EDITORIAL ANDES COGNITIO

FUNDAMENTOS MATEMÁTICOS DE INVESTIGACIÓN OPERATIVA:

Modelos Matemáticos Aplicados al Diseño Experimental y la Optimización de Procesos

Moisés Arreguín Sámano
Ángel Leyva Ovalle
Johanna Dueñas Durán
Daniel Santiago Paredes Gaibor

Editorial Andes Cognitio

FUNDAMENTOS MATEMÁTICOS DE INVESTIGACIÓN OPERATIVA

Modelos Matemáticos Aplicados al Diseño Experimental y la Optimización de Procesos

Moisés Arreguín Sámano
Correo: marreguin@ueb.edu.ec
Orcid: https://orcid.org/0000-0001-9324-9400
Institución: Universidad Estatal de Bolívar

Ángel Leyva Ovalle
Correo: aleyvao@chapingo.mx
Orcid: https://orcid.org/0000-0003-1873-9797
Institución: Universidad Autónoma Chapingo (UACh-DiCiFo), México

Johanna Dueñas Durán
Correo: johanna.duenas@ueb.edu.ec
Orcid: https://orcid.org/0009-0002-2442-314X
Institución: Universidad Estatal de Bolívar, Bolívar

Daniel Santiago Paredes Gaibor
Correo: daniel.paredes@ueb.edu.ec
Orcid: https://orcid.org/0009-0005-3591-7008
Institución: Universidad Estatal de Bolívar, Bolívar

Editorial "ANDES COGNITIO S.A.S."
DEPARTAMENTO DE EDICIÓN

Editado y Distribuido por:

Editorial:	Andes Cognitio
Sello Editorial:	978-9942-7408
Teléfono:	0995805659
Web:	https://andescognitio.org
ISBN:	978-9942-7408-1-6
DOI:	https://doi.org/10.64230/92bnfm76

© Primera Edición
© Julio 2025
Impreso en Ecuador

Revisión de Ortografía
Lcda. Cristina Paola Chamorro Ortega

Diseño de Portada
Ing. Pamela Rosa Taco Hernández Mgs

Diagramación
Ing. Yoselyn Andrea Rogel Gaibor

Director Editorial
Ec. Juan F. Villacis U. Mgs.

Comité Científico Académico

Ec. Ligia Ximena Tapia Hermida Mgs.
Universidad Nacional de Chimborazo

Ing. Paula Alejandra Abdo Peralta Mgs.
Escuela Superior Politécnica de Chimborazo

Ing. Catherine Gabriela Frey Erazo Mgs.
Escuela Superior Politécnica de Chimborazo

Ing. Juan Enrique Ureña Moreno Mgs.
Escuela Superior Politécnica de Chimborazo

Ing. José Fernando Esparza Parra Mgs.
Escuela Superior Politécnica de Chimborazo

Ing. Alexis Gabriel Reinoso Haro Mgs.
Universidad Estatal de Bolívar

Constancia de Arbitraje

La Editorial Andes Cognitio, hace constar que este libro proviene de una investigación realizada por los autores, siendo sometido a un arbitraje bajo el sistema de doble ciego, de contenido y forma por jurados especialistas. Además, se realizó una revisión del enfoque, paradigma y método investigativo; desde la matriz epistémica asumida por los autores, aplicándose las normas APA, Séptima Edición, proceso de anti plagio en línea Compilatio, garantizándose así la cientificidad de la obra.

Comité Editorial

PRÓLOGO

El presente texto surge como una respuesta a la creciente necesidad de integrar herramientas estadísticas, econométricas y matemáticas avanzadas en la toma de decisiones dentro del ámbito de la ingeniería, las ciencias económicas y las ciencias aplicadas. Su contenido ha sido cuidadosamente estructurado para brindar a estudiantes, docentes e investigadores una guía clara y rigurosa en el diseño de experimentos, la modelación de procesos y la optimización de sistemas bajo restricciones.

Es importante destacar que este texto no solo pretende ser un manual de consulta técnica, sino también una invitación a reflexionar sobre la forma en que los modelos matemáticos pueden contribuir a mejorar la eficiencia, la calidad y la capacidad predictiva en diversos campos del conocimiento. A lo largo de estas páginas, el lector encontrará no solo procedimientos, fórmulas y demostraciones, sino también fundamentos conceptuales y propuestas de análisis que promueven el pensamiento crítico y la solución estructurada de problemas complejos.

Agradecemos a todos los lectores que se acercan a esta obra con ánimo de aprender, aplicar y transformar.

Los autores

ÍNDICE GENERAL

INTRODUCCIÓN

El desarrollo del pensamiento cuantitativo y el dominio de técnicas matemáticas aplicadas constituyen pilares esenciales en la formación universitaria de las ciencias naturales, sociales e ingenierías. En este contexto, el presente libro titulado Fundamentos Matemáticos de Diseño Experimental y Optimización de Procesos tiene como propósito principal ofrecer una herramienta académica sólida y actualizada que facilite la comprensión, formulación y aplicación de modelos matemáticos en la solución de problemas reales.

Este texto se estructura en seis capítulos interrelacionados que permiten una progresión lógica del conocimiento. El primer capítulo introduce al lector en los principios fundamentales de la estimación estadística, abordando con profundidad el enfoque de máxima verosimilitud, los mínimos cuadrados ordinarios y su relación con la función de regresión poblacional. En el segundo capítulo, se desarrolla el análisis de modelos de superficie de respuesta, mostrando la utilidad de funciones lineales, cuadráticas y cúbicas para modelar relaciones complejas entre variables.

El tercer capítulo se enfoca en las transformaciones matemáticas y el uso de espacios vectoriales, fundamentos indispensables para comprender las restricciones en procesos de optimización. Posteriormente, en el cuarto capítulo, se aborda el cálculo de integrales múltiples aplicado a la estimación de parámetros, así como el análisis de varianza y la bondad de ajuste en modelos de regresión.

Los capítulos quinto y sexto presentan herramientas para la predicción mediante series temporales y el análisis de regresión múltiple, integrando

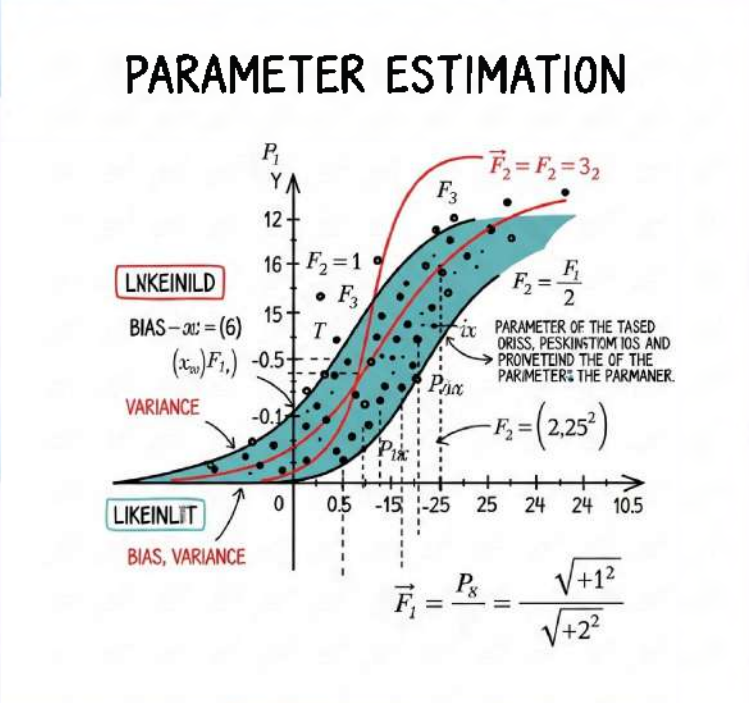
conceptos clave como multicolinealidad, valores atípicos y transformación de datos. Estas secciones se acompañan del uso de software como Microsoft Excel, que facilita la implementación práctica de los modelos presentados.

Este libro está diseñado tanto para estudiantes de pregrado como de posgrado, investigadores y profesionales que buscan mejorar sus competencias en el análisis cuantitativo. Al ofrecer una combinación equilibrada de teoría y aplicación, se espera que contribuya significativamente al desarrollo de habilidades analíticas y a la toma de decisiones fundamentadas en evidencia.

EDITORIAL ANDES COGNITIO

CAPÍTULO I

ESTADÍSTICA APLICADA AL DISEÑO EXPERIMENTAL Y ESTIMACIÓN DE PÁRAMETROS



CAPÍTULO I

ESTADÍSTICA APLICADA AL DISEÑO EXPERIMENTAL Y ESTIMACIÓN DE PARÁMETROS

1. Máxima Verosimilitud (MV)

1.1. Definición

La estimación de máxima verosimilitud del modelo de regresión con dos variables parte de suponer que el modelo con dos variables es $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y_i son independientes y normalmente distribuidas con media $\beta_1 + \beta_2 X_i$ y varianza $= \sigma^2$, pues $Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2)$. Como resultado, la función de densidad de probabilidad conjunta de $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$ dadas las medias y varianzas anteriores, se escribe como:

$$f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2)$$

Dada la independencia de las Y, está función de densidad de probabilidad conjunta se escribe como el producto de las n funciones de densidad individuales como:

$$f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2) =$$

$$f(Y_1 | \beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_2 | \beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_3 | \beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_4 | \beta_1 + \beta_2 X_i, \sigma^2) \dots f(Y_n | \beta_1 + \beta_2 X_i, \sigma^2)$$

Donde:

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\}}$$

La anterior ecuación, es la función de densidad de una variable normalmente distribuida con media y varianza dadas. Sin embargo, al sustituir la ecuación anterior en ecuación

$$f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n|\beta_1 + \beta_2 X_i, \sigma^2) =$$

$f(Y_1|\beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_2|\beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_3|\beta_1 + \beta_2 X_i, \sigma^2)\, f(Y_4|\beta_1 + \beta_2 X_i, \sigma^2) \dots f(Y_n|\beta_1 + \beta_2 X_i, \sigma^2)$ se tiene:

$$f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n|\beta_1 + \beta_2 X_i, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^2} e^{\left\{-\frac{1}{2}\ \Sigma\frac{(Y_i-\beta_1-\beta_2 X_i)^2}{\sigma^2}\right\}}$$

Si se conocen o están dadas $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$, pero no se conocen β_1, β_2 y σ^2, la función en ecuación anterior se llama Función de Verosimilitud, denotada con $FV(\beta_1, \beta_2, \sigma^2)$ y es escrita como, pues si se conocen β_1, β_2 y σ^2, pero no las Y_i, la siguiente ecuación representa la función de densidad de probabilidad conjunta (probabilidad de observar conjuntamente las Y_i):

$$FV(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^2} e^{\left\{-\frac{1}{2}\ \Sigma\frac{(Y_i-\beta_1-\beta_2 X_i)^2}{\sigma^2}\right\}}$$

1.2. Método de Máxima Verosimilitud

El Método de Máxima Verosimilitud consiste en estimar los parámetros desconocidos de manera que la probabilidad de observar las Y dadas sea lo más alta (máxima) posible. Entonces, se tiene que encontrar el máximo de la función en ecuación anterior.

Es un ejercicio sencillo de cálculo diferencial. Sin embargo, la diferenciación es más fácil expresar esta ecuación en términos de la función logaritmo o log de la siguiente manera, pues log es una función monótona y Ln FV alcanza su máximo valor en mismo punto que FV:

$$\text{Ln FV} = -n\text{Ln}\sigma - \frac{n}{2}\text{Ln}(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$$

$$= -\frac{n}{2}\text{Ln}\sigma^2 - \frac{n}{2}\text{Ln}(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$$

Al diferenciar parcialmente la ecuación anterior respecto a β_1, β_2 y σ^2 se obtiene:

$$\frac{\partial\ \text{Ln FV}}{\partial\beta_1} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)\,(-1)$$

$$\frac{\partial\ \text{Ln FV}}{\partial\beta_2} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)\,(X_i)$$

$$\frac{\partial\ \text{Ln FV}}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(Y_i - \beta_1 - \beta_2 X_i)^2$$

Se iguala estas ecuaciones a cero, primera condición de primer orden para optimización se deja que $\tilde{\beta}_1$, $\tilde{\beta}_2$ y $\tilde{\sigma}^2$, se usa $\tilde{}$ (tilde) para estimadores de MV mientras que $\hat{}$ (acento circunflejo) para estimadores de MCO, denotan estimadores de MV para obtener:

$$\frac{1}{\tilde{\sigma}^2}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right) = 0$$

$$\frac{1}{\tilde{\sigma}^2}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right) X_i = 0$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right)^2 = 0$$

Después de simplificar, las ecuaciones $\frac{1}{\tilde{\sigma}^2}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right) = 0$ y $\frac{1}{\tilde{\sigma}^2}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right) X_i = 0$ llevan a:

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2\sum X_i$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2$$

1.3. Ecuaciones Normales de teoría de Mínimos Cuadrados

Las que son, precisamente, Ecuaciones Normales de teoría de Mínimos Cuadrados obtenidas en MCO:

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \approx \sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \approx \sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

Por lo tanto, los estimadores de MV, $\tilde{\beta}$, son iguales que estimadores de MCO, $\hat{\beta}$, dados en:

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \overline{Y} - \hat{\beta}_2 \overline{X}$$

Esta igualdad no es casual, pues al examinar la verosimilitud $\text{Ln FV} = -n\text{Ln}\sigma - \frac{n}{2}\text{Ln}(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} = -\frac{n}{2}\text{Ln}\sigma^2 - \frac{n}{2}\text{Ln}(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$ el último término entra con signo negativo.

Entonces, la maximización de esta ecuación equivale a la minimización de este término, que es justo el enfoque de Mínimos Cuadrados en $\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$.

Al sustituir estimadores de MV $(= \text{MCO})$ en $-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum \left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right)^2 = 0$ y simplificar, se obtiene el estimador MV de $\tilde{\sigma}^2$:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum\left(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i\right)^2 = \frac{1}{n}\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 = \frac{1}{n}\sum \hat{u}_i^2$$

Con base en esta ecuación, el estimador de MV $\tilde{\sigma}^2$ difiere del estimador de MCO $\hat{\sigma}^2 = \left[\frac{1}{(n-2)}\right]\sum \hat{u}_i^2$ que, $Y_i = \beta_1 + \beta_2 X_i + u_i \Rightarrow \overline{Y} = \beta_1 + \beta_2 \overline{X}_i + \bar{u}$tal que al restar la última respecto a la primera ecuación se obtiene $y_i = \beta_2 x_i + (u_i - \bar{u})$ aunque $\hat{u}_i = y_i - \hat{\beta}_2 x_i$.

Tal que al sustituir esta penúltima ecuación en última es $\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i$ aunque si se reúnen términos, se eleva al cuadrado y se suman ambos lados se obtiene $\sum \hat{u}_i = \left(\hat{\beta}_2 - \beta_2\right)^2 \sum x_i^2 + \sum (u_i - \bar{u})^2 - 2\left(\hat{\beta}_2 - \beta_2\right)\sum x_i (u_i - \bar{u})$ tal que al tomar valores esperados en ambos lados se tiene $E\left(\sum \hat{u}_i^2\right) = \sum x_i^2 \, E\left(\hat{\beta}_2 - \beta_2\right)^2 + E[\sum (u_i - \bar{u})^2] - 2E\left[\left(\hat{\beta}_2 - \beta_2\right)\sum x_i (u_i - \bar{u})\right] = \sum x_i^2 \, Var\left(\hat{\beta}_2\right) + (n-1)Var(u_i) - 2E[\sum x_i u_i \, (x_i u_i)] = \sigma^2 + (n-1)\sigma^2 - 2E\left[\sum k_i x_i u_i^2\right] = (n-2)\sigma^2$.

Donde sí se usa la definición de k_i en ecuación $\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i \, Y_i \Rightarrow k_i = \frac{x_i}{(\sum x_i^2)}$ y relación dada en ecuación $\hat{\beta}_2 = \sum k_i \, (\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i \, X_i + \sum k_i u_i = \beta_2 + \sum k_i u_i$ tal que al obtener valores esperados de ecuación anterior para ambos lados y advertir que las k_i al ser no estocásticas pueden tratarse de como constantes para obtener $E\left(\hat{\beta}_2\right) = \beta_2 + \sum k_i E(u_i) = \beta_2$ pues $E(u_i) = 0$ por suposición tal que $\hat{\beta}_2$ es un estimador insesgado de β_2 así, de la misma manera, se demuestra que $\hat{\beta}_1$ es un estimador insesgado de β_1, es un estimador insesgado de σ^2.

Por tanto, el estimador de MV de σ^2 es sesgado aunque su magnitud se determina fácilmente pues se toma la esperanza matemática de $\tilde{\sigma}^2 = \frac{1}{n}\sum\left(Y_i - \tilde{\beta}_1 -$

$\tilde{\beta}_2 X_i)^2 = \frac{1}{n}\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 = \frac{1}{n}\sum \hat{u}_i^2$ en ambos lados de la ecuación y se obtiene:

$$E(u_i) = \frac{1}{n}E\left(\sum \hat{u}_i^2\right) = \left(\frac{n-2}{n}\right)\sigma^2$$

Con la ecuación $E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2$ demuestra que $\tilde{\sigma}^2$ está sesgado hacia abajo (subestima el verdadero valor σ^2) en muestras pequeñas. Sin embargo, a medida que se incrementa indefinidamente n, el tamaño de la muestra, el segundo término en $E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2$, factor sesgo, tiende a cero.

Entonces, asintóticamente (muestra muy grande), $\tilde{\sigma}^2$ también es insesgada. Es decir, el $\lim E(\tilde{\sigma}^2) = \sigma^2$ a medida que $n \to \infty$. Se puede demostrar que $\tilde{\sigma}^2$ es un estimador consistente o a medida que n aumenta indefinidamente, $\tilde{\sigma}^2$ converge hacia su verdadero valor σ^2.

1.4. Estimación de MCO y MV

La Estimación de MCO y MV de Coeficientes de Regresión Parcial implica que para estimar los parámetros del modelo de regresión con tres variables en $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ se considera primero el método de mínimos cuadrados ordinarios (MCO) y método de máxima verosimilitud (MV).

1.5. Función de Regresión Poblacional (FRP)

Para encontrar estimadores de MCO, se escribe primero la función de regresión muestral (FRM: la Función de Regresión Muestral (FRM) aborda una muestra de valores Y correspondientes a valores fijos de X.

Por tanto, su labor es estimar la Función de Regresión Poblacional (FRP) con base en información muestral; aunque, quizá no pueda al calcularse la FRP con "precisión" debido a fluctuaciones muestrales. Entonces, el diagrama de dispersión se usa y, simultáneamente, se trazan dos líneas de regresión muestral con el fin de "ajustar" razonablemente bien las dispersiones FRM_1 se basa en la primera muestra y FRM_2 en la segunda.

Sin embargo, no hay forma de estar completamente seguro de que alguna de las líneas de regresión de esta figura representa la verdadera recta o curva poblacional, pues las líneas de regresión en la figura anterior se conocen como líneas de regresión muestral tal que representan la línea de regresión poblacional, pero por fluctuaciones muestrales son el mejor de los casos sólo una aproximación de la verdadera regresión poblacional (RP).

Entonces, se obtendrán N FRM diferentes para N muestras diferentes y estas FRM no por fuerza son iguales. La FRP en que se basa la línea de RP se desarrolla en el concepto de Función de Regresión Muestral (FRM) para representar la línea de regresión muestral, siendo la contraparte muestral de $E(Y|X_i) = \beta_1 + \beta_2 X_i$ la ecuación $\widehat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ tal que $\widehat{Y}_i$ es estimador de $E(Y|X_i)$.

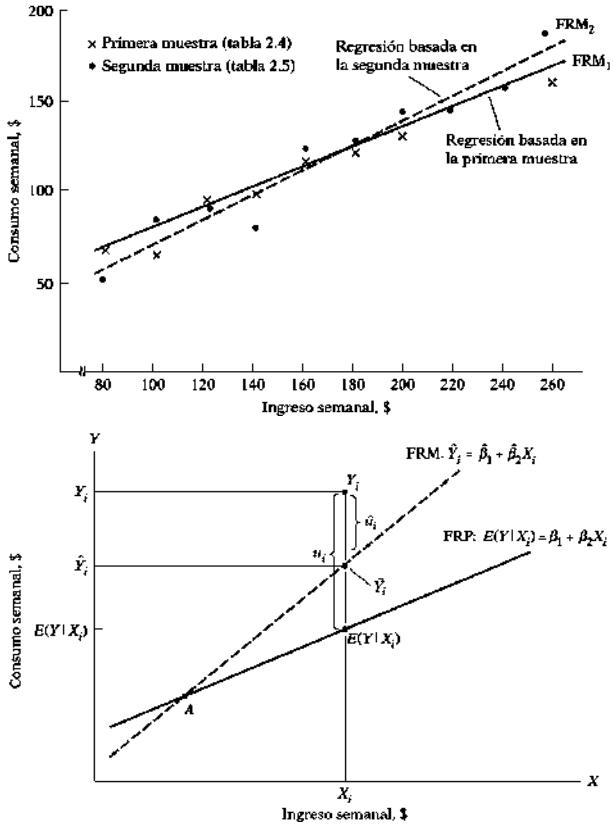
Un estimador o estadístico muestral no es más que una regla, fórmula o método para estimar el parámetro poblacional a partir de información muestral. Un valor numérico particular obtenido por estimador en un análisis se conoce como estimación.

Tal como FRP se expresa en dos formas equitativas: $E(Y|X_i) = \beta_1 + \beta_2 X_i$ y $Y_i = E(Y|X_i) + u_i = \beta_1 + \beta_2 X_i + u_i$, la FRP $\widehat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ se expresa estocásticamente como $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$.

Por lo tanto, el objetivo principal del análisis de regresión es estimar la Función de Regresión Poblacional (FRP) $Y_i = \beta_1 + \beta_2 X_i + u_i$ con base en Función de

Regresión Muestral (FRM) $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$, pues son más frecuentes los casos en que el análisis se basa en una sola muestra tomada de la población tal que debido a fluctuaciones muestrales, la estimación de la FRP basada en FRM es, en el mejor de los casos, una aproximación correspondiente a la FRP (el concepto fundamental del análisis de regresión es el de Función de Esperanza Condicional (FEC) o Función de Regresión Poblacional (FRP).

Figura 1.1. Representación gráfica



Fuente: (Gujarati & Porter, 2010)

Para $X = X_i$ se tiene una observación muestral, $Y = Y_i$ que en términos de FRM, Y_i observada es $Y_i = \widehat{Y}_i + \hat{u}_i$ y en términos de RFP $Y_i = E(Y|X_i) = u_i$. Su objetivo del análisis de regresión es averiguar la forma en que varía el valor promedio de

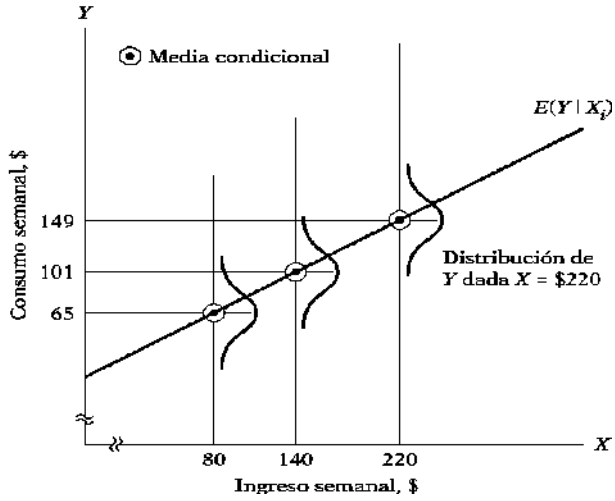
la variable dependiente (regresada) según el valor dado de la variable explicativa (regresora).

1.6. Métodos de estimación

Se estiman dos métodos de estimación frecuentes: Mínimos Cuadrados Ordinarios (MCO) y Máxima Verosimilitud (MV). Hay ocasiones en que la FRP de dos variables tiene la forma $Y_i = \beta_2 + X_i + u_i$ tal que el término del intercepto está ausente o es cero, que explica el nombre Regresión a través del origen e incluso al generalizar la regresión poblacional (FRP) de dos variables $Y_i = E(Y|X_i) + u_i = \beta_1 + \beta_2 X_i + u_i$ se escribe la FRP de tres variables como $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ de $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

Figura 1.2. Método de estimación



Fuente: (Gujarati & Porter, 2010)

De lo anterior, específicamente las figuras, cada media condicional $E(Y|X_i)$ es función de X_i, donde X_i es un valor dado de X. Simbólicamente es $E(Y|X_i) =$

$f(X_i)$ tal que $f(X_i)$ denota alguna función de variable explicativa X. $E(Y|X_i)$ es una función lineal de X_i.

1.7. Función de Esperanza Condicional

La ecuación $E(Y|X_i) = f(X_i)$ se llama Función de Esperanza Condicional (FEC), Función de Regresión Poblacional (FRP) o Regresión Poblacional Lineal (RP). Esta función sólo denota que el valor esperado de distribución de Y dada X_i se relaciona funcionalmente con X_i o, en otras palabras, cómo la media o respuesta promedio de Y varia con X. $E(Y|X_i) = \beta_1 + \beta_2 X_i$.

Donde $\hat{u}_i$ es término residual, contraparte muestral del término de perturbación estocástico u_i. Como se vio anteriormente, el procedimiento MCO consiste en seleccionar valores desconocidos de parámetros de forma que la suma de cuadrados de residuos ($SCR = \sum \hat{u}_i^2$) sea lo más pequeña posible. Simbólicamente:

$$\min \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2$$

Donde la expresión para SCR se obtiene por simple manipulación algebraica de $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$, siendo el procedimiento más directo para obtener estimadores que reducen $\min \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2$ diferenciarla parcialmente respecto de las incógnitas, igualar a cero las expresiones resultantes y resolverlas al mismo tiempo.

De $\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2$ respecto de tres incógnitas e igualar a cero las ecuaciones resultantes se obtiene $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-1)$, $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-1 * X_{2i})$ y $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2 \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-1 * X_{3i})$, simplificando estas ecuaciones se

obtienen las ecuaciones normales, comparables con ecuaciones $\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$ y $\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$, respectivamente:

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}_2 + \hat{\beta}_3 \overline{X}_3$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$$

A partir de ecuación $\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}_2 + \hat{\beta}_3 \overline{X}_3$ se deduce:

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_2 - \hat{\beta}_3 \overline{X}_3$$

Entendido como estimador de MCO del intercepto poblacional $\hat{\beta}_1$. Conforme a permitir que letras minúsculas denoten desviaciones de medias muestrales, se derivan las fórmulas de ecuaciones normales $\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}_2 + \hat{\beta}_3 \overline{X}_3$, $\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$ y $\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$:

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}\, x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}\, x_{3i})^2}$$

Estas ecuaciones, simétricas por naturaleza pues una se obtiene de la otra mediante cambio de roles de X_2 y X_3 mientras que sus denominadores son idénticos y, por lógica, el caso de tres variables es una extensión natural del caso de dos variables, dan estimadores de MCO de coeficientes de regresión parcial poblacionales $\hat{\beta}_2$ y $\hat{\beta}_3$, respectivamente.

Varianzas y errores estándar de estimadores de MCO se obtienen después de los coeficientes de regresión parcial y errores estándar de estimadores, pues $\mathrm{Var}(\hat{\beta}_2) = \mathrm{E}[\hat{\beta}_2 - \mathrm{E}(\hat{\beta}_2)]^2 = \mathrm{E}(\hat{\beta}_2 - \beta_2)^2$, pues $\mathrm{E}(\hat{\beta}_2) = \beta_2$, tal que $= \mathrm{E}(\sum k_i u_i)^2$ con ecuación $\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i = \beta_2 + \sum k_i u_i$ tal que $= \mathrm{E}(k_1^2 u_1^2 + k_2^2 u_2^2 + k_3^2 u_3^2 + k_4^2 u_4^2 + \cdots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + 2k_3 k_4 u_3 u_4 + 2k_5 k_6 u_5 u_6 + 2k_7 k_8 u_7 u_8 + \cdots + 2k_{n-1} k_n u_{n-1} u_n)$ aunque por supuestos de $\mathrm{E}(u_i^2) = \sigma^2 \; \forall \; i$ y $\mathrm{E}(u_i u_j) = 0, i \neq j$ se deduce que $\mathrm{Var}(\hat{\beta}_2) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum x_i^2}$, pero con definición de k_i^2 se concluye que $\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$, pero también $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$. Igual que en caso de dos variables. Se requiere errores estándar para crear intervalos de confianza y probar hipótesis estadísticas.

Las fórmulas pertinentes son:

$$\mathrm{Var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}\right] * \sigma^2$$

$$\mathrm{ee}(\hat{\beta}_1) = +\sqrt{\mathrm{Var}(\hat{\beta}_1)}$$

$$\mathrm{ee}(\hat{\beta}_2) = +\sqrt{\mathrm{Var}(\hat{\beta}_2)}$$

$$\mathrm{ee}(\hat{\beta}_3) = +\sqrt{\mathrm{Var}(\hat{\beta}_3)}$$

$$\mathrm{Var}(\hat{\beta}_2) = \left[\frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\right] * \sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$\mathrm{Var}(\hat{\beta}_3) = \left[\frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}\right] * \sigma^2 = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

$$\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23} * \sigma^2}{(1 - r_{23}^2) * \left(\sqrt{\sum x_{2i}^2}\right) * \left(\sqrt{\sum x_{3i}^2}\right)}$$

En todas las fórmulas, σ^2 es varianza homocedástica de perturbaciones poblacionales u_i. De igual forma, las derivaciones de estas fórmulas son más sencillas con notación matricial, pues los métodos matriciales permiten desarrollar fórmulas no sólo para varianza de $\hat{\beta}_i$, cualquier elemento dado de $\hat{\beta}$, sino también para varianza entre dos elementos $\hat{\beta}$ cuales quiera, como $\hat{\beta}_i$ y $\hat{\beta}_j$ tal que se requieren estas varianzas-covarianzas para fines de inferencias estadística.

