horizontales la razón de varianzas de errores es cero. Reemplazando en (48)

$$EAM = \frac{1}{n b_D} \sum |y_i - \bar{y} - b_D (x_i - \bar{x})|$$

Pero la pendiente en la recta de mínimos cuadrados ordinarios es:

$$b_D = \frac{S_{yy}}{S_{xy}}$$

Reemplazando:

$$EAM = \frac{1}{n} \sum \left| \frac{S_{xy}}{S_{yy}} (y_i - \bar{y}) - (x_i - \bar{x}) \right|$$

CONCLUSIONES

En el plano, se dedujeron los parámetros de las rectas de mínimos cuadrados, asumiendo una pendiente "m" para los errores y considerando varianzas y covarianzas de las variables involucradas.

Para una recta de mínimos cuadrados con una pendiente "m" para los errores, se encontraron las expresiones de SCT, SCM y SCE, en función de la pendiente de los errores, las varianzas y covarianzas. Se demostró que en todos los casos SCT=SCM+SCE. Aparecieron propiedades matemáticas las cuales se expusieron y estudiaron.

Con fundamento en la expresión de la pendiente de la recta de regresión de Deming, se encontró la pendiente de los errores. Puede afirmarse que la recta de regresión de Deming, se puede obtener aplicando mínimos cuadrados a los errores medidos con la pendiente hallada.

Con la teoría desarrollada en el trabajo, se dedujeron las expresiones de SCT, SCM y SCE para la recta de regresión de Deming y se obtuvo además el valor del coeficiente de determinación R^2 respectivo. El estudio de dicho coeficiente fue profundo y se determinó analíticamente que la recta de mínimos cuadrados ortogonales es la que posee bajo todas las circunstancias el máximo valor, valor que logró establecerse. Adicionalmente se estableció la forma como cambia la pendiente de la recta de regresión de Deming con el incremento o decremento de la razón de varianzas de errores de las variables.

Siendo parte del objetivo general del presente trabajo, el estudio geométrico de la recta de regresión de Deming, Se establecieron diferencias entre pendientes y ángulos de rectas notables dentro de la familia de rectas de Deming. Dichas diferencias se cuantificaron, se estudiaron a profundidad y se determinaron valores máximos y las condiciones bajo las cuales se generaban, se establecieron relaciones matemáticas sobre dichos valores.

Contemplando que cuando se trata de regresión simple bajo el método de mínimos cuadrados, las modalidades más frecuentes son las de mínimos cuadrados ordinarios y mínimos cuadrados ortogonales, se realizó un estudio comparativo del coeficiente de determinación entre las dos modalidades, y se determinaron con profundidad las diferencias para todos los casos posibles, se reafirmó que la recta de mínimos cuadrados ortogonales poseía siempre un valor de coeficiente de determinación mayor o igual a la recta de mínimos cuadrados ortogonales.

Otro aspecto matemático que se estudió fue la relación existente entre la pendiente de la recta de regresión de Deming y la recta de regresión ortogonal, determinándose su expresión, pero además se definieron valores máximos y mínimos de esa dicha relación, y el valor para la razón de varianzas de las variables involucradas (k) para la cual se establecía esos puntos críticos. Se practicó un estudio de esos valores críticos (o extremos) en función del coeficiente de correlación de Pearson.

Como complemento al estudio geométrico y matemático de la recta de regresión de Deming (objetivo general del trabajo), Se encontraron los valores de δ de una regresión de Deming que garantizan rectas contenidas en un intervalo de confianza de la recta de regresión ortogonal.

Se estudió el caso en que se conoce un intervalo para las razones de varianzas de errores (δ) y su chequeo para definir si se encuentra en un intervalo de confianza de la recta de regresión ortogonal.

Se establecieron métricas importantes en la recta de regresión de Deming y en las rectas notables que pertenecen a esa familia.

El trabajo demuestra en todos sus apartes que regresiones particulares como la recta de mínimos cuadrados ordinarios, mínimos cuadrados ortogonales y mínimos cuadrados horizontales son simplemente variantes de una familia denominada regresión de Deming.

El estudio demuestra que la recta de regresión de Deming, es simplemente una recta de mínimos cuadrados en la cual los errores tienen una pendiente definida, la cual se puede determinar fácilmente haciendo uso de las expresiones obtenidas.

Este estudio entrega la herramienta para calcular el coeficiente de determinación R^2 para cualquier recta de mínimos cuadrados, en función de la varianza, la covarianza y el cociente de varianzas de errores (δ) de las variables.

En rectas de mínimos cuadrados, las diferencias angulares y de pendientes entre dos modalidades definidas por su respectivo valor de (δ) es de suma importancia. Con las herramientas que suministra este trabajo, el investigador podrá precisar la diferencia cuando decide usar indistintamente y sin mayor argumentación una de ellas. Por el contrario, la evaluación de esa discrepancia, en caso de ser pequeña, permitirá al investigador escoger una modalidad cualquiera con la seguridad que la estimación no afectará significativamente el resultado. En la mayoría de los casos los investigadores, cuando deciden realizar regresión lineal de mínimos cuadrados. Si se trabaja con las expresiones deducidas, podrá observarse que las diferencias al usar desprevenidamente una de las modalidades de la regresión de Deming, puede llegar teóricamente hablando, hasta.

El capítulo 6, suministra elementos para decidir sobre el empleo de la regresión ortogonal, sobre otra modalidad de recta de Deming (más complicada y laboriosa) y permite conocer hasta que valores de cociente de varianzas de los errores las variables (δ) puede la regresión ser sustituida por el tipo ortogonal sin salir del intervalo de confianza. También lo demostrado en este aparte, suministra al investigador la teoría necesaria para optar por la regresión ortogonal, al inferir que su regresión estará contenida en el intervalo de confianza de esta posibilidad, cuando tiene conocimiento del intervalo del cociente de las varianzas de los errores de las variables (δ)

A parte de la métrica del coeficiente de determinación, R^2 para cualquier recta de regresión de Deming, se dedujo y se relacionaron otras métricas que serán útiles en la valoración del modelo.

Toda la teoría expuesta puede aplicarse a las regresiones de mínimos cuadrados en el plano.

En el desarrollo, aparece una recta, que se denominó de los máximos cuadrados ortogonales. Está recta fue objeto de un estudio en un trabajo previo del autor en [13] Villota A. Hidalgo A. Este lugar geométrico, corresponde a la recta que, pasando por el centroide de los puntos, se aleja mayormente de dichos puntos. Su importancia practica puede radicar por ejemplo en circunstancias en la cuales, si bien una red debe pasar por el centro de una población, dicha línea debe alejarse de las viviendas para evitar efectos de contaminación o radiación.

En general se ha practicado un estudio matemático y geométrico a la recta de regresión de Deming amplio inédito y novedoso que no ha sido abordado en tratados publicados.

RECOMENDACIONES

Como recomendaciones y expectativas para próximos trabajos relacionados con el presentado se plantean las siguientes:

1.- Realizar el estudio multivariado de regresiones inspiradas en la de Deming,

2.- El objeto de este estudio era fundamentalmente un estudio geométrico y matemático a la recta de regresión de Deming, propósito que se cumplió, sin embargo, surge como inquietud a futuro abordar con mayor profundidad temáticas inferenciales.

.

BIBLIOGRAFÍA

[1] Lehmann, Geometría Analítica. (1989) Edit. Limusa.

[2] Wexler Ch, Geometría analítica. Un enfoque vectorial, (1977) Montaner y Simon, S.A. Barcelona

[3] Apostol T.M. (2006) Cálculo de funciones de varias variables y algebra lineal con aplicaciones a las ecuaciones diferenciales y probabilidad. Editorial Reverte. Barcelona España.

[4] Canavos G.C. Probabilidad y Estadística (2007), Editorial Mc Graw Hill. México

[5] Kenney J.F. Keeping E.S. Mathematics of Statistics. Editorial Van Nostrand

[6] Walpole, Myers, Myers, Ye, Probabilidad y Estadística, Octava Edición, Pearson, Pretnice Hall,2007

[7] Deming W.L, Statistical Adjusment of data. Wiley. NY Statistical adjustment of data.Wiley, NY. (Dover Publications edition 1985)

[8] Adcock, R. J. (1878). "A problem in least squares". The Analyst. Annals of Mathematics.

[9] Coolidge, JL. (1913), Dos aplicaciones geométricas de las matemáticas de mínimos cuadrados. El mensual matemático estadounidense. Jstor.

[10] Jensen Ch.A. (2007), Deming regression. Diabetes Center, Gentofte, Denamark

[11] Soporte de Mintiab, Métodos y fórmulas para regresión ortogonal (2020).

[12] Zylberberg A.D, (2006), Probabilidad y Estadística. Edit. Nueva Librería.

[13] Villota A, Hidalgo A. Estudio comparativo de rectas de regresión obtenidas mediante el método de mínimos cuadrados ortogonales. Tesis de Maestría Universidad Tecnológica de Pereira. 2018.

[14] Willmott, Cort J.; Matsuura, Kenji (19 de diciembre de 2005). "Ventajas del error absoluto medio (MAE) sobre el error cuadrático medio (RMSE) en la evaluación del rendimiento medio del modelo". Climate Research. Vol 1.

[15] 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)

[16] Habibollah Khajehsharifi. Masoumeh Sadeghi. Eslam Pourbasheer. (2009) Spectrophotometric simultaneous determination of ceratine, creatinine, and uric acid in real samples by orthogonal signal correction–partial least squares regression. Monatshefte für Chemie - Chemical Monthly. June, Volume 140

[17] Rainer Haeckel, Werner Wosniok and Rainer Klauke (2013). Comparison of ordinary linear regression, orthogonal regression, standardized principal component analysis, Deming and Passing-Bablok approach for method validation in laboratory. Laboratoriums Medizin June. Journal of Laboratory Medicine.

[18] Sandra Donevska, Eva Fiserova, Karel Hron. (2011). On the Equivalence between Orthogonal Regression and Linear Model. Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica 50, 2 -19–27.

[19] Taliha Keles. (2018) Comparison of Classical Least Squares and Orthogonal Regression in Measurement Error Models, International Online Journal of Educational Sciences, 10(3), 200-214.