Por definición, matriz varianza-covarianza de $\hat{\beta}$ es:

$$E(uu') = \begin{bmatrix} \mathrm{Var}(u_1^2) & E(u_1u_2) & \dots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \dots & E(u_2u_n) \\ \dots & \dots & \dots & \dots \\ E(u_nu_1) & E(u_nu_2) & \dots & E(u_n^2) \end{bmatrix}$$

Tal que:

$$\mathrm{Var} - \mathrm{Cov}(\hat{\beta}) = E\left\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\right\}$$

$$= \begin{bmatrix} \mathrm{Var}(\hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_2) & \dots & \mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \mathrm{Var}(\hat{\beta}_k) \end{bmatrix}$$

Esto se demuestra por:

$$\mathrm{Var} - \mathrm{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \Rightarrow \underset{(k*1)}{\hat{\beta}} = \underset{(k*k)}{(X'X)^{-1}} * \underset{(k*n)}{X'} \underset{(n*1)}{y}$$

Por consiguiente:

$$\hat{u}'\hat{u} = (y - X\hat{\beta})' * (y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

Con reglas de derivación:

I. $a' = [a_1 \quad a_2 \quad \cdots \quad a_n]$ es $\vec{u}$ renglón de números y $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ es $\vec{\mu}$ columna de variables $x_1, x_2, \ldots, x_n$ tal que $\frac{\partial(a'X)}{\partial x} = a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$.

II. Considera matriz $x'A_x = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ n_n \end{bmatrix}$

Tal que, $\frac{\partial(x'A_x)}{\partial x} = 2A_x$ que es $\vec{u}$ columna de n elementos, o $\frac{\partial(x'A_x)}{\partial x} = 2x'A$ que es $\vec{\mu}$ renglón de n elementos.

Con base en esto, $\underset{(k*k)}{\text{Var} - \text{Cov}(\hat{\beta})} = \sigma^2 \underset{(k*k)}{(X'X)^{-1}} \Rightarrow \underset{(k*1)}{\hat{\beta}} = \underset{(k*k)}{(X'X)^{-1}} * \underset{(k*n)}{X'} \underset{(n*1)}{y}$, $\hat{u}'\hat{u} = (y - X\hat{\beta})' * (y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$, reglas de diferenciación matricial y $\frac{\partial(\hat{u}'\hat{u})}{\partial x} = -2X'y + 2X'X\hat{\beta} = 0 \Rightarrow (X'X)\hat{\beta} = X'y$ tal que $\hat{\beta} = (X'X)^{-1}X'y$ siempre que exista su inversa, la matriz anterior de Varianza-Covarianza se obtiene a partir de la fórmula:

$$\text{Var} - \text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Donde σ^2 es varianza homocedástica de u_i y $(X'X)^{-1}$ es matriz inversa que aparece en ecuación $\underset{(k*1)}{\hat{\beta}} = \underset{(k*k)}{(X'X)^{-1}} * \underset{(k*n)}{X'} \underset{(n*1)}{y}$, que da estimador de MCO ($\hat{\beta}$), se obtiene:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Se sustituye $y = X\beta + u$ en expresión anterior se obtiene:

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u \Rightarrow$$
$$\hat{\beta} - \beta = (X'X)^{-1}X'u.$$

Por definición,

$$\text{Var} - \text{Cov}(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = E\left\{\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'\right\} =$$

$E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$ aprovechando que $(AB)' = A'B'$ y las X no son estocásticas, al tomar el valor esperado de ecuación anterior implica:

$$\text{Var} - \text{Cov}(\hat{\beta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}$$

Tal que, al derivar el resultado anterior se emplea el supuesto que $E(uu') = \sigma^2 I$.

Con base en todo lo anterior, se verifica que un estimador insesgado de σ^2 está dado por:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3}$$

Los grados de libertad son $(n - 3)$, pues para calcular $\sum \hat{u}_i^2$ se debe estimar primero β_1, β_2 y β_3, que consumen 3 gl. Tal que en caso de cuatro variables, los gl será $(n - 4)$.

Entonces, el estimador $\hat{\sigma}^2$ se estima mediante $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$ una vez que se dispone de residuos, pero también se obtiene más rápido con la relación siguiente, pues la derivación de esta ecuación es $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$ (letras minúsculas indican desviaciones respecto de valores de media)

entonces $\sum \hat{u}_i^2 = \sum(\hat{u}_i\hat{u}_i) = \sum \hat{u}_i(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) = \sum \hat{u}_i y_i = \sum y_i \hat{u}_i = \sum y_i(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$ tal que $\sum \hat{u}_i x_{2i} = \sum \hat{u}_i x_{3i} = 0$:

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

Esta última ecuación sería la contraparte de tres variables de la relación dada en, recordando que $\sum \hat{u}_i^2$ (SCR) $=$ SCT $-$ SCE, en $\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$.

Complementariamente, la derivación de k ecuaciones normales o simultáneas señalan que al diferenciar $\sum \hat{u}_i^2 = \sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \hat{\beta}_4 X_{4i} \dots - \hat{\beta}_k X_{ki})^2$ parcialmente respecto a β_1, β_2, β_3, ... β_k, se obtiene:

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \hat{\beta}_4 X_{4i} \dots - \hat{\beta}_k X_{ki})(-1)_{\text{(Derivada interna)}}$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \hat{\beta}_4 X_{4i} \dots - \hat{\beta}_k X_{ki})(-1 * X_{2i})_{\text{(Derivada interna)}}$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \hat{\beta}_4 X_{4i} \dots - \hat{\beta}_k X_{ki})(-1 * X_{3i})_{\text{(Derivada interna)}}$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \hat{\beta}_4 X_{4i} \dots - \hat{\beta}_k X_{ki})(-1 * X_{4i})_{\text{(Derivada interna)}}$$

… … … …

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_k} = 2\sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{ki} \dots - \hat{\beta}_k X_{ki}\right)(-1 * X_{ki})_{\text{(Derivada interna)}}$$

Se igual a cero las derivadas parciales anteriores, se reordena los términos y se obtiene las k ecuaciones normales siguientes:

$$\begin{aligned}
n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \hat{\beta}_4 \sum X_{4i} + \cdots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\
\hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i}X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i}X_{ki} &= \sum X_{2i}Y_i \\
\hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{3i}X_{2i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i}X_{ki} &= \sum X_{3i}Y_i \\
\dots\ \dots\ \dots\ \dots & \\
\hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki}X_{2i} + \hat{\beta}_3 \sum X_{ki}X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki}Y_i
\end{aligned}$$

Observe que a partir de la primera ecuación del conjunto de ecuaciones anteriores, resulta $\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_2 - \hat{\beta}_3\overline{X}_3 - \hat{\beta}_4\overline{X}_4 - \cdots - \hat{\beta}_k\overline{X}_k$, estimador de MCO del intercepto poblacional β_1. En forma matricial las ecuaciones anteriores equivalen a:

$$\underbrace{\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \dots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \dots & \sum X_{3i}X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \dots & \sum X_{ki}^2 \end{bmatrix}}_{(X'X)} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \dots \\ \hat{\beta}_k \end{bmatrix}}_{(\hat{\beta})}$$

$$= \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{32} & \dots & X_{3n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k1} & \dots & X_{k1} \end{bmatrix}}_{(X')} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{bmatrix}}_{(y)}$$

O, en forma compacta:

$$(X'X)(\hat{\beta}) = X'y \Rightarrow (X'X)^{-1}(X'X)(\hat{\beta}) = (X'X)^{-1}X'y$$

Como $(X'X)^{-1}(X'X) = I$ es matriz idéntica de orden $k * k$, se obtiene:

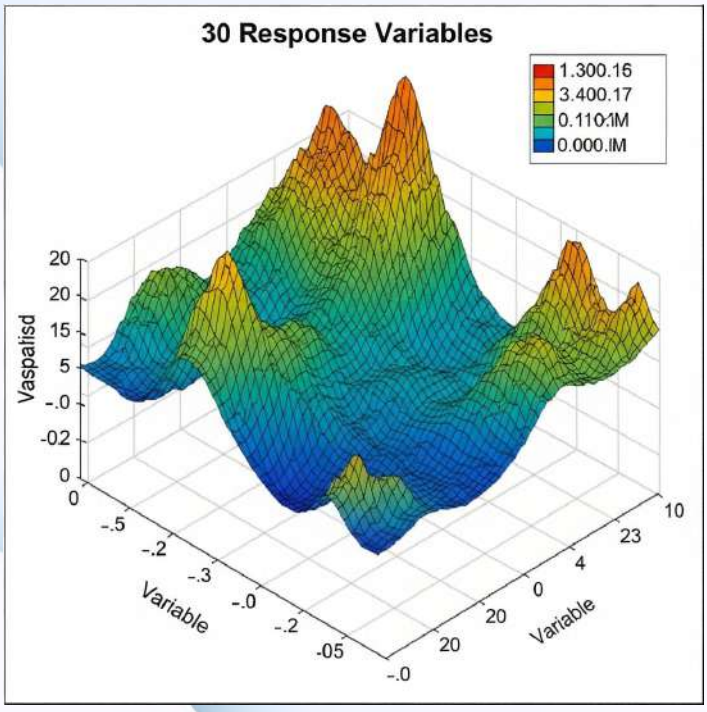
$$I\hat{\beta} = (X'X)^{-1}X'y \Rightarrow \therefore \underset{(k*1)}{\hat{\beta}} = \underset{(k*k)}{(X'X)^{-1}} \underset{(k*n)}{X'} \underset{(n*1)}{y}$$

Por desgracia, no es fácil obtener fórmulas como las anteriores. Muy a menudo los científicos o los economistas tienen que trabajar con grandes cantidades de datos para encontrar relaciones entre las variables de un problema. Una manera común de hacer esto es ajustar una curva entre los distintos puntos de datos.

EDITORIAL ANDES COGNITIO

CAPÍTULO II

MODELOS DE SUPERFICIE DE RESPUESTA



CAPÍTULO II

MODELOS DE SUPERFICIE DE RESPUESTA

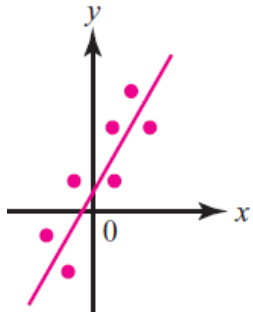
2. Modelo matemático-econométrico

Esta curva puede ser un modelo matemático-econométrico, suponiendo en cada caso que existen n puntos de datos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$ y la principal razón de su uso es que permiten obtener máximos o mínimos en las curvas (extremos relativos y/o extremos locales):

2.1. Recta o Lineal.

Suponga que se busca la recta de la forma $y = a + bx$ o $y = b + mx$ que mejor represente a n datos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$:

Figura 2.1. Modelo matemático-econométrico lineal



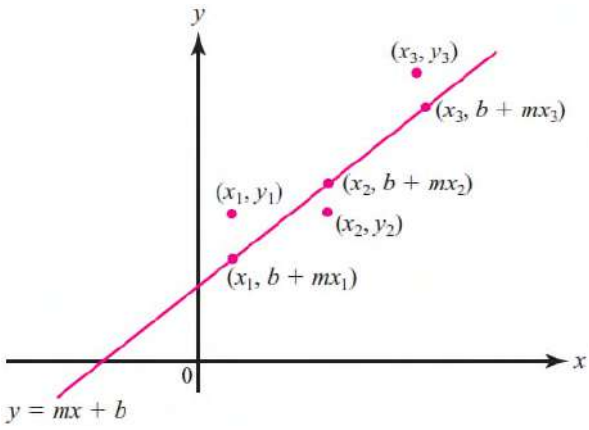
Fuente: (Mora, 2016)

La siguiente figura muestra lo que sucede usando tres datos, suponiendo que las variables x e y están relacionadas por la fórmula $y = b + mx$.

Por ejemplo, para $x = x_1$ el valor correspondiente de y es $b + mx_1$.

Sin embargo, esto es diferente del valor "real", $y = y_1$:

Figura 2.2. Distancia entre puntos



Fuente: (Mora, 2016)

En $\mathbb{R}^2$ la distancia entre puntos (a_1, b_1) y (a_2, b_2) está dada por la ecuación $d = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}$. En consecuencia, al determinar la manera de elegir la recta $y = a + bx$ o $y = b + mx$ que mejor se aproxima a datos dados, es razonable usar el criterio de seleccionar aquella que minimiza la suma de cuadrados de diferencias entre valores y de puntos y el valor y correspondientes a la recta.

La distancia entre (x_1, y_1) y $(x_1, b + mx_1)$ es $y_1 - (b + mx_1)$, el problema para n datos se puede establecer como "problema de mínimos cuadrados en caso de una recta", en que se encuentra números m y b tales que la suma $[y_1 - (b + mx_1)]^2 + [y_2 - (b + mx_2)]^2 + [y_3 - (b + mx_3)]^2 + [y_4 - (b + mx_4)]^2 + \cdots + [y_n - (b + mx_n)]^2$ sea mínima.

Para estos valores de m y b, la recta recta $y = a + bx$ o $y = b + mx$ se llama "Aproximación por recta de Mínimos Cuadrados a datos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \ldots, (x_n, y_n)$".

Una vez definido el problema se busca un método para encontrar la aproximación de mínimos cuadrados. Lo más sencillo es escribir en forma matricial. Si los

puntos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$ están todos sobre la recta $y = a + bx$, llamados colineales (dos o más elementos que se encuentran en una misma línea), entonces se tiene:

$$y_1 = a + bx_1 \qquad y_1 = b + mx_1$$

$$y_2 = a + bx_2 \qquad y_2 = b + mx_2$$

$$y_3 = a + bx_3 \qquad y_3 = b + mx_3$$

ó

$$y_4 = a + bx_4 \qquad y_4 = b + mx_4$$

$$\dots \qquad \dots$$

$$y_n = a + bx_n \qquad y_n = b + mx_n$$

O, con otra nomenclatura, $\mathbf{y} = \mathbf{Au}$. Donde:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \text{ y } u = \begin{pmatrix} b \\ m \end{pmatrix}$$

Si los puntos no son colineales, entonces $\mathbf{y} - \mathbf{Au} \neq \mathbf{0}$ y el problema se convierte en "forma vectorial del problema de mínimos cuadrados", que encuentra un vector $\vec{u}$ tal que la forma euclideana $|y - Au|$ es mínima.

Es importante mencionar que en $\mathbb{R}^2$, la magnitud, longitud, módulo o norma de un $\vec{\mu}$ es $|(x,y)|_{\text{(Nomenclatura Física)}}$ o $\|(x,y)\|_{\text{(Nomenclatura Matemática)}} = \sqrt{x^2 + y^2}$, mientras que en $\mathbb{R}^3$, la magnitud, longitud, módulo o norma de un $\vec{\mu}$ es $|(x,y,z)|_{\text{(Nomenclatura Física)}}$ o $\|(x,y)\|_{\text{(Nomenclatura Matemática)}} = \sqrt{x^2 + y^2 + z^2}$, etcétera.

Entonces, minimizar $|y - Au|$ equivale a minimizar la suma de cuadrados en $[y_1 - (b + mx_1)]^2 + [y_2 - (b + mx_2)]^2 + [y_3 - (b + mx_3)]^2 + [y_4 - (b + mx_4)]^2 + \cdots + [y_n - (b + mx_n)]^2$.

Encontrar el vector $\vec{u}$ que minimiza no es tan difícil como parece. Como A es una matriz de $n_{(\text{Número de filas})} * 2_{(\text{Número de columnas})}$ y u una matriz de $2_{(\text{Número de filas})} * 1_{(\text{Número de columnas})}$, Au es un vector en $\mathbb{R}^n$ perteneciente a la imagenA, entendida como un subespacio de $\mathbb{R}^n$ cuya dimensión es a lo sumo dos, pues cuando mucho dos columnas de A son linealmente independientes.

2.1.1. Teorema de Aproximación

Por Teorema de Aproximación de Norma $\mathbb{R}^n$, "sea H un subespacio de $\mathbb{R}^n$, υ un vector en $\mathbb{R}^n$. Entonces $\text{proy}_{H\ (\text{Imagen de A})}\upsilon$ es la mejor aproximación para υ en H tal que si h es cualquier otro vector en H implica que $|\upsilon - \text{proy}_{H\ (\text{Imagen de A})}\upsilon| < |\upsilon - h|$", Teorema de Proyección, "sea H un subespacio de $\mathbb{R}^n$ y $\upsilon \in \mathbb{R}^n$.

Entonces existe un par único de vectores h y p tales que $h \in H$, $p \in H$ y $\upsilon = h + p$. En particular, $h = \text{proy}_{H\ (\text{Imagen de A})}\upsilon$ y $p = \text{proy}_{H\ (\text{Imagen de A})\perp}\upsilon$ tal que $\upsilon = h + p = \text{proy}_{H\ (\text{Imagen de A})}\upsilon + \text{proy}_{H\ (\text{Imagen de A})\perp}\upsilon$.

Su **demostración** implica que sea $h = \text{proy}_{H\ (\text{Imagen de A})}\upsilon$ y $p = \upsilon - h$ y por **Definición Proyección**, "sea H un subespacio de $\mathbb{R}^n$ con base ortonormal $\{u_1, u_2, u_3, u_4, \ldots, u_k\}$. Si $\upsilon \in \mathbb{R}^n$, entonces la proyección ortogonal de υ sobre H, denotada por $\text{proy}_{H\ (\text{Imagen de A})}\upsilon$ está dada por $\text{proy}_{H\ (\text{Imagen de A})}\upsilon = (\upsilon * u_1)u_1 + (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k$, siendo que $\text{proy}_{H\ (\text{Imagen de A})}\upsilon \in H$", se tiene que $h \in H$.

Demostrando que $p \in H^{\perp}$, sea $\{u_1, u_2, u_3, u_4, \dots, u_k\}$ una base ortonormal para H: $h = (\upsilon * u_1)u_1 = (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k$ y sea x un vector H tal que existen constantes $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_k$ tales que $x = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \cdots + \alpha_k u_k$ entonces $p * x = (\upsilon - h) * x = [\upsilon - (\upsilon * u_1)u_1 - (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k] * [\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \cdots + \alpha_k u_k]$ aunque como $\mathbf{u_i} * \mathbf{u_j} = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$ es sencillo verificar que el producto escalar $p * x = (\upsilon - h) * x = [\upsilon - (\upsilon * u_1)u_1 - (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k] * [\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \cdots + \alpha_k u_k]$ está dado por $\mathbf{p} * \mathbf{x} = \sum_{i=1}^{k} \alpha_i(\upsilon * u_i) - \sum_{i=1}^{k} \alpha_i(\upsilon * u_i) = 0$, así $\mathbf{p} * \mathbf{x} = 0 \; \forall \; x \in H$ indicando que $p \in H^{\perp}$.

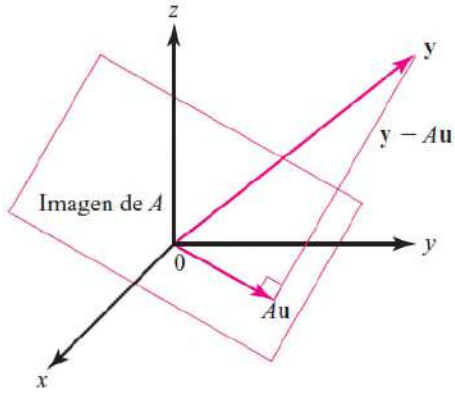
Para demostrar que $p = \text{proy}_{H\ (\text{Imagen de A})^{\perp}}\upsilon$ se amplia $\{u_1, u_2, u_3, u_4, \dots, u_k\}$ a base ortonormal (base ortogonal en que la norma de cada elemento componente es unitaria o tiene vectores unitarios) en $\mathbb{R}^n$: $\{u_1, u_2, u_3, u_4, \dots, u_k, \upsilon_{k+1}, \dots, u_n\}$.

Entonces, $\{\upsilon_{k+1}, \dots, u_n\}$ es una base para $H^{\perp}$", tal que se puede decir que $\upsilon - \text{proy}_{H\ (\text{Imagen de A})}\upsilon \in H$ se escribe $\upsilon - h = \left(\upsilon - \text{proy}_{H\ (\text{Imagen de A})}\upsilon\right) + \left(\text{proy}_{H\ (\text{Imagen de A})}\upsilon - h\right)$ siendo que el primer término de la derecha como el segundo está en H tal que $\left(\upsilon - \text{proy}_{H\ (\text{Imagen de A})}\upsilon\right) * \left(\text{proy}_{H\ (\text{Imagen de A})}\upsilon - h\right) = 0$ y por **Teorema** que afirma "sea $B = \{u_1, u_2, u_3, u_4, \dots, u_n\}$ una base ortonormal para $\mathbb{R}^n$ y sea $\upsilon \in \mathbb{R}^n$ tal que $\upsilon = (\upsilon * u_1)u_1 + (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k$ o $\upsilon = \text{proy}_{\mathbb{R}^n}y$, siendo su **demostración** que considera B como una base, puede escribir υ de manera única como $\upsilon = c_1u_1 + c_2u_2 + c_3u_3 + c_4u_4 + \cdots + c_nu_n$ por lo que $\upsilon * u_i = c_1(u_1 * u_i) = c_2(u_2 * u_i) + c_3(u_3 * u_i) + c_4(u_4 * u_i) + \cdots + c_i(u_i * u_i) + \cdots + c_n(u_n * u_i) = c_i$, pues los vectores $\vec{u}_i$ son ortonormales y como esto se cumple para $i = 1, 2, 3, 4, \dots, n$ la demostración queda completa, se puede decir que $\upsilon = (\upsilon * u_1)u_1 + (\upsilon * u_2)u_2 + (\upsilon * u_3)u_3 + (\upsilon * u_4)u_4 + \cdots + (\upsilon * u_k)u_k + (\upsilon *$

$u_{k+1})u_{k+1} + \cdots + (\upsilon * u_n)u_n = proy_H\,\upsilon + proy_{H^\perp}\,\upsilon$ aunque para probar la unicidad, suponga que $\upsilon = h_1 - p_1 = h_2 - p_2$ donde $h_1, h_2 \in H$ y $p_1, p_2 \in H^\perp$ entonces $h_1 - h_2 = p_1 - p_2$ aunque $h_1 - h_2 \in H$ y $p_1 - p_2 \in H^\perp$ de manera que $h_1 - h_2 \in H \in \cap\ H^\perp = \{0\}$, así que $h_1 - h_2 = 0$ y $p_1 - p_2 = 0$ completando la prueba", por lo que ", tal que $|y - Au|$ es un mínimo cuadrado pues $Au = proy_{H\ (\text{Imagen de A})}y$ donde H es la imagen de A.

En $\mathbb{R}^3$ la imagen de A será un plano o recta que pasa por el origen, pues son los únicos subespacios de dimensión uno o dos. En la siguiente figura, el vector que minimiza se denota por u:

Figura 2.3. Representación gráfica en $\mathbb{R}^3$



Fuente: (Mora, 2016)

Con base en esta figura y el Teorema de Pitágoras, "en todo triángulo rectángulo, el cuadrado de la longitud de la hipotenusa es igual a la suma de los cuadrados de las respectivas longitudes de los catetos", se deduce que $|y - Au|$ es mínima cuando $y - Au$ es ortogonal a imagen A. Es decir, si $\vec{u}$ es vector que minimiza implica que para todo vector $u \in \mathbb{R}^2$ por lo que $Au \perp (y - A\vec{u})$.

Usando la definición de producto escalar en $\mathbb{R}^n$, se encuentra que $Au \perp (y - A\vec{u})$ se vuelve $Au * (y - A\vec{u}) = 0$, $(Au)^\intercal * (y - A\vec{u}) = 0$ debido a que $a * b = a^\intercal b$, $(A^\intercal u^\intercal) * (y - A\vec{u}) = 0$ pues por **Teorema ii)** que supone en i) $(A^\intercal)^\intercal$,

ii) $(\mathbf{AB})^{\intercal} = \mathbf{A}^{\intercal}\mathbf{B}^{\intercal}$, iii) Si A y B son de $n * m$, entonces $(AB)^{\intercal} = A^{\intercal}+B^{\intercal}$ y iv) Si A es invertible, entonces $A^{\intercal}$ es invertible y $(A^{\intercal})^{-1} = (A^{-1})^{\intercal}$ o $a^{\intercal}(A^{\intercal}y - A^{\intercal} * A\vec{u}) = 0$ se cumple $\forall\, u \in \mathbb{R}^2 \iff A^{\intercal}y - A^{\intercal} * A\vec{u} = 0$, pero al despejar $\vec{u}$ de $A^{\intercal}y - A^{\intercal} * A\vec{u} = 0$ se obtiene que $y - Au \perp Au$.

Entonces, la solución al problema de mínimos cuadrados para ajuste por línea recta se da si A y y son:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdots \\ y_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix} \text{ y } u = \begin{pmatrix} b \\ m \end{pmatrix}$$

Entonces, la recta $y = a + bx$ o $y = b + mx$ da el mejor ajuste en sentido de mínimos cuadrados para puntos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \ldots, (x_n, y_n)$ cuando:

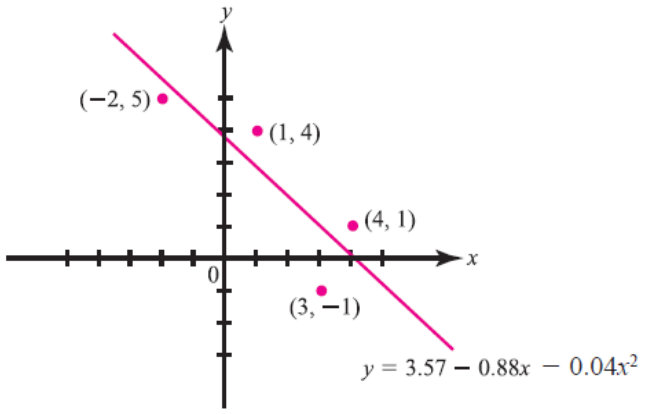
$$\begin{pmatrix} b \\ m \end{pmatrix} = \vec{u} = (A^{\intercal}A)^{-1}A^{\intercal}y \; y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdots \\ y_n \end{pmatrix}$$

Se ha puesto que $A^{\intercal}A$ (matriz A transpuesta por matriz A) es invertible. Éste siempre es el caso si nos n datos no son colineales.

2.2. Cuadrática

Ahora se desea ajustar una curva cuadrática a n datos. Una curva cuadrática en x es cualquier expresión de forma $y = a + bx + cx^2$ o $y = b + mx + cx^2$ tal que es la ecuación de una parábola en el plano. Si los n datos estuvieran sobre la parábola se tendría:

Figura 2.4. Representación gráfica de la función cuadrática



Fuente: (Mora, 2016)

$$y_1 = a + bx_1 + cx_1^2 \qquad y_1 = b + mx + cx_1^2$$

$$y_2 = a + bx_1 + cx_2^2 \qquad y_2 = b + mx + cx_2^2$$

$$y_3 = a + bx_1 + cx_3^2 \qquad y_3 = b + mx + cx_3^2$$

ó

$$y_4 = a + bx_1 + cx_4^2 \qquad y_4 = b + mx + cx_4^2$$

$$\ldots \qquad \ldots$$

$$y_n = a + bx_n + cx_n^2 \qquad y_n = b + mx + cx_n^2$$

Aunque, este sistema se puede escribir como $\mathbf{y} - \mathbf{Au}$ con:

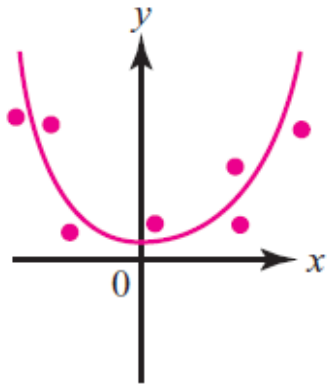
$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdots \\ y_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ \cdots & \cdots & \cdots \\ 1 & x_n & x_n^2 \end{pmatrix} \text{ y } u = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Al igual que antes, si todos los datos no se encuentran sobre la misma parábola implica que $y - Au \neq 0$ para cualquier vector $\vec{u}$ y, de nuevo, el problema es

"encontrar un vector u en $\mathbb{R}^2$ tal que $|y - Au|$ es mínima". Usando un razonamiento similar a éste, se puede demostrar que cuando menos tres de las x_i son diferentes, entonces A^TA es invertible y el vector que minimiza al vector $\vec{u}$ está dado por $\vec{\mathbf{u}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$.

La siguiente imagen muestra la figura de un modelo matemático-econométrico cuadrático:

Figura 2.5. Modelo matemático-econométrico cuadrático



Fuente: (Mora, 2016)

Teorema: Sea $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$ puntos en $\mathbb{R}^2$, suponga que no todas las x_i son iguales. Entonces, si A está dada como la siguiente forma es matriz A^TA es invertible de $2 * 2$:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdots \\ y_n \end{pmatrix}, A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix} y\ u = \begin{pmatrix} b \\ m \end{pmatrix}$$

Es importante mencionar que si $x_1 = x_2 = x_3 = x_4 = \cdots = x_n$, entonces todos los datos sobre la recta vertical $x = x_1$ y la mejor aproximación lineal es dicha recta.

Su **demostración** es que se tiene:

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix}$$

Como no todas las x_i son iguales, las columnas de A son linealmente independientes. Ahora bien:

$$A^{\intercal}_{(\text{Matriz A transpuesta})}A = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

Si $A^{\intercal}_{(\text{Matriz A transpuesta})}A$ no es invertible, entonces $A^{\intercal}A = 0$. Esto significa que:

$$n\sum_{i=1}^{n} x_i^2 = \left(\sum_{i=1}^{n} x_i\right)^2$$

Sea $u = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix}$ y $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \cdots \\ x_n \end{pmatrix}$. En consecuencia, $|u|^2 = u * u = n, |x|^2 = \sum_{i=1}^{n} x_i^2$ y $u * x = \sum_{i=1}^{n} x_i$. Tal que la ecuación $n\sum_{i=1}^{n} x_i^2 = (\sum_{i=1}^{n} x_i)^2$ puede establecerse como $|u||x|^2 = |u * x|^2$, sacando raíz cuadrada se puede obtener $|u * x| = |u||x|$.

La desigualdad de Cauchy-Schwarz en $\mathbb{R}^n$ sostiene que si u y υ son dos vectores en $\mathbb{R}^n$:

i. $|u * v| \leq |u||x|$. Su **demostración** es si $u = 0$, $v = 0$ o ambos, entonces $|u * v| \leq |u||x|$ se cumple, pues ambos lados son iguales a 0. Si se supone que $u \neq 0$ y $v \neq 0$, entonces:

$$0 \leq \left|\frac{u}{|u|} - \frac{v}{|v|}\right|^2 = \left(\frac{u}{|u|} - \frac{v}{|v|}\right) * \left(\frac{u}{|u|} - \frac{v}{|v|}\right) = \frac{u * u}{|u|^2} - \frac{2u * v}{|u||v|} + \frac{v * v}{|v|^2}$$

$$= \frac{|u|^2}{|u|^2} - \frac{2u * v}{|u||v|} + \frac{|v|^2}{|v|^2} = 1 - \frac{2u * v}{|u||v|} + 1 = 2 - \frac{2u * v}{|u||v|} \Rightarrow 2$$

$$\leq \frac{2u * v}{|u||v|}, \frac{u * v}{|u||v|} \leq 1 \text{ y } u * v \leq |u||v|$$

De manera semejante con:

$$0 \leq \left|\frac{u}{|u|} - \frac{v}{|v|}\right|^2 \Rightarrow \frac{u * v}{|u||v|} \geq -1 \text{ ó } u * v \geq |u||v|$$

Tal que con estas dos desigualdades se obtiene:

$$-|u||v| \leq u * v \leq |u||v| \text{ ó } |u * v| \leq |u||v|$$

ii. $|u * v| = |u||x| \Leftrightarrow u = 0$ o $v = \lambda u$ para algún número $\mathbb{R}\,\lambda$. Su **demostración** indica que si $u = \lambda v$, entonces $|u * v| = |\lambda v * v| = |\lambda||v|^2$ y $|u||v| = |\lambda v||v| = |\lambda||v||v| = |\lambda||v|^2 = |u * v|$. Inversamente, suponga que $|u * v| = |u||v|$ con $u \neq 0$ y $v \neq 0$, entonces:

$$\left|\frac{u * v}{|u||v|}\right| = 1 \Rightarrow \frac{u * v}{|u||v|} = \pm 1$$

Entonces:

- ✓ **Caso 1**: $\frac{u*v}{|u||v|} = 1$.

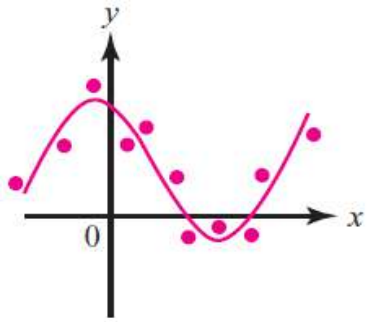
$$\left|\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right|^2=\left(\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right)*\left(\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right)=\frac{u*u}{|u|^2}-\frac{2u*\upsilon}{|u||\upsilon|}+\frac{\upsilon*\upsilon}{|\upsilon|^2}$$
$$=\frac{|u|^2}{|u|^2}-\frac{2u*\upsilon}{|u||\upsilon|}+\frac{|\upsilon|^2}{|\upsilon|^2}=1-\frac{2u*\upsilon}{|u||\upsilon|}+1=2-\frac{2u*\upsilon}{|u||\upsilon|}$$
$$=2-2(1)=0\Rightarrow\frac{u}{|u|}=\frac{\upsilon}{|\upsilon|}\text{ ó }u=\frac{u}{|u|}*\upsilon=\lambda\upsilon$$

✓ **Caso 2**: $\frac{u*\upsilon}{|u||\upsilon|}=-1$.

$$\left|\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right|^2=\left(\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right)*\left(\frac{u}{|u|}-\frac{\upsilon}{|\upsilon|}\right)=\frac{u*u}{|u|^2}+\frac{2u*\upsilon}{|u||\upsilon|}+\frac{\upsilon*\upsilon}{|\upsilon|^2}$$
$$=\frac{|u|^2}{|u|^2}+\frac{2u*\upsilon}{|u||\upsilon|}+\frac{|\upsilon|^2}{|\upsilon|^2}=1+\frac{2u*\upsilon}{|u||\upsilon|}+1=2+\frac{2u*\upsilon}{|u||\upsilon|}$$
$$=2-2(1)=0\Rightarrow\frac{u}{|u|}=-\frac{\upsilon}{|\upsilon|}\text{ ó }u=-\frac{u}{|u|}*\upsilon=\lambda\upsilon$$

2.3. Cúbica

Figura 2.6. Modelo matemático-econométrico cúbico



Fuente: (Mora, 2016)

El objetivo es encontrar la curva del tipo específico que se ajuste "mejor" a los datos dados. Complementariamente, un sistema de ecuaciones se escribe $A\mathbf{x}=\mathbf{b}$ tal que A es una matriz de mxn, $\mathbf{X}\in\mathbb{R}^n$ y $\mathbf{b}\in\mathbb{R}^n$. Asimismo, la función $A\mathbf{x}=\mathbf{b}$

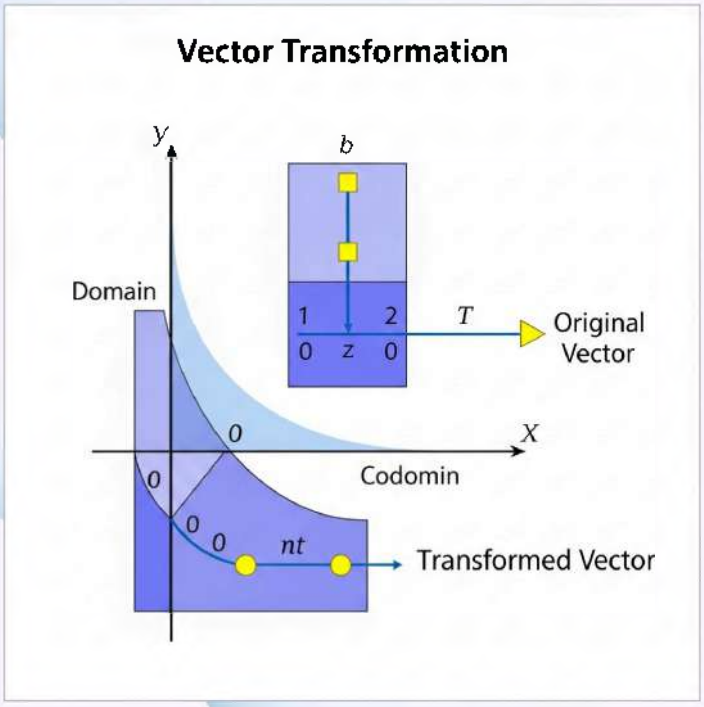
tiene las propiedades que caracteriza las transformaciones lineales, $A(\alpha\mathbf{x}) = \alpha A\mathbf{x}$ si α es una escalar y $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.

Entonces, la definición de Transformación Lineal es que sean V y W espacios vectoriales reales tal que una transformación lineal T de V en W es una función que asigna a cada vector $v \in V$ un vector único $T\mathbf{v} \in W$ y que satisface, para cada $\mathbf{u}$ y $\mathbf{v}$ en V y cada escalar α, $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}$ y $T(\alpha\mathbf{v}) = \alpha T\mathbf{v}$.

EDITORIAL ANDES COGNITIO

CAPÍTULO III

TRANSFORMACIONES Y ESTRUCTURAS VECTORIALES EN LA OPTIMIZACIÓN MATEMÁTICA



CAPÍTULO III

TRANSFORMACIONES Y ESTRUCTURAS VECTORIALES EN LA OPTIMIZACIÓN MATEMÁTICA

3. Transformación

Es importante mencionar que las definiciones y teoremas se cumplen también para espacios vectoriales complejos (espacios vectoriales donde escalares son números complejos); aunque, sólo se manejan espacios vectoriales reales y, por ende, se eliminará la palabra "real" en análisis de espacios vectoriales y transformaciones lineales, conocidas como operadores lineales.

Ejemplos:

3.1. Transformación lineal de $\mathbb{R}^2$ en $\mathbb{R}^3$

Sea T: $\mathbb{R}^2 \to \mathbb{R}^3$ definida por $T\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X+Y \\ X-Y \\ 3Y \end{pmatrix}$. Por ejemplo: $T\begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \\ -9 \end{pmatrix} \Rightarrow$

$$T\left[\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right] = T\begin{pmatrix} X_1+X_2 \\ Y_1+Y_2 \end{pmatrix} = \begin{pmatrix} X_1+X_2+Y_1+Y_2 \\ X_1+X_2-Y_1-Y_2 \\ 3Y_1+3Y_2 \end{pmatrix} = \begin{pmatrix} X_1+Y_1 \\ X_1-Y_1 \\ 3Y_1 \end{pmatrix} +$$

$$\begin{pmatrix} X_2+Y_2 \\ X_2-Y_2 \\ 3Y_2 \end{pmatrix} \text{ Pero } \begin{pmatrix} X_1+Y_1 \\ X_1-Y_1 \\ 3Y_1 \end{pmatrix} = T\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \text{ y } \begin{pmatrix} X_2+Y_2 \\ X_2-Y_2 \\ 3Y_2 \end{pmatrix} = T\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} \text{ Así, } T\left[\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} +\right.$$

$$\left.\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}\right] = T\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + T\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} \approx T\left[\alpha\begin{pmatrix} X \\ Y \end{pmatrix}\right] = T\begin{pmatrix} \alpha X \\ \alpha Y \end{pmatrix} = \begin{pmatrix} \alpha X+\alpha Y \\ \alpha X-\alpha Y \\ 3\alpha Y \end{pmatrix} =$$

$$\alpha\begin{pmatrix} X+Y \\ X-Y \\ 3Y \end{pmatrix} = \alpha T\begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \therefore \textbf{ T es una transformación lineal.}$$

3.2. Transformación cero

Sean V y W espacios vectoriales y definida $T: V \to W$ por $T\mathbf{v} = \mathbf{0}$ para $\forall\ \mathbf{v}$ en $V \Rightarrow T(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{0} + \mathbf{0} = T\mathbf{v}_1 + T\mathbf{v}_2$ y $T(\alpha\mathbf{v}) = \mathbf{0} = \alpha\mathbf{0} = \alpha T\mathbf{v}$. En este caso, T se denomina **transformación cero**.

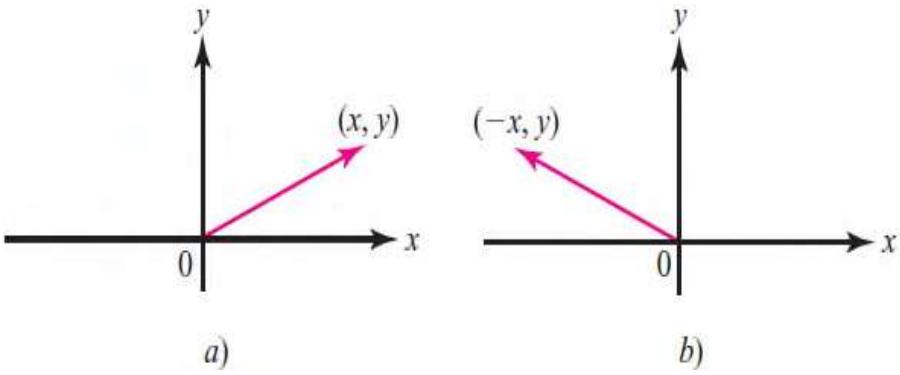
3.3. Transformación identidad

Sea V un espacio vectorial y definida $I: V \to V$ por $I\mathbf{v} = \mathbf{v}$ para $\forall\ \mathbf{v}$ en V. ES obvio que I es una transformación lineal, llamada transformación identidad u operador identidad.

3.4. Transformación de reflexión

Sea $T: \mathbb{R}^2 \to \mathbb{R}^3$ definida por $T\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} -X \\ Y \end{pmatrix}$. ES fácil verificar que T es lineal. En términos geométricos, T toma un vector en $\mathbb{R}^2$ y lo refleja respecto al eje Y (vector $(-X, Y)$ es reflexión respecto a eje Y del vector (X, Y)):

Figura 3.1. Transformación de reflexión



Fuente: (Mora, 2016)

3.5. Transformación lineal de $\mathbb{R}^n \to \mathbb{R}^m$ dada por multiplicación de matriz mxn

Sea A una matriz de mxn, definida $T: \mathbb{R}^n \to \mathbb{R}^m$ por $T\mathbf{x} = A\mathbf{x}$. Como $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$ y $A(\alpha\mathbf{x}) = \alpha A\mathbf{x}$ si $\mathbf{x}$ y están en $\mathbb{R}^n$, se observa que T es una transformación lineal. Entonces, toda matriz A de mxn se puede usar para definir una transformación de $\mathbb{R}^n$ en $\mathbb{R}^m$. Por lo tanto, toda transformación lineal entre espacios vectoriales de dimensión finita se puede representar por una matriz.

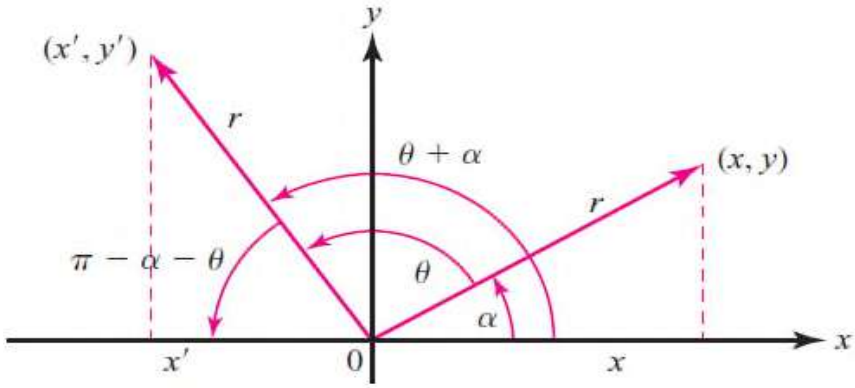
3.6. Transformación de rotación

Suponga que $\vec{\mathbf{v}} = \begin{pmatrix} X \\ Y \end{pmatrix}$ en $\Pi_{(XY)}$ se rota un $\measuredangle\, \theta$, medido en grados (^{0}) y radianes (rad) en sentido contrario a manecillas del reloj.

Llame a este nuevo vector rotado $\mathbf{v}' = \begin{pmatrix} X' \\ Y' \end{pmatrix}$ (si r denota longitud de $\mathbf{v}$, que no cambia por rotación tal que $x = r\,\mathrm{Cos}(\alpha) \Rightarrow x' = r\,\mathrm{Cos}(\theta + \alpha)$ y $y = r\,\mathrm{Sen}(\alpha) \Rightarrow y' = r\,\mathrm{Sen}(\theta + \alpha)$, pues la definición estándar de $\mathrm{Cos}(\theta)$ y $\mathrm{Sen}(\theta)$ como las coordenadas X y Y de un punto en círculo unitario.

Si (X, Y) es un punto en círculo de radio r con centro en origen, entonces $x = r\,\mathrm{Cos}(\varphi)$ y $y = r\,\mathrm{Sen}(\varphi)$, donde $\varphi_{(\text{Phi o Fi})}$ es ángulo que forma $\vec{\mathbf{v}}(X, Y)$ con lado positivo del eje X. Además, (X', Y') se obtiene rotando (X, Y) un ángulo θ):

Figura 3.2. Transformación de rotación



Fuente: (Mora, 2016)

Pero $\mathrm{r\,Cos}(\theta+\alpha) = \mathrm{r\,Cos}(\theta)\mathrm{Cos}(\alpha) - \mathrm{r\,Sen}(\theta)\mathrm{Sen}(\alpha) \Rightarrow \mathrm{x}' = \mathrm{x\,Cos}(\theta) - \mathrm{y\,Sen}(\theta)$. Paralelamente, $\mathrm{r\,Sen}(\theta+\alpha) = \mathrm{r\,Sen}(\theta)\mathrm{Cos}(\alpha) + \mathrm{r\,Cos}(\theta)\mathrm{Sen}(\alpha) \Rightarrow \mathrm{y}' = \mathrm{x\,Sen}(\theta) + \mathrm{y\,Cos}(\theta)$. Sea $\mathbf{A}_\theta = \begin{pmatrix} \mathrm{Cos}(\theta) & -\mathrm{Sen}(\theta) \\ \mathrm{Sen}(\theta) & \mathrm{Cos}(\theta) \end{pmatrix} \Rightarrow \mathrm{x}' = \mathrm{x\,Cos}(\theta) - \mathrm{y\,Sen}(\theta)$ y $\mathrm{y}' = \mathrm{x\,Sen}(\theta) + \mathrm{y\,Cos}(\theta)$ denota que $\mathbf{A}_\theta \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \end{pmatrix} = \begin{pmatrix} \mathrm{X}' \\ \mathrm{Y}' \end{pmatrix}$. La transformación lineal $\mathrm{T}: \mathbb{R}^2 \to \mathbb{R}^2$ definida por $\mathrm{T}\mathbf{v} = \mathbf{A}_\theta\mathbf{v}$, donde $\mathbf{A}_\theta$ está dado por $\mathbf{A}_\theta = \begin{pmatrix} \mathrm{Cos}(\theta) & -\mathrm{Sen}(\theta) \\ \mathrm{Sen}(\theta) & \mathrm{Cos}(\theta) \end{pmatrix}$ y recibe el nombre de transformación de rotación.

3.7. Transformación de proyección ortogonal

Sea H un subespacio de $\mathbb{R}^n$ tal que la transformación de proyección ortogonal está dada por $\mathrm{P}: \mathrm{V} \to \mathrm{H}$ se define por $\mathrm{P}\mathbf{v} = \mathrm{Proy}_\mathrm{H}\mathbf{v}$. Sea la sucesión $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \dots, \mathbf{u}_\mathrm{k}\}$ una base ortogonal para H.

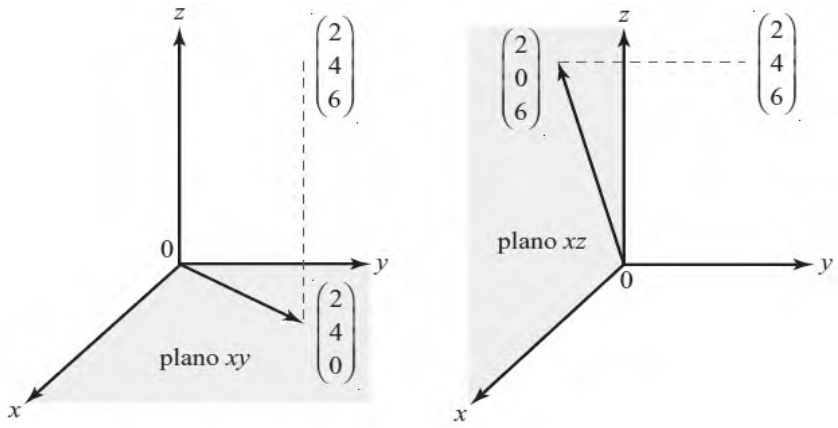
Entonces, con base en Definición de Proyección Ortogonal, que indica sea H un subespacio de $\mathbb{R}^n$ con base ortonormal $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \dots, \mathbf{u}_\mathrm{k}\}$ tal que si $\mathbf{v} \in \mathbb{R}^n \Rightarrow$ proyección ortogonal de $\mathbf{v}$ sobre H, denotada por $\mathrm{Proy}_\mathrm{H}\mathbf{v}$ está dada por $\mathrm{Proy}_\mathrm{H}\mathbf{v} = (\mathbf{v} * \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v} * \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v} * \mathbf{u}_3)\mathbf{u}_3 + (\mathbf{v} * \mathbf{u}_4)\mathbf{u}_4 + \cdots + (\mathbf{v} * \mathbf{u}_\mathrm{k})\mathbf{u}_\mathrm{k}$ tal que $\mathrm{Proy}_\mathrm{H}\mathbf{v} \in \mathrm{H}$, se tiene que $\mathrm{P}\mathbf{v} = (\mathbf{v} * \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v}_1 * \mathbf{u} * \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v} * \mathbf{u}_3)\mathbf{u}_3 +$

$(\mathbf{v} * \mathbf{u}_4)\mathbf{u}_4 + \cdots + (\mathbf{v} * \mathbf{u}_k)\mathbf{u}_k$, como $(\mathbf{v}_1 + \mathbf{v}_2) * \mathbf{u} = \mathbf{v}_1 * \mathbf{u} + \mathbf{v}_2 * \mathbf{u}$ y $(\alpha\mathbf{v}) * \mathbf{u} = \alpha(\mathbf{v} * \mathbf{u}) \Rightarrow$ P es una transformación lineal.

3.7.1. Dos operadores de proyección

Se define $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ por $T\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} \Rightarrow$ T es operador de proyección que toma un $\vec{\mu}$ de tres dimensiones y se proyecta en $\Pi_{(XY)}$. Análogamente, $T\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X \\ 0 \\ Z \end{pmatrix}$ proyecta un $\vec{v}$ en espacio en $\Pi_{(XZ)}$:

Figura 3.3. Transformación de dos operadores de proyección



Fuente: (Mora, 2016)

3.7.2. Operador de transposición

Se define $T: M_{mn} \rightarrow M_{nm}$ por $T(A) = A^T$. Tal que, $(A + B)^T = A^T + B^T$ y $(\alpha A)^T = \alpha A^T \Rightarrow$ T, denominado **operador de transposición** es una transformación lineal.

3.7.3. Operador integral

Sea $J: C[0,1] \to \mathbb{R}$ definida por $Jf: \int_0^1 f(x)\,dx$ tal que para $f, g \in C[0,1]$, como $\int_0^1 [f(x)+g(x)]\,dx = \int_0^1 f(x)\,dx + \int_0^1 g(x)\,dx$ y $\int_0^1 \alpha f(x)\,dx = \int_0^1 f(x)\,dx \Rightarrow J$ es un operador lineal. Por ejemplo: $J(X^3) = \frac{1}{4} \Rightarrow J$ se denomina operador integral.

3.7.4. Operador diferencial.

Suponga que $D: C^1[0,1] \to C[0,1]$ se define por $Df = f'$. Para $f, g \in C^1[0,1]$, como $(f+g)' = f' + g'$ y $(\alpha f')' = \alpha f'$ se denota que D es un operador lineal, por lo que D se denomina operador diferencial.

3.8. Transformación no lineal.

Suponga que $T: C[0,1] \to \mathbb{R}$ está definida por $Tf = f(0) + 1 \Rightarrow T$ no es lineal. Es decir: $T(f+g) = (f+g) + 1 = f(0) + g(0) + 1 \Rightarrow Tf + Tg = [f(0) + 1] + [g(0) + 1] = f(0) + g(0) + 2$.

Sin embargo, este ejemplo da una idea clara que una transformación puede parecer lineal, pero en realidad no es así. No toda transformación que parece lineal lo es. Por ejemplo: $T: \mathbb{R} \to \mathbb{R}$ por $Tx = 2x + 3 \Rightarrow$ la gráfica de $\{(x, Tx)\}: x \in \mathbb{R}$ es línea recta en en $\Pi_{(XY)}$, pero T no es lineal por que $T(x+y) = 2(x+y) + 3 = 2x + 2y + 3$ y $Tx + Ty = (2x+3) + (2y+3) = 2x + 2y + 6$.

Por lo tanto, las únicas transformaciones lineales de $\mathbb{R}$ en $\mathbb{R}$ son funciones de forma $f(x) = mx$ para algún $\mathbb{R}$ m y, en consecuencia, entre todas las funciones cuyas gráficas son rectas, las únicas que son transformaciones lineales son aquellas que pasan por el origen.

En álgebra y cálculo una función lineal con dominio $\mathbb{R}$ está definida como una función que tiene la forma $f(x) = mx + b$ tal que se puede decir que una función

lineal es una transformación de $\mathbb{R}$ en $\mathbb{R}$ si y sólo si $b = 0$, ordenada al origen vale 0.

3.9. Transformaciones de Box − Cox

Complementariamente, las transformaciones de Box − Cox son una familia de transformaciones potenciales usadas en estadística para corregir sesgos en distribución de errores, corregir varianzas desiguales (diferentes valores de la variable predictora) y, especialmente, corregir la no linealidad en la relación (mejorar correlación entre las variables). Esta transformación recibe el nombre de los estadísticos George Box y David Cox.

La transformación potencial está definida como una función continua que varía respecto a la potencia lambda (λ). Para datos $(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)$ se realiza la transformación $Y_i^{'} = Y_i^{\lambda}$ tal que $Y_i^{(\lambda)} = \begin{cases} K_1\left(Y_i^{\lambda} - 1\right) \text{ si } \lambda \neq 0, \\ K_2 \operatorname{Ln}(Y_i) \text{si } \lambda = 0 \end{cases}$ tal que K_2 es media geométrica de valores $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$; en consecuencia, $K_2 = \left(\prod_{i=1 \text{ (Producto desde 1 hasta n)}}^{n} Y_i\right)^{\frac{1}{n}} = (Y_1 * Y_2 * Y_3 * Y_4 *, \dots, * Y_n)^{\frac{1}{n}}$, mientras que $K_1 = \frac{1}{\lambda * K_2^{\lambda - 1}}$.Entonces, para realizar una transformación potencial, dado un valor de lambda λ, se calcula primero la media geométrica de valores $Y_1(K_2)$. Después se sustituye este valor para calcular el parámetro K_1. El procedimiento para seleccionar el mejor valor λ consiste en que primero se selecciona el rango de valores lambda λ, se selecciona el que logra que la transformación se acerque al máximo a los datos tal que para cada valor de λ se hace la transformación del paso anterior.

Por último, se sustituyen de la variable (s) explicativas en diferentes funciones y se calculan los cuadrados de residuales estadísticos. Aquella que tenga el menor valor de suma de residuales será la mejor opción tal que K_2 es valor fijo para todos los casos y que sólo hay que calcular de nuevo el valor K_1.

Según (Grossman & Flores, 2012),

$$V_{(\text{Espacio Vectorial})} = \mathbb{R}^n = \left\{ \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{pmatrix} : X_j \, \varepsilon \, \mathbb{R} \text{ para } j = 1, 2, 3, 4 \ldots n \right\}$$

tal que cada vector $\mathbb{R}^n$ es una matriz n*1, siendo $\mathbb{R}^n$ e símbolo que denota al conjunto de todos los vectores de dimensión n: $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{pmatrix}$ donde cada a_i es un número real $(\mathbb{R})$.

Según la definición de matrices, una matriz A de $m * n$ es un arreglo rectangular de mn números dispuestos en m renglones y n columnas tal que

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2j} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3j} & \cdots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4j} & \cdots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & a_{i4} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}, \; \mathbf{x} + \mathbf{y}$$ es una matriz de nX1 si **x** y **y** son matrices nx1. Haciendo $0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ y $-x = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{pmatrix}$, se observa que los axiomas ii) a x) se obtienen de la definición de suma de vectores (matrices) y el **Teorema** que indica que A, B y C son tres matrices de mxn y sean $\alpha - \beta$ dos escales implica los siguientes puntos:

i) $A + 0 = A$

ii) $0A = 0$

iii) $A + B = B + A$ (**ley conmutativa para suma de matrices**)

iv) $(A + B) + C = A + (B + C)$ (**ley asociativa para suma de matrices**)

v) $\alpha(A + B) = \alpha A + \alpha B$ (**ley distributiva para multiplicación por un escalar**)

vi) $1A = A$

vii) $(\alpha + \beta)A = \alpha A + \beta A$

Con base en esto, se puede decir que el conjunto de puntos $\mathbb{R}^n$ que se encuentran en una recta que pasa por el origen constituye un $V_{(\text{Espacio Vectorial})}$, pues sea $V_{(\text{Espacio Vectorial})} = \{(x, y)\}: y = mx$ donde m es un número fijo real y x número real arbitrario. Es decir, V consiste en todos los puntos que están sobre la recta $y = mx$ que pasa por el origen y tiene pendiente m. Su demostración consiste en que un espacio vectorial real $(V +, *)$ es un conjunto de objetos, denominados vectores, junto con dos operaciones binarias llamadas suma (Ley de Composición Interna en V) y multiplicación (Ley de Composición Externa en V) por un escalar, que satisfacen 8 axiomas:

$$\left.\begin{array}{c} 1.\ \vec{\mu} + \vec{\upsilon} = \vec{\upsilon} + \vec{\mu} \\ 2.\ \vec{\mu} + (\vec{\upsilon} + \vec{w}) = (\vec{\mu} + \vec{\upsilon}) + \vec{w} \\ 3.\ \vec{w} + \vec{0} = \vec{w} \\ 4.\ \vec{\mu} + (-\vec{\mu}) = \vec{0} \end{array}\right\} \quad \forall\, \vec{\mu}, \vec{\upsilon}, \vec{w} \in \mathbb{R} \ \text{ y}$$

$$\left.\begin{array}{c} 1.\ \alpha\,(\vec{\mu} + \vec{\upsilon}) = \alpha\vec{\mu} + \alpha\vec{\upsilon} \\ 2.\ (\alpha + \beta)\vec{\mu} = \alpha\vec{\mu} + \beta\vec{\mu} \\ 3.\ (\alpha\beta)\vec{\mu} = \alpha(\beta\vec{\mu}) \\ 4.\ 1 * \vec{\mu} = \vec{\mu} \end{array}\right\} \quad \begin{array}{l} \forall\, \alpha, \beta \in \vec{\mu} \\ \forall\, \vec{\mu}, \vec{\upsilon} \in \mathbb{R}^2 \end{array}$$

3.10. Espacio vectorial

Se llama espacio vectorial V sobre $\mathbb{R}$ a la terna $(V+,*) \Rightarrow V = \{\vec{\mu}, \vec{\upsilon}, \vec{w}, \dots\}$ donde sus elementos son vectores y se cumplen los 8 axiomas anteriores. Entonces, unsub espacio vectorial $(H+,*) \subset (V+,*)$ tal que H es un bus espacio vectorial de V si H es un subconjunto no vacio de V y, a la vez, H es un espacio vectorial, junto con las operaciones de suma entre vectores y multiplicación por un escalar definidas para ∇.

Por teorema, un sub conjunto no vacío H de un espacio vectorial V es un sub espacio de V si se cumplen las dos reglas de cerradura: I) Si $\vec{\mu} \in H$ y $\vec{\upsilon} \in H \Rightarrow \vec{\mu} + \vec{\upsilon} \in H$ y II) Si $\vec{\mu} \in H$ y $\alpha\vec{\mu} \in H \Rightarrow \forall\, \alpha_{(\text{Escalar})} \in \mathbb{R}$ tal que la caracterización de un sub espacio vectorial está dado por que $H \subset V \Leftrightarrow \left.\begin{matrix}\vec{\mu}, \vec{\upsilon} \in H \\ \alpha, \beta \in \mathbb{R}\end{matrix}\right\} \Rightarrow (\alpha\vec{\mu} + \beta\vec{\upsilon}) \in H$; por lo que, un Subespacio Trivial es $\{\vec{0}\} \Rightarrow \vec{0} = (0,0,0,0,\dots,0)$ y un Subespacio de $\mathbb{R}^2$ es $H = \{(X,Y)|y = mx\}$ que es un conjunto de puntos que pasan por el origen $(0,0)$.

3.10.1. Combinación lineal

La combinación lineal se demuestra a partir que sean $\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_n$ vectores en un espacio vectorial V tal que cualquier vector de la forma $\alpha_1\upsilon_1 + \alpha_2\upsilon_2 + \alpha_3\upsilon_3 + \alpha_4\upsilon_4 \dots + \alpha_n\upsilon_n$ se denomina *Combinación Lineal* de $\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_n$, donde $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n$ son escalares. Por último, un conjunto generador se representa a partir que vectores $\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_n$ de un espacio vectorial V genera a V si todo vector en V se puede escribir como una combinación lineal de los mismos; es decir, $\forall\, \upsilon \in V_{(\text{Espacio Vectorial})}$ ε escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n \Rightarrow \upsilon = \alpha_1\upsilon_1 + \alpha_2\upsilon_2 + \alpha_3\upsilon_3 + \alpha_4\upsilon_4 \dots + \alpha_n\upsilon_n$.

Además, un espacio generado por un conjunto de vectores se genera a partir que sea $\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_k$ k vectores de un espacio vectorial V. El espacio generado

por $\{\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_k\}$ es el conjunto de combinaciones lineales $\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_k$; es decir, $\text{gen}\{\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \dots, \upsilon_k\} = \{\upsilon | \upsilon = \alpha_1\upsilon_1 + \alpha_2\upsilon_2 + \alpha_3\upsilon_3 + \alpha_4\upsilon_4 \dots + \alpha_k\upsilon_k\}$.

Para demostrar que V es un espacio vectorial se puede verificar que se cumple cada uno de los siguientes Axiomas de un Espacio Vectorial son, siendo los primero cinco para definir a un grupo abeliano, mientras que axiomas vi) a x) describen interacción de escalares y vectores mediante operación binaria de un escalar y un vector:

i) Si $X \varepsilon V$ y $Y \varepsilon V$, entonces $X + Y \varepsilon V$ (**cerradura bajo suma)**

ii) Para $\forall x, y$ y z en V, $(x + y) + z = x + (y + z)$ (**ley asociativa de suma de vectores**)

iii) $\in$ un vector $0 \varepsilon V \Rightarrow \forall X \varepsilon V, X + 0 = 0 + X = X$ (**0 es llamado vector cero o idéntico aditivo**)

iv) Si $X \varepsilon V$, existe un vector $-X$ en εV tal que $X + (-X) = 0$ (**$-$x se llama inverso aditivo de X**)

v) Si **X** y **Y** están en V, entonces $\mathbf{X + Y = Y + X}$ (**Ley conmutativa de suma de vectores**)

vi) Si $X \varepsilon V$ y α es una escalar, entonces $\alpha X \varepsilon V$ (**cerradura bajo multiplicación por un escalar**)

vii) Si **X** y **Y** están en **V** y α es una escalar, entonces $\alpha(\mathbf{X + Y}) = \alpha\mathbf{X} + \alpha\mathbf{Y}$ (**primera ley distributiva**)

viii) Si $X \varepsilon V$ y $\alpha - \beta$ son escalares, entonces $(\alpha + \beta)\mathbf{X} = \alpha\mathbf{X} + \beta\mathbf{X}$ (**segunda ley distributiva**)

ix) Si $X \varepsilon V$ y $\alpha - \beta$ son escalares, entonces $\alpha(\beta\mathbf{X}) = (\alpha\beta)\mathbf{X}$ (**ley asociativa de multiplicación por escalares**)

x) Para $\forall$ vector $X \varepsilon V$, $1\mathbf{X} = \mathbf{X}$

Es importante observar que vectores en $\mathbb{R}^2$ se escriben como renglones en lugar de columnas, pues esencialmente es equivalente. **Demostración**:

i) $\mathbf{X} = (X_1, Y_1)$ y $\mathbf{Y} = (X_2, Y_2)$ están en $V \Rightarrow Y_1 = mX_1, Y_2 = mX_2$ y $\mathbf{X} + \mathbf{Y} = (X_1, Y_1) + (X_2, Y_2) = (X_1, mX_1) + (X_2, mX_2) = (X_1 + X_2, mX_1 + mX_2) = \big(X_1 + X_2, m(X_1 + X_2)\big)\, \varepsilon\, V$ ∴ se cumple este axioma.

ii) $(X, Y)\, \varepsilon\, V \Rightarrow Y = mX - (X, Y) = -(X, mX) = (-X, m(-X))$ tal que $-(X, Y)\, \varepsilon\, V$ y $(x, mX) + (-X, m(-X)) = (X - X, m(X - X)) = (0,0)$.

Por lo tanto, ∀ vector en V es vector en $\mathbb{R}^2$ y $\mathbb{R}^2$ es V, como se muestra enseguida: $V_{(\text{Espacio Vectorial})} = \mathbb{R}^n =$

$$\left\{ \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{pmatrix} : X_j\, \varepsilon\, \mathbb{R} \text{ para } j = 1, 2, 3, 4 \ldots n \right\}$$ tal que cada vector $\mathbb{R}^n$ es una matriz n*1, siendo $\mathbb{R}^n$ e símbolo que denota al conjunto de todos los vectores de dimensión n: $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{pmatrix}$ donde cada a_i es un número real $(\mathbb{R})$. Como $(0,0) = 0$ está en V por estar en el origen, todas las demás propiedades se deducen del ejemplo anterior.

Entonces, V es $V_{(\text{Espacio Vectorial})}$. Complementariamente, el conjunto de puntos en $\mathbb{R}^3$ que se encuentran en un plano (Π) que pasa por el origen constituye un $V_{(\text{Espacio Vectorial})}$, pues si $V = \{(X, Y, Z): aX + bY + cZ = 0\}$ debido a que V es

el conjunto de puntos en $\mathbb{R}^3$ que está en Π con vector $\mathbf{n}$ (a, b, c) que pasa por el origen $(0, 0, 0)$.

Demostración: (X_1, Y_1, Z_1) y (X_2, Y_2, Z_2) están en $V \Rightarrow (X_1, Y_1, Z_1) + (X_2, Y_2, Z_2) = (X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2)\, \varepsilon\, V$ debido a que $a(X_1 + X_2) + b(Y_1 + Y_2) + c(Z_1 + Z_2) \Rightarrow (aX_1 + bY_1 + cZ_1) + (aX_2 + bY_2 + cZ_2) = 0 + 0 = 0$ $\therefore$ El axioma se cumple. El resto de los axiomas mencionados se verifican fácilmente. Entonces, el conjunto de puntos que se ubican en un Π $\mathbb{R}^3$ que pasa por el origen constituye un $V_{(\text{Espacio Vectorial})}$.

Sin embargo, el conjunto de puntos en $\mathbb{R}^2$ que se ubican sobre una recta que no pasa por el origen no constituye un $V_{(\text{Espacio Vectorial})}$ pues $V = \{(X, Y): Y = 2X + 1, X\, \varepsilon\, V\}$, en otras palabras V es el conjunto de puntos que están sobre la recta $Y = 2X + 1$, pero V no es un espacio vectorial debido a que no se cumple la cerradura bajo la suma (Si $X\, \varepsilon\, V$ y $Y\, \varepsilon\, V$, entonces $X + Y\, \varepsilon\, V$).

Por ejemplo:

Sea $V = \{1\}$, tal que V consiste sólo en el número 1, por lo que no es un espacio vectorial debido a que viola el axioma de vectores i) -axioma de cerradura-, inclusive lo hace con otros axiomas aunque es suficiente con demostrar que lo hace al menos con uno de los diez mencionados y, en consecuencia, queda probado que V no es un espacio vectorial e, incluso, basta con observar que $1 + 1 = 2 \notin V$.

Demostración: Si (X_1, Y_1) y (X_2, Y_2) están en V, entonces $(X_1, Y_1) + (X_2, Y_2) = (X_1 + X_2, Y_1 + Y_2)$ tal que si el vector del lado derecho estuviera en V se tendría $Y_1 + Y_2 = 2(X_1 + X_2) + 1 = 2X_1 + 2X_2 + 1$, pero $Y_1 = 2X_1 + 1$ y $Y_1 = 2X_2 + 1$, de manera que $Y_1 + Y_2 = (2X_1 + 1) + (2X_2 + 1) = 2X_1 + 2X_2 + 2$. Por lo tanto, $(X_1 + X_2, Y_1 + Y_2) \notin V \Leftrightarrow (X_1, Y_1)\, \varepsilon\, V$ y $(X_2, Y_2)\, \varepsilon\, V$.

Por ejemplo:

$(0,1)$ y $(3,7)$ están en V, pero $(0,1)+(3,7)=(3,8)$ no está en V porque $8 \neq 2*(3)+1$ otra manera forma sencilla de comprobarlo es observar que $\mathbf{0}=(0,0)$ no se encuentra en V porque $0 \neq 2*(0)+1$. En conclusión, no es difícil demostrar que el conjunto de puntos en $\mathbb{R}^2$ ubicado sobre cualquier recta que no pasa por $(0,0)$ no constituyen un $V_{(\text{Espacio Vectorial})}$.

El Espacio Vectorial P_n existe tal que sea $V=P_n$ el conjunto de polinomios con coeficientes reales de grado $\leq n$. Si $p \in P_n$ entonces $p(X)=a_nX^n+a_{n-1}X^{n-1}+a_{n-2}X^{n-2}+a_{n-3}X^{n-3}+a_{n-4}X^{n-4}+\cdots+a_1X+a_0 \ \forall\, a_i \in \mathbb{R}$.

La suma de $p(X)+q(X)$ está definida usualmente, si $q(X)=b_nX^n+b_{n-1}X^{n-1}+b_{n-2}X^{n-2}+b_{n-3}X^{n-3}+b_{n-4}X^{n-4}+\cdots+b_1X+b_0 \ \forall\, b_j \in \mathbb{R} \Rightarrow\therefore$ $p(X)+q(X)=(a_n+b_n)X^n+(a_{n-1}+b_{n-1})X^{n-1}+(a_{n-2}+b_{n-2})X^{n-2}+(a_{n-3}+b_{n-3})X^{n-3}+(a_{n-4}+b_{n-4})X^{n-4}+\cdots+(a_1+b_1)X+(a_0+b_0)$ siendo obvio que la suma de dos polinomios de grado $\leq n$ es otro polinomio de $\leq n$, por lo tanto se cumple el axioma i). Las propiedades ii) y v) son claras. Si se define el polinomio $\mathbf{0}=0X^n+0X^{n-1}+0X^{n-2}+0X^{n-3}+0X^{n-4}+\cdots+0X+0 \Rightarrow 0 \in P_n$ y, con ello, el axioma iii) se cumple, con lo que P_n es un $V_{(\text{Espacio Vectorial})} \rightarrow \mathbb{R}$ o un espacio vectorial real.

Además, las funciones constantes, incluyendo función $f(X)=0$, son polinomios de grado cero.

Los Espacios Vectoriales $C[0,1]$ y $C[a,b]$ existen si $V=C[0,1]$ el conjunto de funciones continuas de valores reales definidas en intervalo $[0,1]$. Se define $(f+g)X=f(X)+g(X)$ y $(\alpha f)(X)=\alpha[f(X)]$ como suma de funciones continuas es continua, por lo que el axioma i) se cumple y el resto de los axiomas son verificables fácilmente con $\mathbf{0}=$ función cero y $(-f)(X)=-f(X)$.

Igualmente, $C[0,1]$, el conjunto de funciones de valores reales definidas y continuas en $[a,b]$ constituye un $V_{(\text{Espacio Vectorial})}$.

El Espacio Vectorial M_{mn} existe si $V = M_{mn}$ denotada por conjunto de matrices mxn con componentes reales, entonces con suma de matrices y multiplicación por un escalar usuales se verifica que M_{mn} es $V_{(\text{Espacio Vectorial})}$ cuyo neutro aditivo es la matriz de dimensiones mxn. Sin embargo, un conjunto de matrices invertibles puede no formar un $V_{(\text{Espacio Vectorial})}$, pues si S_3 es conjunto de matrices invertibles de 3x3.

Se define la "suma" $A + B$ por $A + B = AB$, pues si A y B son invertibles, entonces AB es invertible por el Teorema que indica sean A y B son invertibles de mxn $\Rightarrow$ AB es invertible y $(AB)^{-1} = A^{-1}B^{-1}$.

Demostración:

Con base en **Teorema** que afirma que si una matriz A es invertible, entonces su inversa es única, pues si B y C son dos inversas de A se pude demostrar que $B = C$. Por definición se tiene que $AB = BA = I$ y $AC = CA = I$.

Por ley asociativa de multiplicación de matrices se tiene que $B(AC) = (AB)C \Rightarrow B = BI = (AB)C = IC = C \Rightarrow \therefore B = C$ quedando demostrado el teorema y, en consecuencia, $A^{-1}B^{-1} = (AB)^{-1} \Leftrightarrow A^{-1}B^{-1}(AB) = (AB)(A^{-1}B^{-1}) = I$ tratándose, únicamente, de una consecuencia, pues $(A^{-1}B^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}(I)B = B^{-1}B = I$ debido a $ABCD = A[B(CD)] = [(AB)C]D = A(BC)D = (AB)(CD)$ y, por lo tanto, $(AB)(A^{-1}B^{-1}) = A(B^{-1}B)A^{-1} = AIA^{-1} = AA^{-1} = I$ de manera que el axioma i) se cumple, mientras que axioma ii) es la ley asociativa para multiplicación de matrices, pues el Teorema de Ley Asociativa de Multiplicación de Matrices lo confirma al ser $A = (a_{ij})$ una matriz de mxn, $B = (b_{ij})$ matriz de mxp y $C = (c_{ij})$ matriz de pxq,

entonces esta ley indica que $A(BC) = (AB)C$ se cumple y ABC, definida por cualquiera de ambos lados de esta ecuación, es una matriz de nxq.

Los axiomas iii) y iv) se satisfacen con $\mathbf{0} = I_3$ y $-A = A^{-1}$. No obstante, $AB \neq BA$ en general debido a que, generalmente, el producto de matrices no es conmutativo ($AB \neq BA$) e, incluso, puede que AB esté definida mientras que BA no lo esté y, por ende, en ocasiones ocurre que $AB = BA$ como una excepción, no de una regla, tal que si $AB = BA$ se dice que A y B conmutan aunque el orden de multiplicación de dos matrices debe hacerse con cuidado. Con base en esto, el axioma v) no se cumple y, por lo tanto, S_3 no es $V_{(\text{Espacio Vectorial})}$.

Además, según **Teorema** indica que si V es un espacio vectorial:

i) $\alpha\mathbf{0} = \mathbf{0}$ para $\forall \alpha_{(\text{Escalar})}$

ii) $\mathbf{0} * \mathbf{X} = \mathbf{0}$ para $\forall X \in V$

iii) $\alpha\mathbf{X} = \mathbf{0} \Rightarrow \alpha = 0, X = 0$ o ambos

iv) $(-1)\mathbf{X} = -\mathbf{X}$ para $\forall X \in V$

Demostración:

i) Por axioma iii), $\mathbf{0} + \mathbf{0} = \mathbf{0}$ y axioma vii), $\alpha\mathbf{0} = \alpha(\mathbf{0} + \mathbf{0}) = \alpha\mathbf{0} + \alpha\mathbf{0}$ si se suma $-\alpha\mathbf{0}$ en ambos lados de esta ecuación y axioma ii), ley asociativa, se obtiene $\alpha\mathbf{0} + (-\alpha\mathbf{0}) = [\alpha\mathbf{0} + \alpha\mathbf{0}] + (-\alpha\mathbf{0}) \Rightarrow \mathbf{0} = (\alpha\mathbf{0}) + [\alpha\mathbf{0} + (-\alpha\mathbf{0})] = \alpha\mathbf{0} + \mathbf{0} = \alpha\mathbf{0}$.

ii) Esencialmente, se una la misma prueba anterior. Se inicia con $0 + 0 = 0$ y se usa axioma vii) para comprobar que $0\mathbf{X} = (0 + 0)\mathbf{X} = 0\mathbf{X} + 0\mathbf{X}$ o $0\mathbf{X} + (-0\mathbf{X}) = 0\mathbf{X} + [0\mathbf{X} + (-0\mathbf{X})]$ o $\mathbf{0} = 0\mathbf{X} + \mathbf{0} = 0\mathbf{X}$

iii) Si $\alpha\mathbf{X} = \mathbf{0}$. Si $\alpha \neq 0$ se multiplican ambos lados de esta ecuación por $1/\alpha = \alpha^{-1}$ para obtener $(\alpha^{-1})(\alpha\mathbf{X}) = (\alpha^{-1})\mathbf{0} = \mathbf{0}$ por parte de axioma i), pero $(\alpha^{-1})(\alpha\mathbf{X}) = 1\mathbf{X} = \mathbf{X}$ por axioma ix), tal que $\mathbf{X} = \mathbf{0}$

iv) Se usa el hecho que $1 + (-1) = 0$. Luego, usando axioma ii), se obtiene $\mathbf{0} = 0\mathbf{X} = [1 + (-1)]\mathbf{X} = 1\mathbf{X} + (-1)\mathbf{X} = \mathbf{X} + (-1)\mathbf{X}$. Si se suma $-\mathbf{X}$ en ambos lados de esta ecuación se obtiene $-\mathbf{X} = \mathbf{0} + (-\mathbf{X}) = \mathbf{X} + (-1)\mathbf{X} + (-\mathbf{X}) = \mathbf{X} + (-\mathbf{X}) + (-1)\mathbf{X} = \mathbf{0} + (-1)\mathbf{X} = (-1)\mathbf{X} \Rightarrow \therefore -\mathbf{X} = (-1)\mathbf{X}$. Además, el orden de la suma en esta ecuación se puede invertir mediante axioma v), ley conmutativa.

3.10.2. Subespacios Vectoriales

Los Subespacios Vectoriales se explica debido a que si H es un subespacio vectorial de V si H es un subconjunto no vacío de V, H es un espacio vectorial, junto con operaciones de suma entre vectores y multiplicación por escalar definida para V. Es decir, el subespacio H hereda las operaciones del $V_{(\text{Espacio Vectorial})}$ "padre".

El Teorema Subespacio Vectorial indica que un subconjunto no vacío H de un V es un subespacio V es un subconjunto de V si cumple las reglas de cerradura si un subconjunto no vacío es un subespacio:

i) Si $\mathbf{X} \in$ H y $\mathbf{Y} \in$ H $\Rightarrow \mathbf{X} + \mathbf{Y} \in$ H

ii) Si $\mathbf{X} \in$ H $\Rightarrow \alpha\mathbf{X} \in$ H para $\forall\alpha$

Demostración:

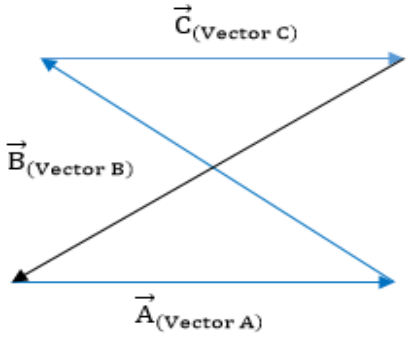
Si H es un V, entonces las dos reglas de cerradura se cumplen; es decir, las dos operaciones de axiomas i)-iv), operaciones de cerradura, se cumplen por hipótesis. Como vectores en H son vectores en V, los axiomas ii), v), vii), viii), ix) y x), identidades asociativa, conmutativa, distributiva y multiplicativa, se cumplen. Sea $\mathbf{X} \in$ H por hipótesis de axioma ii), pero el **Teorema** que indica que si V es un espacio vectorial tal que $\mathbf{0} \in$ H, cumpliéndose axioma iii).

Finalmente, por axioma ii), $(-1)\mathbf{X}$ H para $\forall\, \mathbf{X} \in$ H.

Por **Teorema** que indica que si V es un espacio vectorial, axioma iv), $-\mathbf{X} = (-1)\mathbf{X} \in$ H, cumpliéndose axioma iv) y, con ello, la prueba está completa.

En una ecuación de física se puede deducir que vector posición de un punto en un determinado instante en el tiempo en física se representa mediante la ecuación $\vec{r}_{(\text{Vector Posicion})} = \vec{A}_{(\text{Vector A})} + \vec{B}_{(\text{Vector B})} + \vec{C}_{(\text{Vector C})}$, mientras que gráficamente es:

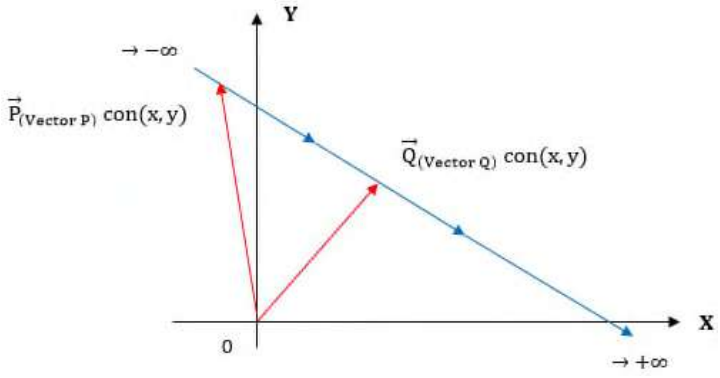
Figura 3.4. Representación gráfica



Fuente: (Mora, 2016)

Entonces, $\overrightarrow{OQ}_{(\text{Vector OQ})} = \overrightarrow{OP}_{(\text{Vector OP})} + PQ_{(\text{Vector PQ})}$ nace de la suma de vectores por el método del polígono:

Figura 3.5. Método del polígono



Fuente: (Mora, 2016)

Sin embargo, esta sumatoria de vectores puede transformarse en la forma de Regresión Lineal Simple $(Y = b + mx)$ y su nomenclatura matemática sería $\vec{r}_{(\text{Vector Posicion})} = \vec{P}_{(\text{Vector P})} +$

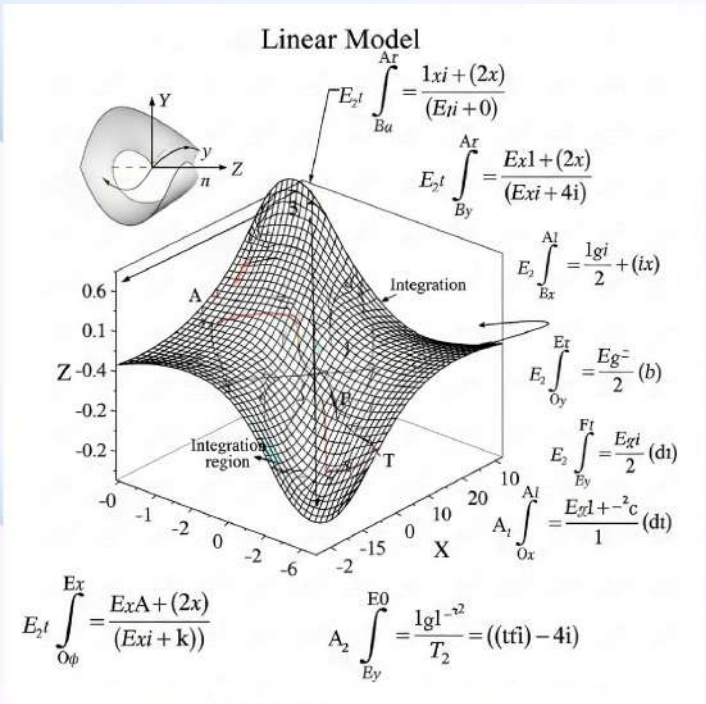
$t_{(\text{n veces la multiplicación de vector } \overrightarrow{PQ})}\overrightarrow{PQ}_{(\text{Vector PQ})} \Rightarrow \vec{r}_{(\text{Vector Posicion})} = \begin{pmatrix} X \\ Y \end{pmatrix} +$

$t\begin{pmatrix} \mu \\ \upsilon \end{pmatrix} \in \mathbb{R} \Rightarrow \vec{r}_{(\text{Vector Posicion})} = (x, y) + t(\mu, \upsilon).$

EDITORIAL ANDES COGNITIO

CAPÍTULO IV

MODELOS LINEALES Y CÁLCULO DE PARÁMETROS MEDIANTE INTEGRALES MÚLTIPLES



CAPÍTULO IV

MODELOS LINEALES Y CÁLCULO DE PARÁMETROS MEDIANTE INTEGRALES MÚLTIPLES

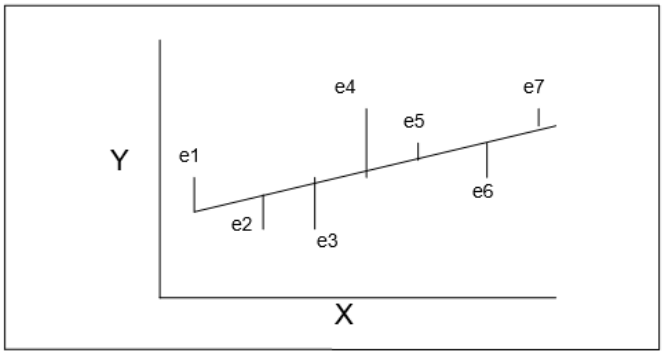
4. Modelos lineales y su estimación de parámetros

Los modelos lineales y su estimación de parámetros por Mínimos Cuadrados (MCO o OLS en inglés) propone que una relación funcional entre Y y X es una línea recta de la forma $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ tal que el paso siguiente es encontrar estimadores para β_0, conocido como intersección o coeficiente de intercepto y β_1, coeficiente de pendiente de la línea recta. Se pueden diseñar diversos métodos algebraicos para realizar esta estimación.

Por ejemplo: se puede aplicar el principio que la suma de residuos sea igual a cero; es decir, $\sum e_i = 0$. Este criterio aseguraría que aquellos residuos iguales en magnitud y signo tendrían la misma importancia; aunque, este método tiene la desventaja que aquellos errores de igual magnitud, pero distinto signo, se cancelan mutuamente.

Sin embargo, el criterio con que trabaja este método es encontrar los estimadores de β_0 y β_1 que minimicen la suma del cuadrado de residuos, sin perder el comportamiento del término error. Entonces, cada residuo (error) es la distancia que existe entre valores observados de Y_i y valores predichos por el modelo.

Figura 4.1. Gráfica de residuos



Fuente: (Gujarati & Porter, 2010)

También es posible aplicar el criterio de minimizar la suma de los valores absolutos de los errores, pero como se puede comprobar, los cálculos son sumamente engorrosos y lógicamente la importancia de errores estará en función de su magnitud y, por lo tanto, una predicción con un error igual a dos unidades será considerada peor que una predicción con dos errores de una unidad cada uno.

Además, se ha desarrollado un método algebraico, relativamente fácil de computar, que penaliza más los errores grandes que los pequeños y también asegura que habrá igual número de errores positivos que negativos, este es el método conocido con el nombre de mínimos cuadrados ordinarios.

Se adoptará convencionalmente la notación de letras minúsculas para indicar estimadores y las letras griegas para denotar los parámetros. La flecha $\rightarrow$ indica "estima a..". De esta manera podemos decir que $b_0 \rightarrow \beta_0$ (se lee b cero estima a beta cero), $b_0 \rightarrow \beta_0$, $e_i \rightarrow \varepsilon_i$ y que $y_i \rightarrow Y_i$ para un X_i dado. Entonces, $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ es el modelo y $y_i = b_0 + b_1 X_i + e_i$ es la ecuación de la recta de mejor ajuste, cuyos coeficientes han sido estimados por medio del método de los mínimos cuadrados ordinarios. Si se define $Y_i - y_i = e_i$ como un residuo, que es la diferencia entre el valor observado y el estimado de Y.

Entonces, $\sum(Y_i - y_i)^2 = \sum e_i^2$ es sumatoria del cuadrado de estos residuos, función que hay que minimizar para asegurar que la suma de cuadrados del error sea un mínimo. Si se reemplaza Y_i por $b_0 + b_1X_i$en esta expresión, el resultado es $\sum(Y_i - b_0 + b_1X_i)^2 = \sum e_i^2$

Tal que la sumatoria va de 1 a n, el tamaño de la muestra. Se sabe por análisis matemático que esta suma de cuadrados depende de los valores de b_0 y b_1; por lo tanto, hay que derivarla parcialmente con respecto a cada uno de estos estimadores e igualar cada ecuación resultante a cero para encontrar su valor mínimo. Entonces, $\frac{\sum e_i^2}{b_0} = 2\sum(Y_i - b_0 + b_1X_i)(-1) = 0$ y $\frac{\sum e_i^2}{b_1} = 2\sum(Y_i - b_0 + b_1X_i)(-X_i) = 0$ tal que arreglando términos se obtienen las ecuaciones normales: $\sum Y_i = nb_0 + b_1\sum X_i$ y $\sum X_iY_i = b_0\sum X_i + b_1\sum X_i^2$.

En consecuencia, se trata de un sistema de dos ecuaciones y dos incógnitas. Para despejar b_0 y b_1 se multiplica la primera ecuación por $\sum X_i$,la segunda por n y se le resta la primera a la segunda:

$$n\sum X_iY_i = nb_0\sum X_i + nb_1\sum X_i^2$$

$$\underline{\sum X_i\sum Y_i = nb_0\sum X_i + b_1\left(\sum X_i\right)^2}$$

$$n\sum X_iY_i - \sum X_i\sum Y_i = nb_1\sum X_i^2 - b_1\left(\sum X_i\right)^2$$

Despejando b_1 de esta última ecuación queda $b_1 = \frac{[n\sum X_iY_i - \sum X_i\sum Y_i]}{[n\sum X_i^2 - (\sum X_i)^2]}$ o, en forma más conocida, se divide por n tanto arriba como abajo $b_1 = \frac{\left[\sum X_iY_i - \left(\frac{\sum X_i\sum Y_i}{n}\right)\right]}{\left[\sum X_i^2 - \left(\frac{(\sum X_i)^2}{n}\right)\right]}$ y la intersección b_0 se puede obtener conociendo b_1 y dividiendo por n la primera ecuación normal, lo que de paso demuestra que la recta de regresión pasa por el punto medio de las observaciones en Y y X: $Y = b_0 + b_1X$. Tal que Y y X son las medias de Y y de X, respectivamente.

Por lo tanto, $b_0 = Y - b_1X$ obtiene la ecuación de ajuste que permite obtener cualquier valor de Y dado un X y resolviendo, de acuerdo con los coeficientes estimados, se obtiene $Y_i = b_0 + b_1X_i$, definida como **ecuación de predicción**. En consecuencia, sea el modelo de regresión lineal simple $Y_i = \beta_0 + \beta_1X_i + \varepsilon_i$ y si se cumplen los supuestos en torno a ε_i, entonces los estimadores mínimo-cuadráticos de β_0 y β_1 son estimadores lineales insesgados y mínima varianza (eficientes).

También, los estimadores son MELI; es decir, los mejores estimadores lineales insesgados. Por lo tanto, un estimador lineal es aquel que puede escribirse como una función lineal de las variables X y Y del modelo.

La prueba de hipótesis ayuda a resolver, junto con partición de suma de cuadrados y análisis de varianza, uno de los problemas en la construcción de modelos econométricos debido a que permite evaluar la capacidad que éstos tienen de representar la realidad contenida en la estructura sugerida por las observaciones muestrales disponibles.

Para ello existen numerosas técnicas, revisaremos aquí una de las más populares denominada análisis de varianza (ANOVA, por sus siglas en inglés). El objetivo es saber si X es un buen predictor de Y, para ello hay que probar $H_0: \beta_1 = 0$ (pendiente cero) versus $H_a: \beta_1 \neq 0$. También, es importante comprobar o "docimar" $H_0: \beta_0 = 0$ (recta para por el origen) versus $H_a: \beta_0 \neq 0$. Tal que, si la pendiente es cero, X no tiene capacidad explicativa de variación de Y.

4.1. Análisis de varianza

En consecuencia, la técnica de análisis de varianza parte del hecho de que la suma de cuadrados totales (SCT) se puede particionar en porciones con significado preciso, pues una parte se debe a la media (SCM), al modelo de regresión (SCR) y, la última no explicada, suma de cuadrados de los residuos (SCE):

Análisis de Varianza (ANOVA ó ADEVA ó ANVA) para Regresión Lineal Simple (RLS)					
Fuente de Variación (F.V.)	Grados de libertad (g.l)	Suma de Cuadrados (SC)	Suma de Cuadrados Medios (SCM)	$F_{Calculado}$	F_{Tablas}
Total	n grados de libertad (total de observaciones)	$\sum Y_i^2$			
Media	1 grado de libertad (siempre)	$\left(\sum Y_i\right)^2/n$			
Regresión	1 grado de libertad (sólo 1 variable explicativa)	$\left[\sum X_i Y_i - \left(\sum X_i \sum Y_i\right)/n\right]^2 / \left[\sum X_i^2 - \left(\left(\sum X_i\right)^2\right)/n\right]$	$\frac{SC_{Regresión}}{gl_{Regresión}}$	$\frac{SCM_{Regresión}}{SCM_{(Error)}}$	
Error	$n-2$	$\sum e_i^2$	$\frac{SC_{Error}}{gl_{Error}}$		

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010) y (Hurtado & Gómez, 2010)

4.1.1. Estimadores

Los estimadores que interesan son b_0, b_1 y Y_i para un X dado. Como todo estimador, constituyen variables aleatorias con sus respectivas varianzas:

Estimador	Parámetro	Varianza
b_1	β_1	$\sigma^2 b_1 = \sigma_e^2\left[\frac{1}{\sum(X_i-\bar{X})^2}\right]$
b_0	β_0	$\sigma^2 b_0 = \sigma_e^2\left[\frac{\sum X_i^2}{n\sum(X_i-\bar{X})^2}\right]$
$\hat{y}_i = b_0 + b_1 X_i$	$Y_i = \beta_0 + \beta_1 X_i$	$\sigma^2_{Y/X} = \sigma_e^2\left[\frac{\frac{1}{n}+(X_0-\bar{X})^2}{n\sum(X_i-\bar{X})^2}\right]$

Con esto, se afirma que estimadores b_0, b_1 y Y_i se distribuyen normalmente y, por tanto, cada uno tendrá una distribución "t" de forma:

Para b_0	$\frac{(b_0 - \beta_0)}{\sigma^2_{b_0}} = t_{(n-2,\alpha)}$
Para b_1	$\frac{(b_1 - \beta_1)}{\sigma^2_{b_1}} = t_{(n-2,\alpha)}$
Para $Y_{i/x}$	$\frac{(Y_{i/x} - \mu_{i/x})}{\sigma^2_{Y/X}} = t_{(n-2,\alpha)}$

Con los estadísticos de "t" se realiza pruebas de hipótesis y se construyen intervalos de confianza para parámetros. Generalmente, la prueba que se hace para cada parámetro es demostrar que son diferentes de cero, pues tiene implicaciones muy importantes para el modelo. Por ejemplo: $H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0 \Rightarrow t_c = \frac{(b_1 - 0)}{\sigma^2_{b_1}}$.

Después, se compara este valor "t" calculado con el obtenido por tablas para un valor α (t_t) determinado. Se trata de una prueba de dos colas, tal que para no aceptar o rechazar H_0, el valor t_c debe ser mayor, en términos absolutos, que el valor t_t. Si no se acepta H_0 se interpreta que la variable explicativa X (variable exógena, independiente o explicativa) es un buen predictor de Y (variable endógena, dependiente o explicada) e, igualmente, aplica para b_0.

Para estimar una proyección de punto Y se usa la recta de ajuste $Y_i = b_0 + b_1 X_i$ valorada para un X_0 determinado. Estos son valores predichos o prefijados cuando los X_0 se encuentran dentro de rango de los X observados. También, se puede estimar un valor de Y que esté fuera del rango de observaciones. Si se trata de una serie de tiempo, el estimador tendrá la categoría de pronóstico.

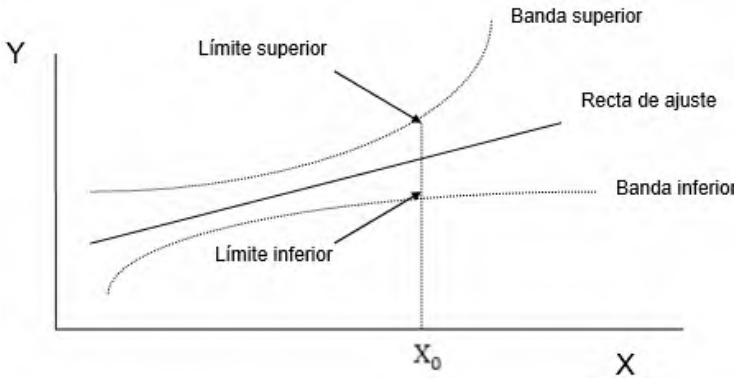
A menudo es de interés obtener una predicción de punto, pero también de un intervalo de confianza para el verdadero valor futuro, que se construye sumando y restándole a la estimación de punto una cantidad que resulta de multiplicar el valor "t" para un α determinado por el error estándar de proyección de punto. Dado que el error estándar de proyección de un punto se incrementa a medida que

X_0 se aleja de la media $\overline{X}$, entonces esta cantidad que se suma y resta a la proyección de punto se irá incrementando, haciendo que la longitud del intervalo sea cada vez más grande.

En consecuencia, aun cuando el modelo haya proporcionado un muy buen ajuste, los intervalos de confianza se van abriendo rápidamente a medida que la proyección se aleja de media de datos. Las curvas que marcan los límites superiores e inferiores de intervalos de confianza se denominan bandas de confianza y se basan en funciones exponencial (conocida formalmente como la función real e^x, donde e es el número de Euler, conocido en ocasiones como número de Euler o constante de Napier, fue reconocido y utilizado por primera vez por el matemático escocés John Napier, quien introdujo el concepto de logaritmo en el cálculo matemático y equivale a 2.71828, aproximadamente.

Esta función tiene por dominio de definición el conjunto de los números reales y tiene la particularidad de que su derivada es la misma función. Se denota equivalentemente como $f(x) = e^x$ o $\exp(x)$, donde e es la base de los logaritmos naturales y corresponde a la función inversa del logaritmo natural) y logarítmica (el método de cálculo mediante logaritmos fue propuesto por primera vez, públicamente, por John Napier (latinizado Neperus) en 1614, en su libro titulado "Mirifici Logarithmorum Canonis Descriptio". Es una función logrítmica básica es la función $y = \log b_x$, donde $b > 0$ y $b \neq 1$):

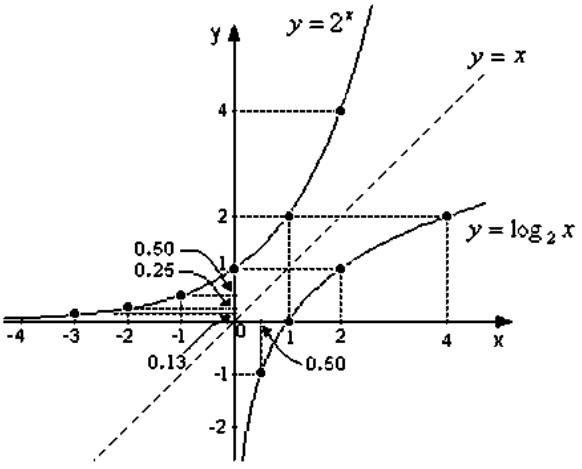
Figura 4.2. Bandas y límites de confianza

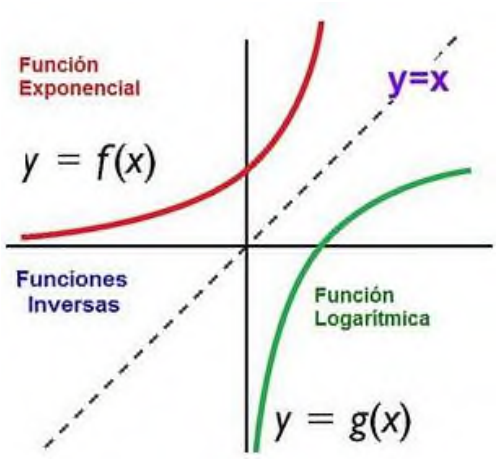


Fuente: (Gujarati & Porter, 2010)

Esta función es la operación inversa a exponenciación de base del logaritmo; es decir, e^x y su dominio es el conjunto de $\mathbb{R}$. Gráficamente:

Figura 4.3. Función exponencial





Fuente: (Gujarati & Porter, 2010) y (Mora, 2016)

Ejemplos:

a) Con base en las siguientes observaciones:

Observaciones		Cálculos		
X	Y	XY	X^2	Y^2
3	2	6	9	4
2	3	6	4	9
2	5	10	4	25
5	10	50	25	100
12	**20**	**72**	**42**	**138**

Haga inferencia estadística sobre los parámetros estimados y las proyecciones Y.

Los errores estándar de estimadores b_0 y b_1 son:

$$\sigma^2_{b_0} = \sigma^2_e \left[\frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}\right] = 7 * \left[\frac{42}{4 * 6}\right] = \sqrt{12.25}$$

$$= \mathbf{3.5}_{(\text{Desviación Estándar o } \sigma_{b_0})}$$

$$\boldsymbol{\sigma_{b_1}^2} = \sigma_e^2\left[\frac{1}{\sum(X_i - \overline{X})^2}\right] = 7 * \left[\frac{1}{6}\right] = \sqrt{1.17} = \mathbf{1.08}_{(\text{Desviación Estándar o } \sigma_{b_1})}$$

Para probar que $b_0 \neq 0$, se divide el estimador b_0 entre su error estándar y se obtiene la "t_c": $\mathbf{t_c} = \frac{b_0}{\sigma_{b_0}} = -\frac{1}{3.5} = \mathbf{-0.2857}$. Después, este valor se compara con el obtenido en tabla de "$t_{(n-2;\ \alpha=0.05)}$" $= 4.303 \Rightarrow |t_c| < |t_t| \therefore$ No se rechaza H_0: $b_0 = 0$, por lo tanto el parámetro de intersección pasa por el origen.

En caso de la pendiente, el valor "t_c" es $\mathbf{t_c} = \frac{b_1}{\sigma_{b_1}} = \frac{2}{1.08} = \mathbf{1.8} \Rightarrow |t_c| < |t_t| \therefore$ No se rechaza H_0: $b_1 = 0$; por lo tanto, X no tiene capacidad predictiva respecto a Y. Suponga que se quiere realizar una proyección para $X_0 = 6 \Rightarrow$ se valora la función estimada:

$$\hat{Y}_1 = -1 + 2(6) = 11$$

$$\hat{Y}_2 = -1 + 2(7) = 13$$

$$\hat{Y}_3 = -1 + 2(8) = 15$$

$$\hat{Y}_4 = -1 + 2(9) = 17$$

$$\hat{Y}_{35} = -1 + 2(40) = 79$$

Si se quiere un intervalo de confianza para esta proyección, se calcula su error estándar:

$\boldsymbol{\sigma_{V/X}^2} = \sigma_e^2\left[\frac{1}{n} * \frac{(X_0 - \overline{X})^2}{n\sum(X_i - \overline{X})^2}\right] = 7 * \left[\frac{(6-3)^2}{6}\right] = 12.25 \Rightarrow \sigma_{V/X} = \sqrt{12.25} =$ $\mathbf{3.50}_{(\text{Desviación Estándar o } \sigma_{b_0})}$. El valor de "$t_{(Gl_{(Error)};\ \alpha=0.05)}$" $= 4.303$ y

"$t_{t\,(Gl_{(Error)};\,\alpha=0.01)}$" = 9.925 ⇒∴ el intervalo de confianza con 95 o 99% de confiabilidad estadística será $\hat{Y}_i \pm \left(t_{t\,(Gl_{(Error)};\,\alpha)} * \sigma_{V/X}\right)$ tal que el verdadero valor de $\hat{Y}_i$ cuando X_0 presenta valores X_i:

X_0	$\hat{Y}_i$	$\sigma^2_{V/X}$	$\sigma_{V/X}$	$t_{(tablas)}$		Intervalo de confianza	
				95%	99%	$\hat{Y}$ (Límite inferior)	$\hat{Y}$ (Límite superior)
6	11	12.25	3.50	4.303	9.925	-4.06	26.06
7	13	20.42	4.52			-2.06	32.44
8	15	30.92	5.56			-0.06	38.93
9	17	43.75	6.61			1.94	45.46
....							
40	79	1598.92	39.99			63.94	251.06

La conclusión es que se trata de un intervalo de confianza demasiado amplio que no tiene aplicación práctica. Esto es, probablemente, por la mala significación estadística del modelo y por el bajo R^2 obtenido.

4.2. Modelo de regresión lineal simple

El modelo de regresión lineal simple puede formalizarse de la siguiente manera:

$$\mathbf{Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i}$$

en donde Y_i es la variable dependiente, X_i es la variable independiente, $\boldsymbol{\beta_0}$ es la intersección de la recta con el eje de las Y, comúnmente llamado también el "intercepto" y $\boldsymbol{\beta_1}$, la pendiente de la recta y finalmente $\boldsymbol{\varepsilon_i}$ es el término error aleatorio.

El modelo supone que para cada valor de X existe una población de Y's, cuyas medias están linealmente relacionadas a X. Por esta razón el modelo también se puede escribir:

$$Y_i = \mu_{y/x} = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Más adelante se verá en detalle que la población de Y's y el término error ε_i se distribuyen igual, es decir, normalmente y con la misma varianza. Aunque la media de ε_i es cero y la de Y es $\mu_{y/x}$.

Cuando el modelo hace referencia a observaciones en el mismo momento en el tiempo, también llamadas observaciones cruzadas (en inglés **cross section**), se utiliza el subíndice i para denotar cada par (X_i , Y_i) donde i = 1, 2, ... , n , siendo n el tamaño de la muestra. Para series temporales, es decir, observaciones en el tiempo, se utiliza el subíndice t (X_t , Y_t) para t = 1, 2, ... , T, siendo T el tamaño de la muestra.

4.2.1. Supuestos del modelo

i. La relación entre Y y X es lineal.

ii. Las X_i son variables no estocásticas cuyos valores son fijos (o controlados por el experimentador).

Uno de los problemas en la construcción de modelos econométricos es poder evaluar la capacidad que éstos tienen de representar la realidad contenida en la estructura sugerida por las observaciones muestrales disponibles. Para ello existen numerosas técnicas, revisaremos aquí una de las más populares denominada **análisis de varianza** (ANOVA, por sus siglas en inglés).

Interesa saber si X es un buen predictor de Y, para ello hay que probar la hipótesis nula de que el parámetro $\beta_1 = 0$ (o sea que la pendiente es cero) versus la alternativa de que es diferente de cero. El lector debe reflexionar sobre las consecuencias que tiene rechazar o no rechazar esta hipótesis (si la pendiente es cero, entonces la X no tiene capacidad explicativa de la variación de Y).

4.2.2. Análisis de varianza

La técnica de análisis de varianza parte del hecho de que la suma de cuadrados totales (SCT) se puede particionar en porciones a las cuales se les puede asignar un significado preciso: una parte debida a la media (SCM), otra debida al modelo de regresión (SCR) y una última no explicada y que es la suma de cuadrados de los residuos (SCE).

SCT	=	SCM	+	SCR	+	SCE
Suma de cuadrados totales		Suma de cuadrados debido a la media		Suma de cuadrados debido a la regresión		Suma de cuadrados debido a los errores

Está claro que la variación total de la variable dependiente se ha particionado en porciones a las cuales se les puede asignar un significado específico. Esta partición es sólo un truco algebraico que parte de una identidad, es decir, no nos comprometemos a asignarles ningún significado estadístico.

4.2.3. Proceso de calculo

El cálculo para cada una de estas sumas de cuadrados es el siguiente:

Suma de cuadrados totales Σy_i^2

Suma de cuadrados debidos a la media $(\Sigma y_i)^2/n$

Suma de cuadrados debidos a la regresión) $[\Sigma y_i x_i - (\Sigma x_i \Sigma y_i)/n]^2/[\Sigma x_i^2 - ((\Sigma x_i)^2)/n]$

Suma de cuadrados debidos a los residuos Σe_i^2

Grados de libertad de estas sumas de cuadrados:

SCT tiene n grados de libertad, es el total de observaciones.

SCM tiene 1 grado de libertad, siempre.

SCR tiene 1 grado de libertad, porque sólo hay una variable explicativa.

SCE tiene los grados de libertad restantes, o sea n - 2.

Si se dividen las sumas de cuadrados por sus respectivos grados de libertad obtendremos las **sumas de cuadrados medios** que pueden considerarse como estimaciones de varianzas. Los únicos que nos interesan son las sumas de los cuadrados medios de la regresión y los del error. Gracias al teorema de Cochrane sabemos que estos cuadrados medios se distribuyen como ji-cuadradas independientes las que divididas por sus propios grados de libertad se distribuyen como una F con los grados de libertad en el numerador correspondientes a los grados de libertad de la regresión y con los grados de libertad en el denominador correspondientes a los del cuadrado medio del error.

De este modo ya se tiene una prueba estadística para probar la hipótesis conjunta respecto a β_1. Los resultados se presentan comúnmente en la llamada **Tabla de Análisis de Varianza** o Tabla ANOVA por sus siglas en inglés.

Tabla Anova

Fuente de variación	Grados de libertad	Suma de cuadrados	Suma de cuadrados medios	F calculada
Total	n	SCT		
Media	1	SCM		

Regresión	1	SCR	SCR/1	SCR/1/SCE/n-2
Error	n - 2	SCE	SCE/n-2	

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010)

Nótese que la SCE/n-2 constituye una estimación de la varianza de la regresión. Es la varianza estimada de los errores que es también la varianza estimada de Y. Su notación es la siguiente:

$$\sigma_e^2 = \Sigma e_i^2/n\text{-}2$$

Para probar la hipótesis de que H_0: $\beta_1 = 0$ versus la alternativa H_a: $\beta_1 \neq 0$ se aprovecha el hecho de que SCR/1/SCE/n-2 sigue una distribución F con 1 grado de libertad en el numerador y n-2 grados de libertad en el denominador.

Si la F calculada en la tabla ANOVA es superior a la dada por las tablas para un α determinado (1-α es la significancia) no se acepta la hipótesis nula y se concluye que el parámetro β_1 es diferente de cero, indicando con ello que X sí tiene capacidad predictiva sobre Y.

La evaluación global del ajuste de la regresión se puede hacer con $SCR_{(Regresión)}$ o, mejor, con varianza muestral de residuos $\left(\frac{1}{n}\right)\sum e_i^2$. Pero residuos no son todos independientes, sino que están ligados por las dos primeras ecuaciones:

i. Suma de residuos es cero: $\sum e_i = 0$.
ii. Suma de residuos ponderada por valores de variable regresora es cero: $\sum x_i e_i = 0$.
iii. $\sum y_i = \sum \hat{y}_i$

iv. Suma de residuos ponderada por predicciones de valores observados es cero: $\sum \hat{y}_i \, e_i = 0$.

Tal que se usa la llamada "varianza residual" o estimación MC de σ^2:

$$\hat{\sigma}^2 = \left[\frac{SCR_{(Residuos)}}{(n-2)}\right]$$

Su raíz cuadrada $\hat{\sigma}$, que tiene las mismas unidades que Y, es llamado "error estándar de regresión". La varianza residual o error estándar son índices de previsión del modelo, pero dependen de unidades de la variable respuesta y no son útiles para comparar rectas de regresión de variables diferentes. Otra medida de ajuste requiere una adecuada descomposición de variabilidad de variable respuesta.

Teorema. Considere el coeficiente de correlación muestral, cuyo significado es convencional:

$$r_{(Coeficiente\ de\ correlación\ muestral)} = \left(\frac{S_{xy}}{S_x S_y}\right) = \left(\frac{S_{xy}}{(S_x S_y)^{1/2}}\right)$$

Entonces, se verifican las relaciones siguientes (Carmona, 2003):

i. $\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$

ii. $SCR_{(Residuos)} = \sum (y_i - \hat{y}_i)^2 = (1 - r^2) \sum (\hat{y}_i - \bar{y})^2 = (1 - r^2) S_y$

iii. $\hat{\sigma}^2 = \left(\frac{\sum e_i^2}{(n-2)}\right) = \left(\frac{(1-r^2)S_y}{(n-2)}\right)$

Demostración: i) $\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 + 2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$, pero $\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \sum (y_i - \hat{y}_i) = 0$ por propiedades anteriores. Además, recuerde la ortogonalidad de sub espacios de errores y estimaciones. Es fácil ver:

$$\sum(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum(x_i - \bar{x})^2 = r^2 \sum(y_i - \bar{y})^2$$

Tal que:

$$\sum(y_i - \bar{y})^2 = \sum(y_i - \hat{y}_i)^2 + r^2 \sum(y_i - \bar{y})^2$$

Entonces:

$$\sum(y_i - \hat{y}_i)^2 = (1 - r^2) \sum(y_i - \bar{y})^2$$

El estimador centrado de varianza σ^2 de modelo, con Y variable aleatoria y X variable controlable; es decir, los valores que toma X son fijados por el experimentador y suponiendo que se calcula Y para diferentes valores de X según el modelo, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \; \forall \, i = 1,2,3,4, \dots, n$ tal que $E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2 \; \forall \, i = 1,2,3,4, \dots, n$ es $\hat{\sigma}^2 = \left[\frac{SCR_{(Residuos)}}{(n-2)}\right] = \left(\frac{(1-r^2)S_y}{(n-2)}\right)$. La descomposición de suma de cuadrados de observaciones en dos términos independientes se interpreta como "la variabilidad de Y se descompone en un primer término que refleja la variabilidad no explicada por regresión, por azar y segundo término que contiene la variabilidad explicada o eliminada por regresión tal que puede interpretarse como la parte determinista de variabilidad de respuesta:

$$VT_{(Variación\ total)} = \sum(y_i - \bar{y})^2 = S_y$$

$$VNE_{(Variación\ no\ explicada)} = \sum(y_i - \hat{y}_i)^2 = SCR_{(Residuos)}$$

$$VE_{(Variación\ explicada)} = \sum(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 S_x$$

Tal que $VT_{(\text{Variación total})}(\sum(y_i - \bar{y})^2) = VNE_{(\text{Variación no explicada})}(\sum(y_i - \hat{y}_i)^2) + VE_{(\text{Variación explicada})}(\sum(\hat{y}_i - \bar{y})^2)$ o, también, $VT_{(\text{Variación total})}(S_y) = VNE_{(\text{Variación no explicada})}(SCR_{(\text{Residuos})}) + VE_{(\text{Variación explicada})}(\hat{\beta}_1^2 S_x)$.

Definición. Una medida del ajuste de recta de regresión a datos es la proporción de variabilidad explicada definida como "coeficiente de determinación":

$$R^2 = \left(\frac{VE_{(\text{Variación explicada})}(\sum(\hat{y}_i - \bar{y})^2)}{VT_{(\text{Variación total})}(S_y)}\right)$$

$$= 1 - \left(\frac{VNE_{(\text{Variación no explicada})}(SCR_{(\text{Residuos})})}{VT_{(\text{Variación total})}(S_y)}\right)$$

Esta medida se puede usar en cualquier tipo de regresión, pero en caso particular de regresión lineal simple con una recta es:

$$R^2 = 1 - \left(\frac{(1 - r^2)\left(VT_{(\text{Variación total})}(S_y)\right)}{VT_{(\text{Variación total})}(S_y)}\right)$$

Entendido como cuadrado del coeficiente de correlación lineal entre dos variables. El coeficiente de determinación R^2 es una medida de bondad del ajuste, $0 \leq R^2 \leq 1$ mientras que el coeficiente de correlación es una medida de dependencia lineal entre dos variables, cuando son aleatorias y sólo hay una variable regresora.

Considere la situación en que dos variables aleatorias, tanto variable respuesta como explicativa o regresora tal que se toma una muestra aleatoria simple de tamaño n formada por pares ordenados $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \ldots, (x_n, y_n)$ de dos variables aleatorias (X, Y) con distribución conjunta normal bivariante:

$$(X, Y)' \sim N_2(\mu, \Sigma) \text{ tal que } \mu = (\mu_1, \mu_2)' \text{ y } \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

Donde $\mathrm{Cov}(X, Y) = \sigma_1\sigma_2\rho$ y ρ es coeficiente de correlación X y Y. La distribución condicionada de Y dado un valor de $X = x$ es:

$$Y|X = x \sim N(\beta_0 + \beta_1 x, \sigma^2_{2*1})$$

Donde:

$$\beta_0 = \mu_1 - \left(\rho\mu_2 * \frac{\sigma_2}{\sigma_1}\right);\ \beta_1 = \left(\rho * \frac{\sigma_2}{\sigma_1}\right) \text{ y } \sigma^2_{2*1} = \sigma^2_2(1 - \rho^2)$$

Tal que, la esperanza de $Y|X = x$ es el modelo de regresión lineal simple:

$$E[Y|X = x] = \beta_0 + \beta_1 x$$

Existe una clara relación entre β_1 y ρ, $\rho = 0$ ssi $\beta_1 = 0$, que no existe regresión lineal; es decir, el conocimiento de $X = x$ no ayuda a predecir Y. El método de máxima verosimilitud proporciona estimadores de β_0 y β_1 que coinciden con estimadores $\mathrm{MC}_{(\text{Mínimos Cuadrados})}$. Es posible plantear inferencias sobre parámetro ρ. En primer lugar, el estimador natural ρ es:

$$r_{(\text{Coeficiente de correlación muestral})} = \left(\frac{S_{xy}}{S_x S_y}\right) = \left(\frac{S_{xy}}{(S_x S_y)^{1/2}}\right)$$

Mientras que:

$$\hat{\beta}_1 = \left(\frac{S_y}{S_x}\right)^{1/2} * r_{(\text{Coeficiente de correlación muestral})}$$

$\hat{\beta}_1$ y $r_{(\text{Coeficiente de correlación muestral})}$ están relacionados, mientras $r_{(\text{Coeficiente de correlación muestral})}$ representa una medida de asociación entre X i Y, $\hat{\beta}_1$ mide el grado de predicción en Y por unidad de X; aunque, cuando X es variable controlada, $r_{(\text{Coeficiente de correlación muestral})}$ tiene un significado

convencional, pues su magnitud depende de la elección del espacio de valores x_i tal que $\rho \notin$ y $r_{(\text{Coeficiente de correlación muestral})}$ no es un estimador.

Además, $r^2 = R^2$ tal que el coeficiente de determinación es cuadrado de correlación. Por último, el principal contraste sobre ρ es correlación $H_0: \rho = 0$ equivalente a $H_0: \beta_1 = 0$ y se estima con estadístico:

$$t = \left(\frac{r_{(\text{Coeficiente de correlación muestral})} \sqrt{n-2}}{\sqrt{1 - r^2_{(\text{Coeficiente de correlación muestral})}} \right)$$

Si H_0 es cierta sigue una distribución $t_{(n-2)}$.

4.2.4. Bondad de ajuste del modelo

La bondad de ajuste del modelo con relación a la muestra utilizada puede medirse también a través del llamado coeficiente de determinación R^2. Este coeficiente mide la proporción de la suma de cuadrados explicada por la variable independiente, una vez descontado el efecto de la media. El resto lo constituye la suma de cuadrados debida al error, no explicada por X.

Una forma económica de calcular el R^2 es utilizando las sumas de cuadrados proporcionadas por la tabla ANOVA, de esta manera se define R^2 como:

$$R^2 = SCR/(SCR+SCE)$$

De esta expresión se puede colegir que si la suma de cuadrados del error es cero, o sea si el ajuste es perfecto, entonces R^2 tomará el valor de 1. Por el contrario, si la SCR decrece con relación a la SCE entonces el R^2 se acerca a cero como límite inferior.

No existe una regla bien demarcada para definir cuando un R^2 es bueno o malo, pero en econometría, de acuerdo con el tipo de modelos que se utilizan y la calidad de la información disponible, generalmente se acepta que un R^2 arriba de 0.8 es bueno. Sin embargo, valores entre 0.6 y 0.8 todavía pueden considerarse aceptables en no pocas ocasiones.

El modelo $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \; \forall \, i = 1,2,3,4,\dots,n$ tal que $E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2 \; \forall \, i = 1,2,3,4,\dots,n$ es $\hat{\sigma}^2 = \left[\frac{SCR_{(Residuos)}}{(n-2)}\right] = \left(\frac{(1-r^2)S_y}{(n-2)}\right)$ es la formulación lineal del problema de hallar la recta de regresión de Y sobre X. Los parámetros β_0, β_1 reciben el nombre de "coeficientes de regresión". El parámetro β_0 es la ordenada en origen o intercept en inglés y β_1 es la pendiente de recta o slope en inglés. La expresión matricial del modelo es:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \vdots \\ \varepsilon_n \end{pmatrix} \text{ regresión } x = 2$$

Suponga que está interesado en decidir i regresión de Y sobre X es realmente lineal tal que se considera la hipótesis:

$$H_0: Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$H_1: Y_i = g(x_i) + \varepsilon_i$$

$g(x_i)$ es una función no lineal desconocida de X. No obstante, el contraste se puede reconducir. Se requiere n_i valores de Y para cada x_i. Con cambio de notación, para cada $i = 1, 2, 3, 4, \dots, k$, sean:

$$x_i: y_{i1}, y_{i2}, y_{i3}, y_{i4}, \dots, y_{in_i};\ \bar{y}_i = \left(\frac{1}{n_i}\right)\sum_j y_{ij}\ ;\ s_{y_i}^2 = \left(\frac{1}{n_i}\right)\sum_j \left(y_{ij} - \bar{y}_i\right)^2\ ;\ \bar{y}$$

$$= \left(\frac{1}{n}\right)\sum_{i,j} y_{ij}\ ;\ s_y^2 = \left(\frac{1}{n}\right)\sum_{i,j} \left(y_{ij} - \bar{y}_i\right)^2\ n$$

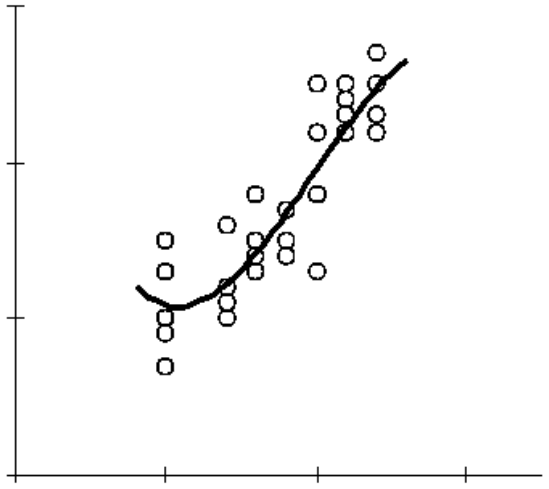
$$= n_1 + n_2 + n_3 + n_4 + \cdots + n_k$$

Se introduce el coeficiente:

$$\hat{\eta}^2 = 1 - \left(\frac{1}{n}\right)\sum_{i=1}^{k}\left(n_i * \frac{s_{y_i}^2}{s_y^2}\right)$$

Que verifica $0 \leq \hat{\eta}^2 \leq 1$,mide el grado de concentración de puntos (x_i, y_{ij}) en curva $y = g(x_i)$ (curva que mejor se ajusta a datos):

Figura 4.4. Representación gráfica



Fuente: (Gujarati & Porter, 2010)

Si se indica $\delta_i = g(x_i)$ con $i = 1,2,3,4, \dots, k$ se convierte hipótesis H_1 en hipótesis lineal con k parámetros. Cuando H_1 es cierta, la estimación de $\delta_i = \bar{y}_i$. La identidad:

$$\mathrm{SCR}_{(\mathrm{Hipótesis})} = \mathrm{SCR}_{(\mathrm{Residuos})} + \left(\mathrm{SCR}_{(\mathrm{Hipótesis})} - \mathrm{SCR}_{(\mathrm{Residuos})}\right)$$

Entonces, es:

$$\frac{\sum_{i,j}\left(y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{n} = \frac{\sum_{i,j}\left(y_{ij} - \bar{y}_i\right)^2}{n} + \frac{\sum_i n_i\left(\bar{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{n}$$

$$\Rightarrow s_y^2\left(1 - r^2_{(\text{Coeficiente de correlación muestral})}\right)$$

$$= s_y^2(1 - \hat{\eta}^2) + s_y^2\left(\hat{\eta}^2 - r^2_{(\text{Coeficiente de correlación muestral})}\right)$$

El contraste que decide si regresión es o no lineal se estima mediante el estadístico:

$$F_{(\text{F de Fisher})} = \left[\frac{\dfrac{\left(\hat{\eta}^2 - r^2_{(\text{Coeficiente de correlación muestral})}\right)}{(k-2)}}{\dfrac{(1-\hat{\eta}^2)}{(n-k)}}\right]$$

Tiene $(k-2)$ y $(n-k)$ grados de libertad tal que si $F_{(\text{F de Fisher})}$ resulta significativa no se acepta su carácter lineal de regresión. Es importante considerar:

i) Solamente se puede aplicar esta prueba si se tiene $n_i > 1$ observaciones de Y para cada x_i $(i = 1,2,3,4, \dots, k)$.

ii) $\hat{\eta}^2$ es una versión muestral de "razón de correlación" entre dos variables aleatorias, representadas por X, Y:

$$\eta^2 = \frac{E[(g(X) - E(Y))^2]}{Var(Y)}$$

Tal que $y = g(X) = E(Y|X = x)$ la curva de media de Y sobre X. Este coeficiente η^2 verifica que:

a) $0 \leq \eta^2 \leq 1$.

b) $\eta^2 = 0 \Rightarrow y = E(Y)$ tal que la curva es la recta $y =$ constante

c) $\eta^2 = 1 \Rightarrow y = g(X)$ tal que Y es función de X

iii) De forma análoga, la hipótesis Y es alguna función, no lineal, de X frente a hipótesis nula que no hay ningún tipo de relación se puede plantear. Las hipótesis son:

$$H_0: y_i = \mu_{(Constante)} + \varepsilon_i$$

$$H_1: y_i = g(x_i) + \varepsilon_i$$

Con igual notaciones:

$$SCR_{(Hipótesis)} = \sum_{i,j} (y_{ij} - \bar{y}_i)^2 \text{ con } n - 1 \text{ grados de libertad}$$

$$SCR_{(Residuos)} = \sum_{i,j} (y_{ij} - \bar{y}_i)^2 \text{ con } n - k \text{ grados de libertad}$$

Tal que:

$$F_{(F\ de\ Fisher)} = \left[\frac{\frac{(\hat{\eta}^2)}{(k-1)}}{\frac{(1-\hat{\eta}^2)}{(n-k)}} \right]$$

Interpretada como prueba de significancia de razón de correlación.

4.3. Estimaciones de coeficientes de regresión

Las estimaciones de coeficientes de regresión con datos observados se estiman mediante estadísticos:

$$\bar{x} = \left(\frac{1}{n}\right)\sum x_i \,;\, s_x^2 = \left(\frac{1}{n}\right)\sum (x_i - \bar{x})^2 \,;\, \bar{y} = \left(\frac{1}{n}\right)\sum y_i \,;\, s_y^2$$
$$= \left(\frac{1}{n}\right)\sum (y_i - \bar{y})^2 \text{ y } s_{xy} = \left(\frac{1}{n}\right)\sum (x_i - \bar{x})\,(y_i - \bar{y})$$

$\bar{x}, \bar{y}, s_x^2, s_y^2$ y s_{xy} son medias, varianzas y covarianzas muestrales; aunque, el significado de s_x^2 y s_{xy} es convencional, pues x no es variable aleatoria. Con esta notación las ecuaciones normales son:

$$X'X\beta = X'Y \Leftrightarrow \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$$

Como:

$$(X'X)^{-1} = \left(\frac{1}{n s_x^2}\right)\begin{pmatrix} \left(\frac{1}{n}\right)\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

La solución es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x};\ \hat{\beta}_1 = \left(\frac{s_{xy}}{s_x}\right) = \left(\frac{s_{xy}}{s_x^2}\right)$$

Tal que:

$$s_{xy} = \sum (x_i y_i) - \left(\frac{1}{n}\right)\sum x_i \sum y_i = \sum (x_i - \bar{x})\,(y_i - \bar{y}) = n s_{xy}$$

$$s_x = \sum x_i^2 - \left(\frac{1}{n}\right)\left(\sum x_i\right)^2 = \sum (x_i - \bar{x})^2 = n s_x^2$$

Tal que, la recta de regresión es $y = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$ o, en forma, $y - \bar{y} = \hat{\beta}_1(x - \bar{x})$. La recta pasa por el par ordenado $(\bar{x}, \bar{y})$ y el modelo es válido en rango de x_i. Esta es la recta que se obtiene del modelo equivalente con datos x_i centrados. Estas

estimaciones son insesgadas y varianza mínima entre todos los estimadores lineales, teorema de Gauss-Markov.

Las varianzas y covarianzas de estimadores son:

$$\mathrm{Var}(\hat{\beta}_1) = \begin{pmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) \end{pmatrix} = \sigma^2 (X'X)^{-1} \Rightarrow \mathrm{E}(\hat{\beta}_0)$$

$$= \beta_0;\ \mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right);\ \mathrm{E}(\hat{\beta}_1) = \beta_1; \mathrm{Var}(\hat{\beta}_1)$$

$$= \left(\frac{\sigma^2}{s_x}\right) \text{ y } \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\left(\sigma^2 * \frac{\bar{x}}{s_x}\right)$$

Para contraste de hipótesis $H_0: \beta_0 = b_0$ se usa estadístico:

$$t_{(\text{Student})} = \left(\frac{\hat{\beta}_0 - b_0}{\left(\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right)\right)^{1/2}} \right)$$

Si no se rechaza la hipótesis, sigue una distribución $t_{(\text{Student});(n-2)\text{grados de libertad}}$.

Justificadamente, suponga que el investigador propone el modelo de regresión simple que carece de parámetro intercepto β_0:

$$y_i = \beta_1 x_i + \varepsilon_i \ \ i = 1,2,3,4, \ldots, n$$

Tal que, el estimador MC de parámetro β_1 es:

$$\hat{\beta}_1 = \left(\frac{\sum x_i y_i}{\sum x_i^2}\right)$$

Y su varianza es:

$$\mathrm{Var}(\hat{\beta}_1) = \left(\frac{1}{(\sum x_i^2)^2}\right)\sum x_i^2 * \mathrm{Var}(y_i) = (\sigma^2)\left(\frac{1}{\sum x_i^2}\right)$$

El estimador de σ^2 es:

$$\hat{\sigma}^2 = \left(\frac{\mathrm{SCR}_{(\mathrm{Residuos})}}{n-1}\right) = \left(\frac{1}{n-1}\right)\left(\sum y_i^2 - \hat{\beta}_1 * \sum x_i y_i\right)$$

Con base en hipótesis de normalidad se construye intervalos de confianza $100 * (1-\alpha)\%$ para β_1:

$$\hat{\beta}_1 \pm t_{(\mathrm{Student})(n-1)}(\alpha) * \hat{\sigma}\sqrt{\left(\frac{1}{\sum x_i^2}\right)}$$

Para $E[Y|x_0]$:

$$\hat{y}_0 \pm t_{(\mathrm{Student})(n-1)}(\alpha) * \hat{\sigma}\sqrt{\left(\frac{x_0^2}{\sum x_i^2}\right)}$$

Para predicción de observación:

$$\hat{y}_0 \pm t_{(\mathrm{Student})(n-1)}(\alpha) * \hat{\sigma}\sqrt{\left(1 + \frac{x_0^2}{\sum x_i^2}\right)}$$

Frecuentemente, la relación entre variables respuesta o endógena Y y regresora o exógena x varía cerca del origen, como química. El diagrama de dispersión ayuda a decidir mejor modelo. Si no está seguro, se recomienda usar el modelo completo y contrastar $H_0: \beta_0 = 0$. *Una medida de ajuste de modelo a datos es error cuadrático medio* $\hat{\sigma}^2 = \left(\frac{\mathrm{SCR}_{(\mathrm{Residuos})}}{n-1}\right) = \left(\frac{1}{n-1}\right)\left(\sum y_i^2 - \hat{\beta}_1 * \sum x_i y_i\right)$ *que es*

viable compararlo con $\hat{\sigma}^2 = \left(\frac{SCR_{(Residuos)}}{n-2}\right) = \left[\frac{\left(1-r^2_{(Coeficiente\ de\ correlación\ muestral)}\right)S_{y(Variación\ total)}}{(n-2)}\right]$.

El coeficiente de determinación R^2 *no es buen índice para comparar dos tipos de modelos. El modelo sin* β_0 la descomposición es:

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

Justifica la definición de coeficiente de determinación es:

$$R_0^2 = \left(\frac{\sum \hat{y}_i^2}{\sum y_i^2}\right)$$

No es comparable con coeficiente de determinación $R^2 = \left(\frac{VE_{(Varianza\ explicada)}}{VT_{(Varianza\ total)}}\right) = 1 - \left(\frac{SCR_{(Residuos)}}{s_{y\ (Varianza\ total)}}\right)$, que es una medida del ajuste de recta de regresión a datos es proporción de variabilidad explicada y puede usarse en cualquier tipo de regresión, en caso de regresión lineal simple con una recta es $R^2 = 1 - \left(\frac{\left(1-r^2_{(Coeficiente\ de\ correlación\ muestral)}\right)S_{y(Variación\ total)}}{S_{y(Variación\ total)}}\right) = r^2_{(Coeficiente\ de\ correlación\ muestral)}$ interpretada como cuadrado del coeficiente de correlación lineal entre dos variables, tal que $R_0^2 > R^2$; aunque, $CME_{0\ (Error)} < CME_{(Error)}$.

Suponga que modelo $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ \forall\ i = 1,2,3,4, \dots, n$ tal que $E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2\ \forall\ i = 1,2,3,4, \dots, n$ es $\hat{\sigma}^2 = \left[\frac{SCR_{(Residuos)}}{(n-2)}\right] = \left(\frac{\left(1-r^2_{(Coeficiente\ de\ correlación\ muestral)}\right)S_{y(Variación\ total)}}{(n-2)}\right)$ es lineal normal tal que

Teorema (sea $Y \sim N(X\beta, \sigma^2 I_n)$ con rango $X = m$ tal que se verifican las

propiedades: i) Estimación MC de β coincide con estimación de máxima verosimilitud. Además, es insesgada y mínima varianza.

Demostración: función de verosimilitud es $L(Y;\beta,\sigma^2) = \sqrt{2\pi\sigma^2}e^{\left[-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right]}$ de modo que el mínimo de $(Y-X\beta)'(Y-X\beta)$ es máximo de L tal que $\hat{\beta} = (X'X)^{-1}X'Y$ tal que $E(\hat{\beta}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\beta = \beta$ tal que cada $\hat{\beta}_i$ es un estimador de varianza mínima de β_i debido a que es centrado, máxima verosimilitud y suficiente.

Otra forma de demostrarlo es mediante

Teorema: si $\psi = a'\beta$ una función paraamétrica estimable y $\hat{\beta}$ es estimador MC de β tal que el estimador $\widehat{\psi}_{(\text{único})} = a'\hat{\beta}$.

Demostración: si ψ es una función paramétrica estimable, tiene un estimador lineal insesgado $b'Y$, donde b es vector $n * 1$ tal que se considera el sub espacio $\Omega = \langle X \rangle$ de $\mathbb{R}^n$ generado por columnas de X.

El vector b se puede descomponer de forma única: $b = \tilde{b} + c;\ \tilde{b} \in \Omega$ y $c \perp \Omega$ tal que c es ortogonal a todo vector de Ω. Considere el estimador lineal $\tilde{b}'Y$ tal que es insesgado y su valor es único. Se sabe que $b'Y$ es insesgado pues $\psi = a'\beta = E(\tilde{b}'Y) + E(c'Y) = E(\tilde{b}'Y) = \tilde{b}'X\beta$ luego $E(\tilde{b}'Y) = a'\beta$, pues $E(c'Y) = c'E(Y) = c'X\beta = 0\beta = 0$.

Suponga que $b^{*'}Y$ es otro estimador insesgado para ψ y $b^* \in \Omega$ tal que $0 = E(\tilde{b}'Y) - E(b^{*'}Y) = (\tilde{b}' - b^{*'})X\beta$ luego $(\tilde{b}' - b^{*'})X = 0$ tal que $(\tilde{b}' - b^{*'})$ es ortogonal a Ω, como pertenece a Ω debe ser $\tilde{b} - b^* = 0$ tal que $\tilde{b} = b^*$aunque se sabe que para cualquier estimador MC de $\beta e = Y - Y\hat{\beta}$ es ortogonal a Ω tal que

$0 = \tilde{b}'e = \tilde{b}'Y - \tilde{b}'X\hat{\beta}$ así $\tilde{b}'Y = \tilde{b}'X\hat{\beta}$ y se sabe que $\tilde{b}'Y = b'X = a'$ tal que $\tilde{b}'Y = a'\hat{\beta}\ \forall\ \hat{\beta}$; ii) $\hat{\beta} \sim N\left(\beta, \sigma^2(X'X)^{-1}\right)$.

Demostración: $\hat{\beta} = \left[(X'X)^{-1}X'\right]Y$, $\hat{\beta}$ es combinación lineal de una normal y, por ello, tiene distribución normal multivariante con matriz de varianzas-covarianzas $(X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} = (X'X)^{-1}\sigma^2$; iii) $(\hat{\beta} - \beta)'X'X\left(\frac{(\hat{\beta}-\beta)}{\sigma^2}\right) \sim \chi^2_m$.

Demostración: es consecuencia de propiedades de normal multivariante anterior; iv) $\hat{\beta}$ es independiente de $SCR_{(Residuos)}$.

Teorema: la función paramétrica $a'\hat{\beta}$ es estimable ssi $a \in \langle X' \rangle = \langle X'X \rangle$.

Demostración: sabe que función paramétrica $a'\hat{\beta}$ es estimable ssi a es combinación lineal de filas X; es decir, cuando $a \in \langle X' \rangle$ tal que sólo queda probar $\langle X' \rangle = \langle X'X \rangle$, pero $X'X_c = X'd$ para $d = X_c$ tal que $\langle X'X \rangle \subset \langle X' \rangle$.

Además, las dimensiones de ambos sub espacios son iguales, pues regresión $\langle X' \rangle$ = regresión $\langle X'X \rangle$, donde se deduce la igualdad y v) $\left(\frac{SCR_{(Residuos)}}{\sigma^2}\right) \sim \chi^2_{n-m}$), aplicando $\left(\frac{SCR_{(Residuos)}}{\sigma^2}\right) = \left(\frac{z_{m+1}}{\sigma^2}\right)^2 + \left(\frac{z_{m+2}}{\sigma^2}\right)^2 + \left(\frac{z_{m+2}}{\sigma^2}\right)^2 + \left(\frac{z_{m+4}}{\sigma^2}\right)^2 + \cdots + \left(\frac{z_n}{\sigma^2}\right)^2$ se obtiene una suma de cuadrados $n - m$ variables normales independientes; es decir, una distribución χ^2_{n-m}, bajo ciertas condiciones generales se prueba que $\hat{\sigma}^2 = \left(\frac{SCR_{(Residuos)}}{(n-m)}\right)$ es un estimador de varianza mínima de σ^2, se verifica:

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)' \sim N_2(\beta, Var(\hat{\beta}))$$

Donde:

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} = \sigma^2 \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}}{s_x}\right) & -\frac{\bar{x}}{s_x} \\ -\frac{\bar{x}}{s_x} & \frac{1}{s_x} \end{pmatrix}$$

Se sabe $\hat{\beta}$ es independiente de $\mathrm{SCR}_{(\mathrm{Residuos})}$ tal que para contrastar una hipótesis de tipo $H_0: a'\beta = c$ se usa el estadístico $t_{(\mathrm{Student})}$:

$$t_{(\mathrm{Student})} = \left[\frac{a'\hat{\beta} - c}{\left(\hat{\sigma}^2(a'(X'X)^{-1}a)\right)^{1/2}}\right]$$

Esta sigue una distribución $t_{(n-2)}$ si H_0 no se rechaza o no es falsa.

El contraste de hipótesis $H_0: \beta_1 = b_1$ *versus* $H_1: \beta_1 \neq b_1$ *se resuelve no aceptando* H_0 *si:*

$$\left|\frac{\beta_1 - b_1}{\left(\frac{\hat{\sigma}^2}{s_x}\right)^{1/2}}\right| > t_{(n-2)}(\alpha)$$

Donde $P\left[\left|t_{(n-2)}\right| > t_{(n-2)}(\alpha)\right] = \alpha$. En particular, interesa contrastar si la pendiente es cero; es decir, la hipótesis $H_0: \beta_1 = 0$ *tal que si* $H_0: \beta_1 = 0$ *no es falsa, el modelo* $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \; \forall \, i = 1,2,3,4,\ldots,n$ donde $E(\varepsilon_i) = 0, \mathrm{var}(\varepsilon_i) = \sigma^2 \; \forall \, i = 1,2,3,4,\ldots,n$ se simplifica y convierte en:

$$y_i = \beta_0 + \varepsilon_i \; \forall \, i = 1,2,3,4,\ldots,n$$

Donde:

$$\mathrm{SCR}_{(\mathrm{Hipótesis})} = \sum \left(y_i - \hat{\beta}_0 \middle| H_{(\mathrm{Hipótesis})}\right)^2 = \sum (y_i - \bar{y})^2 = s_{y\ (\mathrm{Variación\ total})}$$

Dado que $\hat{\beta}_0 \left| H_{(\text{Hipótesis})} = \bar{y} \right.$.

Por **Teorema** $r_{(\text{Coeficiente de correlación muestral})} = \left(\frac{S_{xy}}{S_x S_{y\,(\text{Variación total})}} \right) = \left(\frac{S_{xy}}{(S_x S_{y\,(\text{Variación total})})^{1/2}} \right)$ se sabe que $SCR_{(\text{Residuos})} = \left(1 - r^2_{(\text{Coeficiente de correlación muestral})}\right) * S_y$ tal que:

$$F_{(\text{F de Fisher})} = \left(\frac{SCR_{(\text{Hipótesis})} - SCR_{(\text{Residuos})}}{\left(\frac{SCR_{(\text{Residuos})}}{n-2} \right)} \right)$$

$$= \left(\frac{S_{y\,(\text{Variación total})} - \left(1 - r^2_{(\text{Coeficiente de correlación muestral})}\right) * S_{y\,(\text{Variación total})}}{\left(\frac{\left(1 - r^2_{(\text{Coeficiente de correlación muestral})}\right) * S_{y\,(\text{Variación total})}}{n-2} \right)} \right)$$

$$= (n-2)\left(\frac{r^2_{(\text{Coeficiente de correlación muestral})}}{1 - r^2_{(\text{Coeficiente de correlación muestral})}} \right) \sim F_{(\text{F de Fisher});(1;n-2)}$$

Además:

$$t_{(\text{Student})} = \sqrt{F_{(\text{F de Fisher})}} = r * \left(\frac{\sqrt{n-2}}{1 - r^2_{(\text{Coeficiente de correlación muestral})}} \right)$$

Sigue una distribución $t_{(\text{Student})}$ con $n-2$ grados de libertad.

4.4. Contraste para significación de regresión

Este contraste $H_0: \beta_1 = 0$ *se llama* "contraste para significación de regresión" *y se formaliza en una tabla de análisis de varianza* (ANOVA, ANVA, ANDEVA, ADEVA) *en que explica la descomposición de suma de cuadrados* $VT_{(\text{Variación total})}(\sum(y_i - \bar{y})^2) = VNE_{(\text{Variación no explicada})}(\sum(y_i - \hat{y}_i)^2) + VE_{(\text{Variación explicada})}(\sum(\hat{y}_i - \bar{y})^2)$

o, igual, $VT_{(\text{Variación total})}(S_y) = VNE_{(\text{Variación no explicada})}\left(SCR_{(\text{Residuos})}\right) + VE_{(\text{Variación explicada})}(\hat{\beta}_1^2 S_x)$:

Fuente de variación	**Grados de libertad**	**Suma de cuadrados**	**Cuadrados medios**	**$F_{(\text{Calculada de Fisher})}$**
Total	$n-1$	$S_{y\ (\text{Variación tot}}$		
Regresión	1	$\hat{\beta}_1 S_{xy}$	$CM_{(\text{Regresión})} = \left(\frac{\hat{\beta}_1 S_{xy}}{1}\right)$	$\left(\frac{CM_{(\text{Regresión})}}{CM_{(\text{Error})}}\right)$
Error	$n-2$	$SCR_{(\text{Residuos})} = S_y - \hat{\beta}_1 S_{xy}$	$CM_{(\text{Error})} = \left(\frac{SCR_{(\text{Residuos})}}{n-2}\right)$	

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010) y (Hurtado & Gómez, 2010)

El hecho de no rechazar $H_0: \beta_1 = 0$ *puede implicar que la mejor predicción para todas las observaciones es* $\bar{y}$, pues la variable x no influye y la regresión es inútil; aunque, la relación podría no ser del tipo recta. No aceptar $H_0: \beta_1 = 0$ *puede implicar que el modelo lineal* $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ \forall\ i = 1,2,3,4,\dots,n$ donde $E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2 \ \forall\ i = 1,2,3,4,\dots,n$ es o no adecuado.

Sin embargo, es importante no confundir la significación de regresión con una prueba de causalidad. Los modelos de regresión únicamente cuantifican la relación lineal entre variables endógena o respuesta y exógenas o explicativas, pero no justifican que éstas sean causa de otras.

Finalmente, la adecuación de modelo $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \; \forall \, i = 1,2,3,4, \dots, n$ donde $E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2 \; \forall \, i = 1,2,3,4, \dots, n$ como de normalidad se estudiarán mediante análisis de residuos.

4.5. Estimadores puntuales

Los estimadores puntuales β_0, β_1 y σ^2, intervalos de confianza para estos parámetros se puede proporcionar con base en distribuciones. Su ancho estará en función de la calidad de recta de regresión. Con la hipótesis de normalidad y según distribuciones de $\hat{\beta}_0$ y $\hat{\beta}_1$, un intervalo de confianza para pendiente β_1 con nivel de confianza $100 * (1 - \alpha)\%$ es:

$$\hat{\beta}_1 \pm t_{(Student);(n-2)}(\alpha) * \left(\frac{\hat{\sigma}^2}{s_x}\right)^{1/2}$$

Donde $t_{(Student);(n-2)}(\alpha)$ es tal que $P\left[\left|t_{(Student);(n-2)}\right| < t_{(Student);(n-2)}(\alpha)\right] = 1 - \alpha$. Análogamente, para β_0 es:

$$\hat{\beta}_0 \pm t_{(Student);(n-2)}(\alpha) * \left(\hat{\sigma}^2 * \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right)\right)^{1/2}$$

Las cantidades:

$$ee(\hat{\beta}_0) = \left(\hat{\sigma}^2 * \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right)\right)^{1/2} \text{ y } ee(\hat{\beta}_1) = \left(\frac{\hat{\sigma}^2}{s_x}\right)^{1/2}$$

Son "errores estándar" de intercepción $\hat{\beta}_0$ y pendiente $\hat{\beta}_1$, respectivamente. Trata de estimaciones de desviación típica de estimadores. Son medidas de precisión de estimación de parámetros. Se sabe:

$$\hat{\sigma}^2 = \left(\frac{SCR_{(Residuos)}}{n-2}\right) = \left(\frac{1}{n-2}\right)\left(S_{y\,(Variación\ total)}\left(1 - r^2_{(Coeficiente\ de\ correlación\ muestral)}\right)\right)$$

Es estimador insesgado de σ^2 y distribución $\left(\frac{SCR_{(Residuos)}}{\sigma^2}\right) \sim \chi^2_{(n-2)}$ tal que el intervalo de confianza $100 * (1-\alpha)\%$ es:

$$\left(\frac{SCR_{(Residuos)}}{\chi^2_{(n-2)}\left(\frac{\alpha}{2}\right)}\right) \leq \sigma^2 \leq \left(\frac{SCR_{(Residuos)}}{\chi^2_{(n-2)}\left(1-\frac{\alpha}{2}\right)}\right)$$

Donde $\left(\chi^2_{(n-2)}\left(\frac{\alpha}{2}\right)\right)$ y $\left(\chi^2_{(n-2)}\left(1-\frac{\alpha}{2}\right)\right)$ son valores de $\chi^2_{(n-2)}$ para que suma de probabilidades de colas sea α.

Los estimadores puntuales β_0, β_1 y σ^2, intervalos de confianza para estos parámetros se puede proporcionar con base en distribuciones. Su ancho estará en función de la calidad de recta de regresión.

Con la hipótesis de normalidad y según distribuciones de $\hat{\beta}_0$ y $\hat{\beta}_1$, un intervalo de confianza para pendiente β_1 con nivel de confianza $100 * (1-\alpha)\%$ es:

$$\hat{\beta}_1 \pm t_{(Student);(n-2)}(\alpha) * \left(\frac{\hat{\sigma}^2}{s_x}\right)^{1/2}$$

Donde $t_{(Student);(n-2)}(\alpha)$ es tal que $P\left[\left|t_{(Student);(n-2)}\right| < t_{(Student);(n-2)}(\alpha)\right] = 1-\alpha$. Análogamente, para β_0 es:

$$\hat{\beta}_0 \pm t_{(Student);(n-2)}(\alpha) * \left(\hat{\sigma}^2 * \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right)\right)^{1/2}$$

Las cantidades:

$$ee(\hat{\beta}_0) = \left(\hat{\sigma}^2 * \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x}\right)\right)^{1/2} \text{ y } ee(\hat{\beta}_1) = \left(\frac{\hat{\sigma}^2}{s_x}\right)^{1/2}$$

Son "errores estándar" de intercepción $\hat{\beta}_0$ y pendiente $\hat{\beta}_1$, respectivamente. Trata de estimaciones de desviación típica de estimadores. Son medidas de precisión de estimación de parámetros. Se sabe:

$$\hat{\sigma}^2 = \left(\frac{SCR_{(Residuos)}}{n-2}\right) = \left(\frac{1}{n-2}\right)\left(S_{y\,(Variación\ total)}\left(1 - r^2_{(Coeficiente\ de\ correlación\ muestral)}\right)\right)$$

Es estimador insesgado de σ^2 y distribución $\left(\frac{SCR_{(Residuos)}}{\sigma^2}\right) \sim \chi^2_{(n-2)}$ tal que el intervalo de confianza $100 * (1-\alpha)\%$ es:

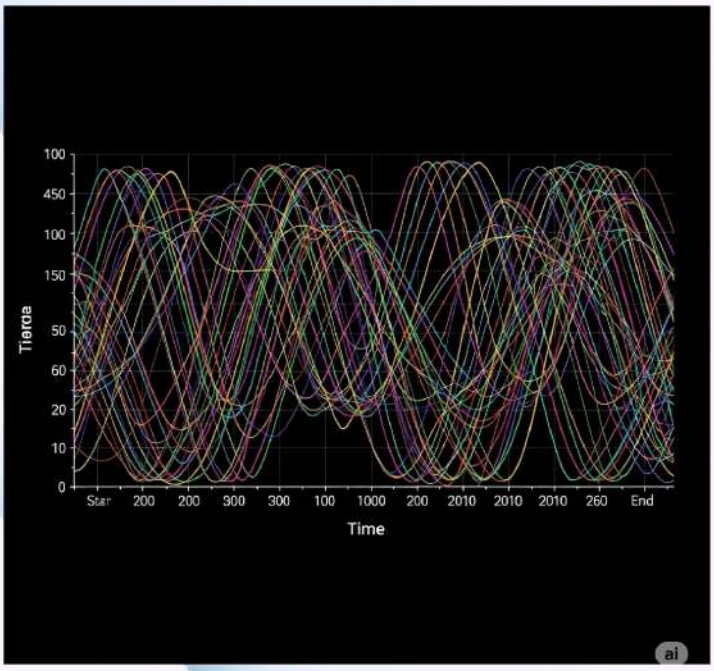
$$\left(\frac{SCR_{(Residuos)}}{\chi^2_{(n-2)}\left(\frac{\alpha}{2}\right)}\right) \leq \sigma^2 \leq \left(\frac{SCR_{(Residuos)}}{\chi^2_{(n-2)}\left(1-\frac{\alpha}{2}\right)}\right)$$

Donde $\left(\chi^2_{(n-2)}\left(\frac{\alpha}{2}\right)\right)$ y $\left(\chi^2_{(n-2)}\left(1-\frac{\alpha}{2}\right)\right)$ son valores de $\chi^2_{(n-2)}$ para que suma de probabilidades de colas sea α.

EDITORIAL ANDES COGNITIO

CAPÍTULO V

SERIES TEMPORALES MODELOS AUTORREGRESIVOS Y PREDICCIÓN ECONÓMICA



CAPÍTULO V

SERIES TEMPORALES, MODELOS AUTORREGRESIVOS Y PREDICCIÓN ECONÓMICA

5. Uso de modelo para estimación y predicción

Para estimar una proyección de punto de Y se utiliza la recta de ajuste:

$$Y_i = b_0 + b_1 X_i$$

valorada para un X_0 determinado. Estos son los valores predichos cuando los X_0 se encuentran dentro del rango de los X observados. También se puede estimar un valor de Y que esté fuera del rango de las observaciones. Si se trata de una serie de tiempo, entonces el estimador tendrá la categoría de pronóstico.

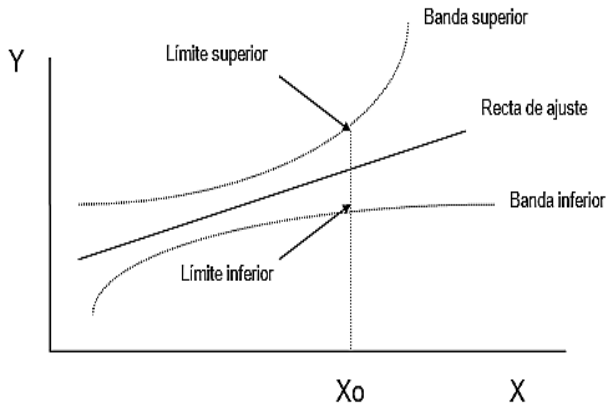
A menudo es de interés no sólo obtener una predicción de punto sino que disponer de un intervalo de confianza para el verdadero valor futuro. Este intervalo de confianza se construye sumándole y restándole a la estimación de punto una cantidad que resulta de multiplicar el valor de t para un α determinado por el error estándar de la proyección de punto.

Dado que el error estándar de la proyección de punto se incrementa a medida que X_0 se aleja de la media de X (ver expresión de la varianza de Y para un X_0 dado), entonces esta cantidad que se suma y se resta a la proyección de punto se irá incrementando, haciendo que la longitud del intervalo sea cada vez más grande.

Por esta razón, aun cuando el modelo haya proporcionado un muy buen ajuste, los intervalos de confianza se van abriendo rápidamente a medida que la proyección se aleja de la media de los datos.

Las curvas que marcan los límites superiores e inferiores de los intervalos de confianza se denominan bandas de confianza y se pueden apreciar en la siguiente figura:

Figura 5.1. Intervalos de confianza



Fuente: (Gujarati & Porter, 2010)

5.1. Modelos de regresión es estimación de respuesta media

Uno de los principales usos de modelos de regresión es estimación de respuesta media $E[Y|x_0]$ para un valor particular x_0 de variable regresora. Asuma que x_0 es un valor dentro del recorrido de datos originales x. Un estimador puntual insesgado de $E[Y|x_0]$ se obtiene con predicción:

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i = \bar{y} + \hat{\beta}_1(x_0 - \bar{x})$$

Se puede interpretar $\beta_0 + \beta_1 x_0$ como función paramétrica estimable:

$$\beta_0 + \beta_1 x_0 = (1, x_0)\beta = x_0'\beta$$

Cuyo estimador es $\hat{y}_0 = x_0'\beta$ tal que:

$$\mathrm{Var}(x_0'\beta) = \sigma^2 x_0' x_0 (X'X)^{-1}$$

Error estándar de $x_0'\beta$ es:

$$ee(x_0'\hat{\beta}) = \left[\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x}\right)\right]^{1/2}$$

El intervalo de confianza para respuesta media $E[Y|x_0]$ es:

$$\hat{y}_0 \pm t_{(Student);[n-2]}(\alpha) * \hat{\sigma}\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x}\right)}$$

Intuitivamente razonable, el ancho de intervalo depende de x_0 y es mínimo para $x_0 = \bar{x}$ tal que crece cuando $|x_0 - \bar{x}|$ aumenta.

5.1.1. Intervalos de confianza

Los intervalos de confianza se dan en forma conjunto para dos parámetros β_0, β_1 de regresión lineal simple. La confianza conjunta de ambos intervalos no es $100 * (1 - \alpha)\%$; aunque, los dos se construyan para verificar el nivel de confianza. Si el nivel de confianza conjunta es $100 * (1 - \alpha)\%$ se construye una región de confianza o, también, llamados "intervalos de confianza simultáneos".

Con base en distribución de $F_{(F\text{ de Fisher})} = \left[\frac{(A\hat{\beta}-c)'(A(X'X)-A')^{-1}\left(\frac{A\hat{\beta}-c}{q}\right)}{\left(\frac{SCR_{(Residuos)}}{n-r}\right)}\right]$ se sabe:

$$F_{(F\text{ de Fisher})} = \left[\frac{(A\hat{\beta} - c)'(A(X'X) - A')^{-1}\left(\frac{A\hat{\beta} - c}{q}\right)}{\left(\frac{SCR_{(Residuos)}}{n - r}\right)}\right] \sim F_{(F\text{ de Fisher});(q.n-r)}$$

Donde $A\hat{\beta} = I\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ y $q = 2$ tal que:

$$\frac{(\hat{\beta}-\beta)^{'}\left(X^{'}X(\hat{\beta}-\beta)\right)}{2*CM_{(Error)}}\sim F_{(F\ de\ Fisher);(2.n-2)}$$

$$X^{'}X=\begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

Con esta distribución se construye una región de confianza $100*(1-\alpha)\%$ para β_0,β_1 de forma conjunta dada por elipse:

$$\frac{(\hat{\beta}-\beta)^{'}\left(X^{'}X(\hat{\beta}-\beta)\right)}{2*CM_{(Error)}}\sim F_{(F\ de\ Fisher);(2.n-2)}(\alpha)$$

Se usan diversos métodos de obtención de intervalos simultáneos de tipo:

$$\hat{\beta}_j\pm\Delta*ee(\hat{\beta}_j)\ j=0,1$$

Método de Scheffé

El método de Scheffé es un ejemplo que estima intervalos simultáneos:

$$\hat{\beta}_j\pm\left(2F_{(F\ de\ Fisher);(2.n-2)}(\alpha)\right)^{1/2}*ee(\hat{\beta}_j)\ j=0,1$$

Una aplicación importante de modelos de regresión es la predicción de nuevas observaciones para un valor x_0 de variable regresora. El intervalo definido anteriormente es adecuado para el valor esperado de respuesta tal que un intervalo de predicción para una respuesta individual concreta se requiere. Esto intervalos reciben el nombre de "intervalo de predicción" en vez de "intervalos de confianza", pues se reserva el nombre de intervalo de confianza para los que se construyen como estimación de un parámetro.

Los intervalos de predicción tienen en cuenta la variabilidad en predicción de valor medio y probabilidad al exigir una respuesta individual tal que si x_0 es valor de interés, entonces:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Es estimador puntual de un nuevo valor de respuesta $Y_0 = Y|x_0$. Si considera la obtención de un intervalo de confianza para observación futura Y_0, el intervalo de confianza para respuesta media en $x = x_0$ es inapropiado, pues es un intervalo sobre media de $Y_{0\,(\text{Parámetro})}$, no sobre observaciones futuras de distribución. No obstante, un intervalo de predicción para una respuesta concreta de $Y_{0\,(\text{Parámetro})}$ se puede hallar considerando variable aleatoria $Y_{0\,(\text{Parámetro})} - \hat{y}_0 \sim N\left(0, \text{Var}\left(Y_{0\,(\text{Parámetro})} - \hat{y}_0\right)\right)$ tal que:

$$\text{Var}\left(Y_{0\,(\text{Parámetro})} - \hat{y}_0\right) = \sigma^2 + \sigma^2 * \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x}\right)$$

Pues $Y_{0\,(\text{Parámetro})}$ que es observación futura es independiente de $\hat{y}_0$ tal que si se usa valor muestral $\hat{y}_0$ para predecir $Y_{0\,(\text{Parámetro})}$ se obtiene un intervalo de predicción $100 * (1 - \alpha)\%$ para $Y_{0\,(\text{Parámetro})}$:

$$\hat{y}_0 \pm t_{(\text{Student});(n-2)}(\alpha) * \hat{\sigma}\sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x}\right)}$$

Este resultado puede generalizarse en un intervalo de predicción $100 * (1 - \alpha)\%$ para media de k futuras observaciones de variable respuesta cuando $x = x_0$. Si $\bar{y}_0$ es media de k futuras observaciones para $x = x_0$, un estimador de $\bar{y}_0$ es $\hat{y}_0$ tal que el intervalo es:

$$\hat{y}_0 \pm t_{(\text{Student});(n-2)}(\alpha) * \hat{\sigma}\sqrt{\left(\frac{1}{k}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_x}\right)}$$

5.2. Microsoft Excel

No es un paquete estadístico, pero dispone de funciones y macros disponibles en internet para pocos datos y análisis sencillos, sin necesidad de recurrir a otros más complejos y costosos. Se recomienda para análisis puntuales (Guillén, 2013).

Según CJF (2010), Microsoft Excel es una hoja de cálculo que está organizada en una estructura tabular con filas y columnas que permite crear tablas, calcular y analizar datos. Excel permite crear tablas que estiman de forma automática los totales de valores numéricos específicos, imprimir tablas con diseños organizados y creas gráficos simples.

Asimismo, Microsoft Excel está compuesta por 16,384 columnas y 1´048,576 filas que forman una cuadrícula. A la intersección entre columna y fila se denomina "celda". Según EduTecno (2010), dispone de varias hojas consecutivas.

Se considera el modelo general de M ecuaciones con M variables endógenas o conjuntamente dependientes se presenta en ecuación (Gujarati y Porter, 2010):

$$\begin{aligned} Y_{1t} &= \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \cdots + \beta_{1M}Y_{Mt} \\ &\quad + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \cdots + \gamma_{1K}X_{Kt} + \upsilon_{1t} \\ Y_{2t} &= \beta_{21}Y_{1t} + \beta_{23}Y_{3t} + \cdots + \beta_{2M}Y_{Mt} \\ &\quad + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \cdots + \gamma_{2K}X_{Kt} + \upsilon_{2t} \end{aligned}$$

$$
\begin{array}{llllllllll}
Y_{3t} & & & & & & & & & \\
= & \beta_{31}Y_{1t} & +\beta_{32}Y_{2t} & & +\cdots+ & \beta_{3M}Y_{Mt} & & & & \\
 & & & & & +\gamma_{31}X_{1t} & +\gamma_{32}X_{2t} & +\cdots+ & +\gamma_{3K}X_{Kt} & +\upsilon_{3t} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
Y_{Mt} & & & & & & & & & \\
= & \beta_{M1}Y_{1t} & +\beta_{M2}Y_{2t} & +\cdots+ & \beta_{M,M1}Y_{M-1} & & & & & \\
 & & & & & +\gamma_{M1}X_{1t} & +\gamma_{M2}X_{2t} & +\cdots+ & +\gamma_{MK}X_{Kt} & +\upsilon_{Mt}
\end{array}
$$

Pueden adoptarse dos enfoques que estiman ecuaciones estructurales:

- **Métodos uni ecuacionales o información limitada**

Cada ecuación en sistema, ecuaciones simultáneas, se estima individualmente, considerando las restricciones impuestas sobre ella, como exclusión de algunas variables, sin preocuparse de restricciones sobre las otras ecuaciones en sistema tal que de ahí "métodos de información limitada".

- **Métodos de sistemas o información completa**

En métodos de sistemas, se estiman todas las ecuaciones en modelo tal que, teniendo en cuenta restricciones ocasionadas por omisión o ausencia de algunas variables, sobre dichas ecuaciones, son esenciales para su identificación, tal que el nombre "métodos de información completa". Por ejemplo:

$$
\begin{array}{llllllll}
Y_{1t} & \beta_{10} & & +\beta_{12}Y_{2t} & +\beta_{13}Y_{3t} & & +\gamma_{11}X_{1t} & & +\upsilon_{1t} \\
= & + & & & + & & + & & \\
Y_{2t} & \beta_{20} & & & +\beta_{23}Y_{3t} & & +\gamma_{21}X_{1t} & +\gamma_{22}X_{2t} & +\upsilon_{2t} \\
= & + & & & + & & + & + & \\
Y_{3t} & \beta_{30} & \beta_{31} & & & +\beta_{34}Y_{4t} & +\gamma_{31}X_{1t} & +\gamma_{32}X_{2t} & +\upsilon_{3t} \\
= & + & + Y_{1t} & & & + & + & + & \\
 & & + & & & & & &
\end{array}
$$

$$Y_{4t} = \beta_{40} + \beta_{42} + Y_{2t} + \gamma_{43} X_{3t} + \upsilon_{4t}$$

Tal que Y son variables endógenas o explicadas y X sonvariables explicativas o exógenas. Por ejemplo: si está interesado en estimar la tercera ecuación, los métodos uni ecuaciones consideran solamente esta ecuación, observando que variables Y_2 y X_3 están excluidas.

Caso contrario, en métodos de sistemas se trata de estimar cuatro ecuaciones simultáneas teniendo en cuenta todas las restricciones impuestas sobre diversas ecuaciones del sistema. Idealmente, el método de sistemas debería usarse tal como "método de máxima verosimilitud con información completa (MVIC)"; aunque, en la práctica estos métodos no son de uso frecuente por múltiples razones:

1. La carga computacional es basta; por ejemplo: modelo de comparativamente pequeño de 20 ecuaciones de Klein-Goldberger de economía de USA para 1955 tenía 151 coeficientes diferentes de 0, de los cuales autores estimaron sólo 51 usando información de series de tiempo,
2. El modelo econométrico del Brookings Social Science Research Council (SSRC) para economía estadounidense, publicado en 1965, tenía inicialmente 150 ecuaciones,
3. Modelos tan elaborados pueden proporcionar detalles complejos de diversos sectores de economía, los cálculos representan un enorme esfuerzo aun en fechas actuales con computadores de alta velocidad, sin mencionar el costo involucrado,
4. Métodos de sistemas, como MVIC, conducen a soluciones altamente no lineales en parámetros y, por ende, difíciles de determinar,
5. Si existe un error de especificación, como forma funcional equivocada o exclusión de variables relevantes, en una o más ecuaciones del sistema, a

este error es transmitido al resto del sistema y, por consiguiente, los métodos de sistemas son muy sensibles a errores de especificación.

Entonces, los métodos uni ecuacionales son usados con más frecuencia en la práctica, pues son menos sensibles a errores de especificación en sentido que aquellas partes del sistema que tienen una especificación correcta pueden no verse afectadas considerablemente por errores de especificación en otra parte. Por lo tanto, los métodos uni ecuacionales abordan:

5.2.1. Mínimos cuadrados ordinarios (MCO)

Debido a interdependencia entre término de perturbación estocástico y variable(s) explicativa(s) o endógena(s), este método es inapropiado para estimación de una ecuación en un sistema de ecuaciones simultáneas, pues si se aplica erróneamente, los estimadores no sólo resultan sesgados, muestras pequeñas, sino inconsistentes; es decir, sin importar qué tan gran es el tamaño de la muestra, el sesgo no desaparece; aunque, existe una situación en que este método puede ser aplicado apropiadamente, aun en contexto de ecuaciones simultáneas, llamado "modelos recursivos, triangulares o causales". Por ejemplo: para ver su naturaleza, considere el sistema de tres ecuaciones:

$$
\begin{aligned}
Y_{1t} &= \beta_{10} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \upsilon_{1t} \\
Y_{2t} &= \beta_{20} + \beta_{20}Y_{1t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \upsilon_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \upsilon_{3t}
\end{aligned}
$$

Donde Y y X son variables endógenas o explicadas y exógenas o explicativas, respectivamente. Las perturbaciones son tal que $\mathrm{Cov}(\upsilon_{1t}, \upsilon_{2t}) = \mathrm{Cov}(\upsilon_{1t}, \upsilon_{3t}) = \mathrm{Cov}(\upsilon_{2t}, \upsilon_{3t}) = 0$; es decir, las perturbaciones de diferentes ecuaciones en mismo

periodo no están correlacionadas, técnicamente es cero correlación contemporánea. Considere ecuación:

$$Y_{1t} = \beta_{10} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \upsilon_{1t}$$

Contiene variables exógenas o explicativas de lado derecho y, por supuestos, no están correlacionadas con término de perturbación υ_{1t}, esta ecuación satisface el supuesto crítico del método MCO clásico tal que la no correlación entre variables explicativas y perturbaciones estocásticas. Por tanto, MCO puede aplicarse directamente a esta ecuación. Después, considere la segunda ecuación:

$$Y_{2t} = \beta_{20} + \beta_{20}Y_{1t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \upsilon_{2t}$$

Contienevariable endógena Y_{1t} como variable explicativa o exógena junto con X no estocásticas tal que es así, pues υ_{1t}, que afecta a Y_{1t}, por supuestos y no está correlacionada con υ_{2t} tal que para todos los efectos prácticos Y_{1t} es variable predeterminada respecto a Y_{2t}. Se procede a estimación de esta ecuación por MCO, así como a ecuación:

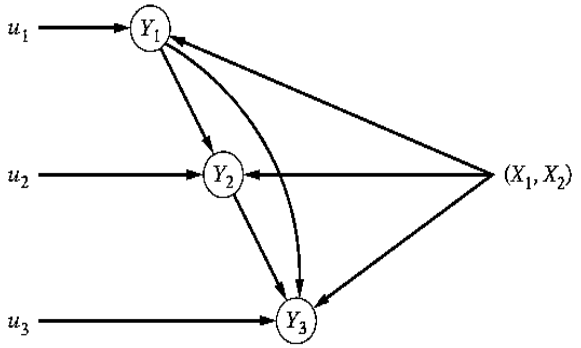
$$Y_{3t} = \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \upsilon_{3t}$$

Esto es porque Y_{1t} y Y_{2t} no están correlacionadas con υ_{3t}. En sistema recursivo puede aplicarse MCO a cada ecuación separadamente tal que no se tiene problema de ecuaciones simultáneas en esta situación. Es claro que no existe interdependencia entre variables endógenas o explicadas.

Entonces, Y_{1t} afecta a Y_{2t}, pero no al revés. En forma semejante, Y_{1t}, Y_{2t} influyen en Y_{3t} sin que esta última influya en las primeras; es decir, cada ecuación presenta una dependencia unilateral tal que de aquí el nombre “modelos causales” o

"triangular", desprendido de que si se forma la matriz de coeficientes de variables endógenas dadas:

Figura 5.1. Modelos causales



Fuente: (Gujarati & Porter, 2010)

$$Y_{1t} = \beta_{10} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \upsilon_{1t}$$

$$Y_{2t} = \beta_{20} + \beta_{20}Y_{1t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \upsilon_{2t}$$

$$Y_{3t} = \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \upsilon_{3t}$$

Se obtiene la matriz triangular:

	Y_1	Y_2	Y_3
Ecuación 1	1	0	0
Ecuación 2	β_{21}	1	0
Ecuación 3	β_{31}	β_{32}	1

Por ejemplo: un sistema recursivo, puede postularse el modelo de determinación de salarios y precios siguiente:

$$\text{Precios: } P_t = \beta_{10} + \beta_{11}W_{t-1} + \beta_{12}R_t(1) + \beta_{12}M_t(2) + \beta_{14}L_t(3) + \upsilon_{1t}$$

$$\text{Salarios: } W_t(4) = \beta_{20} + \beta_{21}UN_t(5) + \beta_{32}P_t(6) + \upsilon_{2t}$$

La ecuación de precios postula que tasa de cambio de precios en periodo actual es una función de las tasas de cambio en los precios del capital y de las materias primas, de la tasa de cambio en la productividad laboral y de la tasa de cambio en los salarios en el periodo anterior. La ecuación de salarios muestra que la tasa de cambio en los salarios en el periodo actual está determinada por la tasa de cambio de los precios en el periodo actual y por la tasa de desempleo.

La cadena causal va de $W_{(t-1)} \Rightarrow P_t \Rightarrow W_{(t)}$ tal que puede aplicarse MCO para estimar parámetros de dos ecuaciones individualmente. Aunque los modelos recursivos han demostrado ser útiles, la mayor parte de los modelos de ecuaciones simultáneas no presentan tal relación unilateral de causa y efecto. Por consiguiente, MCO, en general, resulta inapropiado para estimar una sola ecuación en el contexto de un modelo de ecuaciones simultáneas.

---

[1] Derivada respecto al tiempo de tasa de cambio del precio de capital

[2] Derivada respecto al tiempo de tasa de cambio de precios de importación

[3] Derivada respecto al tiempo de tasa de cambio de productividad laboral

[4] Derivada respecto al tiempo de tasa de cambio de salarios por empleado

[5] Tasa de desempleo en %

[6] Derivada respecto al tiempo de tasa de cambio del precio por unidad de producción

Autores sostiene que MCO generalmente es inaplicable a modelos de ecuaciones simultáneas, se puede usar solamente como estándar o norma de comparación; es decir, se puede estimar una ecuación estructural mediante MCO, con propiedades resultantes de sesgo, inconsistencia, etcétera tal que la misma ecuación puede ser estimada por otros métodos especialmente diseñados para manejar el problema de simultaneidad y resultados de dos métodos pueden comprarse, por lo menos, de manera cualitativa.

En muchas aplicaciones resultados de MCO aplicado de forma inapropiada pueden no diferir mucho de aquellos obtenidos por métodos más complejos. En principio, no debe haber mucha objeción en la presentación de resultados basados en MCO, siempre y cuando las estimaciones hechas con base en métodos alternos diseñados para modelos de ecuaciones simultáneas también sean proporcionadas.

Este método podría dar alguna idea de qué tan malas son las estimaciones de MCO en situaciones en que es aplicado inapropiadamente, teniendo presente que perturbaciones entre ecuaciones no están correlacionadas contemporáneamente. Si no es así, puede ser que se deba recurrir a la técnica de estimación SURE (regresiones aparentemente no relacionadas) de Zellner para estimar los parámetros del sistema recursivo tal que este método puede dar una idea qué tan malas son las estimaciones de MCO en situaciones en que es aplicado inapropiadamente suponiendo que en muestras pequeñas, los estimadores alternativos, al igual que los estimadores por MCO, son sesgados; aunque, el estimador de MCO tiene la "ventaja" sólo para muestras pequeñas de tener varianza mínima entre estos estimadores alternativos.

5.2.2. Mínimos cuadrados indirectos (MCI)

Para una ecuación estructural precisa o exactamente identificada, el método para obtener las estimaciones de los coeficientes estructurales a partir de las estimaciones por MCO de los coeficientes en forma reducida se conoce como

método de mínimos cuadrados indirectos (MCI) y las estimaciones así obtenidas se conocen como estimaciones de mínimos cuadrados indirectos. Pasos de estimación:

a) **Se obtienen primero las ecuaciones en forma reducida**. Éstas se obtienen de las ecuaciones estructurales en forma tal que la variable dependiente en cada ecuación es la única variable endógena y está en función únicamente de las variables predeterminadas (exógenas o endógenas rezagadas) y del (los) término(s) de error(es) estocástico(s).
b) **Se aplica MCO individualmente a ecuaciones en forma reducida**. Esta operación es permisible puesto que variables explicativas en estas ecuaciones están predeterminadas y, por tanto, no están correlacionadas con las perturbaciones estocásticas. Las estimaciones así obtenidas son consistentes.
c) **Se obtienen estimaciones de coeficientes estructurales originales a partir de coeficientes en forma reducida estimados**, **obtenidos en el paso 2**. Si una ecuación está exactamente identificada, hay una correspondencia uno a uno entre los coeficientes estructurales y los coeficientes en la forma reducida; es decir, pueden derivarse estimaciones únicas de los primeros a partir de los últimos.

5.2.3. Mínimos cuadrados en dos etapas

Los mínimos cuadrados en dos etapas (MC2E: Estimación de ecuación sobre identificada).

Considere el modelo:

Función ingreso: $Y_{1t\,(Ingre} = \beta_{10} + \beta_{11}Y_{2t\,(Ti} + \gamma_{11}X_{1t\,(Inv} + \gamma_{12}X_{2t\,(B} + u_{1t}$

Función oferta n $Y_{2t\ (\in din}$ $= \beta_{20} + \beta_{21} Y_{1t\ (Ti} + u_{2t}$

La ecuación de ingreso, un híbrido de los enfoques de las teorías cuantitativa y keynesiana de determinación de ingreso establece que ingreso está determinado por oferta monetaria, gasto de inversión y gasto del gobierno. La función de la oferta monetaria postula que existencias de dinero están determinadas, por el Sistema de la Reserva Federal, con base en nivel del ingreso.

Es obvio, se tiene un problema de ecuaciones simultáneas, que puede verificarse mediante la prueba de simultaneidad. Al aplicar la condición de orden para identificación, la ecuación del ingreso está sub identificada y ecuación de oferta monetaria está sobre identificada. Es poco lo que puede hacerse sobre ecuación del ingreso, a no ser que se altere la especificación del modelo.

5.3. Modelo de demanda y oferta

5.3.1. Función de oferta monetaria

La función de oferta monetaria sobre identificada no puede estimarse mediante MCI, estimadores que heredan todas las propiedades asintóticas de estimadores en forma reducida, como consistencia y eficiencia asintótica, pero propiedades de muestras pequeñas, como insesgamiento generalmente no continúan siendo válidas, pues hay dos estimaciones de β_{21}.

Puede aplicarse MCO a ecuación de oferta monetaria, pero las estimaciones obtenidas por este mecanismo serán inconsistentes en vista de la probable correlación entre la variable explicativa estocástica Y_1 y el término de perturbación estocástico u_2. Suponga que se encuentra una "variable representante" para la variable explicativa estocástica Y_1, tal que,

aunque "se parece" a Y_1 en sentido altamente correlacionada con Y_1, no está correlacionada con u_2.

Tal variable se conoce también como variable instrumental. Si se puede encontrar tal variable representante, puede utilizarse MCO directamente para estimar la función de oferta monetaria. Esta variable se obtiene por el método de mínimos cuadrados en dos etapas (MC2E), desarrollado independientemente por Henri Theil y Robert Basmann.

El problema de identificación surge porque diferentes conjuntos de coeficientes estructurales pueden ser compatibles con el mismo conjunto de información; es decir, una ecuación en una forma reducida dada puede ser compatible con diferentes ecuaciones estructurales o con diferentes hipótesis (modelos) y puede ser difícil decir cuál hipótesis (modelo) particular se está investigando.

Este problema pretende establecer si estimaciones numéricas de parámetros de una ecuación estructural pueden obtenerse de coeficientes en forma reducida estimados. Si es así, se dice que la ecuación particular está identificada; si no, la ecuación bajo consideración está no identificada o sub identificada. Puede estar exactamente (o total o precisamente) identificada o sobre identificada.

Se dice que está exactamente identificada si pueden obtenerse valores numéricos únicos de los parámetros estructurales. Se dice que está sobre identificada si puede obtenerse más de un valor numérico para algunos de los parámetros de las ecuaciones estructurales. Las circunstancias bajo las cuales puede ocurrir cada uno de los casos anteriores se indicarán a continuación:

Considere de forma conjunta con condición de mercado nivelado o equilibrio en que demanda es igual a oferta, el modelo de demanda y oferta siguiente:

Función de demanda:

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} \quad \alpha_1 < 0$$

Función de oferta:

$$Q^{s\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{2t\,(\text{Tiempo})} \quad \beta_1 < 0$$

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = Q^{s\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})}$$

Con base en esto y mediante condición de equilibrio se obtiene:

$$\alpha_0 + \alpha_1 P_t + \upsilon_{1t} = \beta_0 + \beta_1 P_t + \upsilon_{2t}$$

Si resuelve la ecuación anterior, obtiene el precio de equilibrio:

$$P_{t\,(\text{Precio en tiempo})} = \Pi_0 + \upsilon_t$$

Tal que:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right);\ \upsilon_t = \left(\frac{\upsilon_{2t} - \upsilon_{1t}}{\alpha_1 - \beta_1}\right)$$

Al sustituir $P_{t\,(\text{Precio en tiempo})}$ en:

Función de demanda:

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} \quad \alpha_1 < 0$$

Función de oferta:

$$Q^{S(\text{Cantidad ofertada})}_{t_{(\text{En tiempo})}} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{2t\,(\text{Tiempo})} \quad \beta_1 < 0$$

$$Q^{d(\text{Cantidad demandada})}_{t_{(\text{En tiempo})}} = Q^{S(\text{Cantidad ofertada})}_{t_{(\text{En tiempo})}}$$

Se obtiene:

$$Q_{t\,(\text{Cantidad en tiempo})} = \Pi_1 + w_t$$

Donde:

$$\Pi_1 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right);\ w_t = \left(\frac{\alpha_1\upsilon_{2t} - \beta_1\upsilon_{1t}}{\alpha_1 - \beta_1}\right)$$

Es necesario observar que términos de error, υ_t y w_t, son combinaciones lineales de términos de error originales u_1 y u_2. Ecuaciones:

$$P_{t\,(\text{Precio en tiempo})} = \Pi_0 + \upsilon_t\,;\ Q_{t\,(\text{Cantidad en tiempo})} = \Pi_1 + w_t$$

Son ecuaciones en forma reducida. El modelo de demanda y oferta contiene cuatro coeficientes $\alpha_0, \alpha_1, \beta_0$ y β_1, pero no existe una forma única de estimarlos debido a dos coeficientes en forma reducida de ecuaciones:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right);\ \Pi_1 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right)$$

Contienen los cuatro parámetros estructurales, pero no hay forma de estimar las cuatro incógnitas estructurales a partir únicamente de dos coeficientes en forma reducida, pues para estimar cuatro incógnitas se deben tener cuatro ecuaciones independientes y, en general, para estimar k incógnitas se deben tener k ecuaciones, independientes.

Existe una forma alterna y, probablemente, más ilustrativa de considerar el problema de identificación. Multiplique la ecuación siguiente por $\lambda(0 \leq \lambda \leq 1)$:

$$\lambda * Q^{d(\text{Cantidad demandada})}_{t(\text{En tiempo})} = \lambda * \alpha_0 + \lambda\alpha_1 P_{t\ (\text{Precio en tiempo})} + \lambda * \upsilon_{1t\ (\text{Tiempo})}$$

Igual, multiplica siguiente ecuación por $(1 - \lambda)$ y obtiene la ecuación:

$$(1-\lambda) * Q^{s(\text{Cantidad ofertada})}_{t(\text{En tiempo})} = (1-\lambda) * \beta_0 + (1-\lambda) * \beta_1 P_{t\ (\text{Precio en tiempo})} + (1-\lambda) * \upsilon_{2t\ (\text{Tiempo})}$$

Al sumar ambas ecuaciones, se obtiene la combinación lineal de ecuaciones originales:

$$Q_{t\ (\text{Cantidad en tiempo})} = \gamma_o + \gamma_1 P_{t\ (\text{Precio en tiempo})} + w_t$$

Tal que:

$$\gamma_o = \lambda * \alpha_o + (1-\lambda) * \beta_0$$

$$\gamma_1 = \lambda * \alpha_1 + (1-\lambda) * \beta_1$$

$$w_t = \lambda * u_{1t} + (1-\lambda) * u_{2t}$$

La ecuación $Q_{t\ (\text{Cantidad en tiempo})} = \gamma_o + \gamma_1 P_{t\ (\text{Precio en tiempo})} + w_t$ "falsa" o "híbrida", a partir de observación, no se distingue de ecuación $Q^{d(\text{Cantidad demandada})}_{t(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}$ ni $Q^{s(\text{Cantidad ofertada})}_{t(\text{En tiempo})} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}$, pues se consideran regresiones de Q y P tal que si se tiene información de series de tiempo sobre Q y P solamente, cualquier de ecuaciones anteriores o

$Q_{t\,(\text{Cantidad en tiempo})} = \gamma_o + \gamma_1 P_{t\,(\text{Precio en tiempo})} + w_t$ tal que no existe forma de concluir cuál de estas se está verificando.

Para que una ecuación esté identificada; es decir, que sus parámetros sean estimados, debe mostrarse que el conjunto dado de información no producirá una ecuación estructural que sea similar en apariencia a ecuación en que se está interesado. Si pretende estimar la función de demanda se debe demostrar que información dada no es consistente con función de oferta u otro tipo de ecuación híbrida.

La razón por que no fue posible identificar anteriores funciones de demanda u oferta es porque variables Q y P están presentes en ambas funciones y no dispone de información adicional. Sin embargo, suponga siguiente modelo de demanda y oferta:

Función de demanda:

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \alpha_2 I_{t\,(\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} \quad \alpha_1 < 0, \alpha_2 > 0$$

Función de oferta:

$$Q^{S\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{2t\,(\text{Tiempo})} \quad \beta_1 > 0$$

La diferencia entre este modelo y original de demanda-oferta es que existe una variable adicional en función de demanda; es decir, el ingreso de consumidor en el tiempo. Con base en teoría económica de demanda se sabe que ingreso es, en general, un determinante importante de demanda de mayoría de bienes o servicio.

En consecuencia, su inclusión es función de demanda proporcionará información adicional sobre comportamiento del consumidor.

En caso de mayoría de bienes se espera que ingreso de consumidor en tiempo tenga efecto positivo respecto a consumo $(\alpha_2 > 0)$. Si usa mecanismo de nivelación de mercado y $Q^{d(\text{Cantidad demandada})}_{t_{(\text{En tiempo})}} = Q^{s(\text{Cantidad ofertada})}_{t_{(\text{En tiempo})}}$ se obtiene:

$$\alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}$$

Si resuelve ecuación anterior, se obtiene el valor de equilibrio de $P_{t\ (\text{Precio en tiempo})}$:

$$P_{t\ (\text{Precio en tiempo})} = \Pi_0 + \Pi_1 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_t$$

Con base en esto, coeficientes reducidos son:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right); \Pi_1 = -\left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right);\ \nu_t = \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)$$

Si sustituye el valor de equilibrio $P_{t\ (\text{Precio en tiempo})}$ en función de demanda u oferta anterior, se obtiene la cantidad de equilibrio:

$$Q_{t_{(\text{En tiempo})}} = \Pi_2 + \Pi_3 I_{t\ (\text{Ingreso de consumidor en tiempo})} + w_t$$

Tal que:

$$\Pi_2 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right); \Pi_3 = -\left(\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right);\ w_t = \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$$

Las ecuaciones $P_{t\ (\text{Precio en tiempo})} = \Pi_0 + \Pi_1 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_t$ y $Q_{t_{(\text{En tiempo})}} = \Pi_2 + \Pi_3 I_{t\ (\text{Ingreso de consumidor en tiempo})} + w_t$ son ecuaciones en forma reducida, puede aplicarse método MCO para estimar parámetros.

El modelo de demanda y oferta:

Función de demanda:

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \alpha_2 I_{t\,(\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} \quad \alpha_1 < 0, \alpha_2 > 0$$

Función de oferta:

$$Q^{S\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \upsilon_{2t\,(\text{Tiempo})} \quad \beta_1 > 0$$

Estas ecuaciones contiene cinco coeficientes estructurales, $\alpha_0, \alpha_1, \alpha_2, \beta_0$ y β_1, pero sólo dispone de cuatro ecuaciones para estimarlos tal que los cuatro coeficientes en forma reducida son Π_0, Π_1, Π_2 y Π_3 dados en ecuaciones $\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right); \Pi_0 = -\left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right)$; $\nu_t = \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)$ y $\Pi_2 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right); \Pi_3 = -\left(\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right)$; $w_t = \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$.

Por lo tanto, no es posible encontrar una solución única para todos los coeficientes estructurales. No obstante, puede mostrarse con facilidad que los parámetros de función de oferta pueden ser identificados o estimados:

$$\beta_0 = \Pi_2 - \beta_1\Pi_0;\ \beta_1 = \left(\frac{\Pi_3}{\Pi_1}\right)$$

No hay una forma única de estimar los parámetros de la función de demanda; por consiguiente, ésta permanece sub identificada tal que el coeficiente estructural β_1 es una función no lineal de coeficientes en forma reducida, que crea algunos problemas cuando se trata de estimar el error estándar de β_1 estimado.

Para verificar que función de demanda $Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \alpha_2 I_{t\,(\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\,(\text{Tiempo})}$ $\alpha_1 < 0, \alpha_2 > 0$ no puede ser identificada o estimada, multiplíquese

por $\lambda(0 \leq \lambda \leq 1)$ y $Q_{t_{(\text{En tiempo})}}^{S(\text{Cantidad ofertada})} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}$ $\beta_1 > 0$ por $(1-\lambda)$ y asume para obtener ecuación "híbrida":

$$Q_{t_{(\text{En tiempo})}} = \gamma_0 + \gamma_1 P_{t\ (\text{Precio en tiempo})} + \gamma_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + w_t$$

Donde:

$$\gamma_0 = \lambda\alpha_0 + (1-\lambda)\beta_0;\ \gamma_1 = \lambda\alpha_1 + (1-\lambda)\beta_1; \gamma_2 = \lambda\alpha_2; w_t = \lambda u_{1t} + (1-\lambda)\lambda u_{2t}$$

Ecuación $Q_{t_{(\text{En tiempo})}} = \gamma_0 + \gamma_1 P_{t\ (\text{Precio en tiempo})} + \gamma_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + w_t$ es con base en observación indistinguible de función de demanda $Q_{t_{(\text{En tiempo})}}^{d(\text{Cantidad demandada})} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}$ $\alpha_1 < 0, \alpha_2 > 0$; aunque, si es distinguible de función de oferta $Q_{t_{(\text{En tiempo})}}^{S(\text{Cantidad ofertada})} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}$ $\beta_1 > 0$, que no contiene la variable $I_{t\ (\text{Ingreso de consumidor en tiempo})}$ como una variable explicativa o exógena; por tanto, la función de demanda permanece sin identificar.

La presencia de una variable adicional en función de demanda permite identificar la función oferta. La inclusión de variable ingreso en ecuación de demanda proporciona alguna información adicional sobre variabilidad de función tal que con mucha frecuencia la posibilidad de identificar una ecuación depende de si excluye una o más variables que están incluidas en otras ecuaciones del modelo. Suponga que se considera modelo de demanda y oferta:

Función de demanda:

$$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \alpha_2 I_{t\,(\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} \quad \alpha_1 < 0, \alpha_2 > 0$$

Función de oferta:

$$Q^{s\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \beta_2 P_{t-1\,(\text{Precio rezagado en tiempo})} + \upsilon_{2t\,(\text{Tiempo})} \quad \beta_1 > 0, \beta_2 > 0$$

5.3.2. Función de demanda

La función d demanda permanece igual que antes, pero función de oferta incluye variable explicativa adicional, el precio rezagado en un periodo de tiempo $\left(P_{t-1\,(\text{Precio rezagado en tiempo})}\right)$, entendida como variable predeterminada pues su valor se conoce con el tiempo $\left(t_{(\text{Tiempo})}\right)$.

5.3.3. Función de oferta

La función de oferta postula que cantidad de un bien ofrecido depende de su precio actual y precio de periodo anterior, un modelo frecuentemente usado para explicar oferta de muchos bienes agrícolas. Por mecanismo de nivelación de mercado:

$$\alpha_0 + \alpha_1 P_{t\,(\text{Precio en tiempo})} + \alpha_2 I_{t\,(\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\,(\text{Tiempo})} = \beta_0 + \beta_1 P_{t\,(\text{Precio en tiempo})} + \beta_2 P_{t-1\,(\text{Precio rezagado en tiempo})} + \upsilon_{2t\,(\text{Tiempo})}$$

Si resuelve ecuación obtiene el precio de equilibrio:

$$P_{t\ (\text{Precio en tiempo})} = \Pi_0 + \Pi_1 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \Pi_2 P_{t-1\ (\text{Precio rezagado en tiempo})} + \nu_{t\ (\text{Tiempo})}$$

Tal que:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right);\ \Pi_1 = -\left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right);\ \Pi_2 = \left(\frac{\beta_2}{\alpha_1 - \beta_1}\right);\ \nu_{t\ (\text{Tiempo})} = \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)$$

Si sustituye el precio de equilibrio en ecuación de demanda u oferta se obtiene cantidad de equilibrio:

$$Q_{t_{(\text{En tiempo})}} = \Pi_3 + \Pi_4 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \Pi_5 P_{t-1\ (\text{Precio rezagado en tiempo})} + w_t$$

Los coeficientes en forma reducida son:

$$\Pi_3 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right);\ \Pi_4 = -\left(\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right);\ \Pi_5 = \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\right);\ w_{t\ (\text{Tiempo})} = \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right)$$

El modelo de demanda y oferta dado en ecuaciones $Q^{d_{(\text{Cantidad demandada})}}_{t_{(\text{En tiempo})}} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}$ $\alpha_1 < 0, \alpha_2 > 0$ y $Q^{s_{(\text{Cantidad ofertada})}}_{t_{(\text{En tiempo})}} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \beta_2 P_{t-1\ (\text{Precio rezagado en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}$ $\beta_1 > 0, \beta_2 > 0$ contiene seis coeficientes estructurales, $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ y β_2, existen seis coeficientes en forma reducida, $\Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4$ y Π_5 par estimarlos tal que se tienen seis ecuaciones con seis incógnitas y normalmente es posible obtener estimaciones únicas.

Tanto parámetros de ambas ecuaciones, demanda y oferta, como sistema en su totalidad pueden ser identificados. Para verificar que funciones de demanda y oferta anteriores son identificables, se recurre al mecanismo de multiplicar la ecuación de demanda $Q_{t_{(\text{En tiempo})}}^{d_{(\text{Cantidad demandada})}} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}\ \alpha_1 < 0, \alpha_2 > 0$ por $\lambda(0 \leq \lambda \leq 1)$ y función oferta $Q_{t_{(\text{En tiempo})}}^{S_{(\text{Cantidad ofertada})}} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \beta_2 P_{t-1\ (\text{Precio rezagado en tiempo})} + \upsilon_{2t\ (\text{Tiempo})}\ \beta_1 > 0, \beta_2 > 0$ por $(1 - \lambda)$, sumándolas para obtener una ecuación híbrida. Esta ecuación tendrá variables predeterminadas $I_{t\ (\text{Ingreso de consumidor en tiempo})}$ y $P_{t-1\ (\text{Precio rezagado en tiempo})}$; por tanto, será una ecuación por observación diferente tanto de ecuación demanda como oferta, pues la primera contiene $P_{t-1\ (\text{Precio rezagado en tiempo})}$ y, segunda, $I_{t\ (\text{Ingreso de consumidor en tiempo})}$.

Para ciertos bienes y servicios, el ingreso al igual que la riqueza del consumidor, es un determinante importante de la demanda. Por consiguiente, si se modifica función de demanda $Q_{t_{(\text{En tiempo})}}^{d_{(\text{Cantidad demandada})}} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}\ \alpha_1 < 0, \alpha_2 > 0$ tal que se mantenga función oferta como antes:

Función de demanda:

$$Q_{t_{(\text{En tiempo})}}^{d_{(\text{Cantidad demandada})}} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio en tiempo})} + \alpha_2 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \alpha_3 R_{t\ (\text{Riqueza de consumidor en tiempo})} + \upsilon_{1t\ (\text{Tiempo})}\ \alpha_1 < 0, \alpha_2 > 0$$

Función de oferta:

$$Q^{S(\text{Cantidad ofertada})}_{t_{(\text{En tiempo})}} = \beta_0 + \beta_1 P_{t\ (\text{Precio en tiempo})} + \beta_2 P_{t-1\ (\text{Precio rezagado en tiempo})} + \upsilon_{2t\ (\text{Tiempo})} \quad \beta_1 > 0, \beta_2 > 0$$

Se espera que para la mayoría de los bienes y servicios la riqueza, al igual que el ingreso, tenga un efecto positivo sobre el consumo. Si se iguala demanda con oferta, se obtiene precio y cantidad de equilibrio siguientes:

$$P_{t\ (\text{Precio en tiempo})} = \Pi_0 + \Pi_1 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \Pi_2 R_{t\ (\text{Riqueza de consumidor en tiempo})} + \Pi_3 P_{t-1\ (\text{Precio rezagado en tiempo})} + \nu_{t\ (\text{Tiempo})}$$

$$Q_{t_{(\text{En tiempo})}} = \Pi_4 + \Pi_5 I_{t\ (\text{Ingreso de consumidor en tiempo})} + \Pi_6 R_{t\ (\text{Riqueza de consumidor en tiempo})} + \Pi_7 P_{t-1\ (\text{Precio rezagado en tiempo})} + w_{t\ (\text{Tiempo})}$$

Tal que:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right); \Pi_1 = -\left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right); \Pi_2 = -\left(\frac{\alpha_3}{\alpha_1 - \beta_1}\right); \Pi_4 = \left(\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\right); \Pi_5 = -\left(\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\right); \Pi_6 = -\left(\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1}\right); \Pi_7 = \left(\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\right); w_{t\ (\text{Tiempo})} = \left(\frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}\right); \nu_{t\ (\text{Tiempo})} = \left(\frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\right)$$

Este modelo de demanda y oferta contiene siete coeficientes estructurales, pero existe ocho ecuaciones para estimarlos tal que los ocho coeficientes en forma reducida dados en esta ecuación; es decir, el número de ecuaciones es mayor que el número de incógnitas.

En consecuencia, no es posible obtener una estimación única de todos los parámetros del modelo, que puede demostrarse fácilmente. En la forma reducida de coeficientes anteriores:

$$\beta_1 = \left(\frac{\Pi_6}{\Pi_2}\right) = \left(\frac{\Pi_5}{\Pi_1}\right)$$

Con base en esto, existen dos estimaciones de coeficientes de precios en función de oferta y no hay garantía que estos dos valores o soluciones sean idénticos, pues se observa diferencia entre sub identificación, siendo imposible obtener estimaciones de los parámetros estructurales y sobre identificación, puede haber varias estimaciones de uno o más coeficientes estructurales.

Además, β_1 aparece en denominadores de todos los coeficientes en su forma reducida, la ambigüedad en estimación de β_1 será transmitida a demás estimaciones. Sin embargo, no implica que la sobre identificación necesariamente sea mala. Una ecuación en un modelo de ecuaciones simultáneas puede estar sub identificada o identificada, sea sobre identificada o exactamente identificada.

El modelo como un todo está identificado si cada una de sus ecuaciones también lo está. Para asegurar la identificación, se acude a ecuaciones en forma reducida.

Se considera un método alterno y posiblemente menos laborioso para determinar si una ecuación en un modelo de ecuaciones simultáneas está identificada o no, pues las condiciones de orden y rango de identificación aligeran la labor, proporcionando una rutina sistemática. Para entender condiciones de orden y rango, se introduce la notación:

M =	Número de variables endógenas en modelo
m =	Número de variables endógenas en una ecuación
K =	Número de variables predeterminadas en modelo, incluyendo intercepto
k =	Número de variables predeterminadas en una ecuación dada

5.3.4. Condición de orden para identificación

La condición de orden para identificación se refiere a la matriz o al número de filas y columnas que contiene. Una condición necesaria, pero no suficiente, para identificación, conocida como condición de orden, puede expresarse en dos formas diferentes pero equivalentes, condiciones necesaria y suficiente para la identificación se presentan después: ecuaciones simultáneas, para que una

- En modelo de M ecuaciones simultáneas, ecuación esté identificada debe excluir al menos $M-1$ variables, endógenas y predeterminadas, que aparecen en modelo. Si excluye exactamente $M-1$ variables, la ecuación está exactamente identificada. Si excluye más de $M-1$ variables, estará sobre identificada.
- En un modelo de M ecuaciones simultáneas, para que una ecuación esté identificada, el número de variables predeterminadas excluidas de esa ecuación no debe ser menor que el número de variables endógenas incluidas en la ecuación menos 1, es decir:

$$K-k \geq m-1$$

Si $K-k=m-1$, la ecuación está exactamente identificada, pero si $K-k>m-1$ esta ecuación estará sobre identificada.

5.3.5. Condición de rango para identificación

La condición de rango para identificación rango refiere al rango de una matriz y está dado por la matriz cuadrada de máximo rango contenida en matriz dada, cuyo determinante sea diferente de cero tal que, alternativamente, el rango de una matriz es el número máximo de filas o de columnas linealmente independientes de dicha matriz.

La condición de orden analizada anteriormente es una condición necesaria pero no suficiente para la identificación; es decir, aun si se cumple, puede suceder que una ecuación no esté identificada.

En términos más generales, aun si una ecuación cumple la condición de orden $K - k \geq m - 1$ puede no estar identificada porque las variables predeterminadas excluidas de esa ecuación, pero presentes en el modelo, quizá no todas sean independientes de manera que tal vez no exista una correspondencia uno a uno entre los coeficientes estructurales (β) y coeficientes en forma reducida (Π).

Es decir, probablemente no sea posible estimar parámetros estructurales a partir de coeficientes en forma reducida. Entonces, se requiere una condición que sea tanto necesaria como suficiente para la identificación. Ésta es condición de rango para la identificación: En un modelo que contiene M ecuaciones en M variables endógenas, una ecuación está identificada si y sólo si puede construirse por lo menos un determinante diferente de cero, de orden $(M - 1)(M - 1)$, a partir de coeficientes de variables endógenas y predeterminadas excluidas de esa ecuación particular, pero incluidas en otras ecuaciones del modelo.

Con base en condición de rango para identificación, considere el siguiente sistema hipotético de ecuaciones simultáneas (Y variables endógenas o explicadas y X predeterminadas, exógenas o explicativas):

$$
\begin{array}{llllllll}
Y_{1t} & & -\beta_{12}Y_{2t} & -\beta_{13}Y_{3t} & -\gamma_{11}Y_{1t} & & & = u_{1t} \\
-\beta_{10} & & & & & & & \\
Y_{2t} & & & & -\beta_{23}Y_{3t} & -\gamma_{21}X_{1t} & -\gamma_{22}Y_{2t} & = u_{2t} \\
-\beta_{20} & & & & & & & \\
Y_{3t} & -\beta_{31}Y_{1t} & & & -\gamma_{31}X_{1t} & -\gamma_{32}Y_{2t} & & = u_{3t} \\
-\beta_{30} & & & & & & & \\
Y_{4t} & -\beta_{41}Y_{1t} & -\beta_{42}Y_{2t} & & & & -\gamma_{43}Y_{3t} & = u_{4t} \\
-\beta_{40} & & & & & & &
\end{array}
$$

Con el fin de facilitar la identificación, este sistema se escribe en siguiente tabla:

Coeficientes de variables							
1	**Y_1**	**Y_2**	**Y_3**	**Y_4**	**X_1**	**X_2**	**X_3**
$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010) y (Hurtado & Gómez, 2010)

En primera instancia, se aplica la condición de orden para la identificación:

Número de variables predeterminadas excluidas $(K - k)$	**Número de variables endógenas incluidas menos uno $(m - 1)$**	**¿Identificadas?**
2	2	Exactamente
1	1	Exactamente
1	1	Exactamente
2	2	Exactamente

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010) y (Hurtado & Gómez, 2010)

Cada ecuación está identificada por la condición de orden. Verifique esto con condición de rango. Considere la primera ecuación, que excluye las variables Y_4, X_2 y X_3, representada por ceros en primer renglón de primera tabla.

Para que esta ecuación esté identificada, se obtendrá por lo menos un determinante diferente de cero de orden $3 * 3$, a partir de coeficientes de variables excluidas de esta ecuación, pero incluidas en otras. Para conseguir el determinante, se obtiene primero la matriz relevante de coeficientes de variables Y_4, X_2 y X_3, incluidas en otras ecuaciones.

En este caso, sólo existe una matriz, llamada **A**:

$$\mathbf{A} = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

El determinante de la matriz es cero:

$$\det(\mathbf{A}) = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

El determinante es cero, el rango de matriz, $\rho(\mathbf{A}) = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix} < 3$ tal que ecuación:

$$\begin{matrix} Y_{1t} \\ -\beta_{10} \end{matrix} \qquad -\beta_{12}Y_{2t} \quad -\beta_{13}Y_{3t} \quad -\gamma_{11}Y_{1t} \qquad = u_{1t}$$

no satisface la condición de rango y, por tanto, no está identificada. Entonces, condición de rango es tanto necesaria como suficiente para la identificación. Por consiguiente, a pesar de que la condición de orden muestra que ecuación anterior está identificada, su condición de rango muestra que no lo está.

Al parecer, columnas o renglones de la matriz **A** dadas en matriz $\mathbf{A} = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$ no son linealmente independientes, lo que significa que hay alguna relación entre las variables Y_4, X_2 y X_3 tal que puede no haber suficiente información para estimar parámetros de ecuación:

$$\begin{matrix} Y_{1t} \\ -\beta_{10} \end{matrix} \qquad -\beta_{12}Y_{2t} \quad -\beta_{13}Y_{3t} \quad -\gamma_{11}Y_{1t} \qquad = u_{1t}$$

Para el modelo anterior, las ecuaciones en forma reducida mostrarán que no es posible obtener los coeficientes estructurales de esa ecuación a partir de coeficientes en la forma reducida tal que sólo ecuación siguiente sí lo está:

$$Y_{4t} - \beta_{40} \quad -\beta_{41}Y_{1t} \quad -\beta_{42}Y_{2t} \qquad\qquad -\gamma_{43}Y_{3t} \quad = u_{4t}$$

Con base en esto, condición de rango dice si la ecuación bajo consideración está identificada o no, en tanto que la condición de orden expresa si dicha ecuación está exactamente identificada o sobre identificada. El estudio de condiciones de orden y de rango para identificación conduce a principios generales de identificación de una ecuación estructural en sistema de M ecuaciones simultáneas:

a) Si $K - k > m - 1$ con rango de matriz **A** igual a $M - 1$, la ecuación está sobre identificada.
b) Si $K - k = m - 1$ con rango de matriz **A** igual a $M - 1$, la ecuación está exactamente identificada.
c) Si $K - k \geq m - 1$ con rango de matriz **A** igual a $M - 1$, la ecuación está sub identificada.
d) Si $K - k < m - 1$ con rango de matriz $\mathbf{A} < M - 1$, la ecuación no está identificada.

Cuando se hable de identificación, debe entenderse identificación exacta o sobre identificación. No tiene sentido considerar ecuaciones no identificadas o sub identificadas puesto que, no importa qué tan completa sea la información, los parámetros estructurales no pueden ser estimados. Entonces, es posible identificar parámetros de ecuaciones sobre identificadas e igual de ecuaciones exactamente identificadas.

Para modelos grandes de ecuaciones simultáneas, la aplicación de condición de rango es una labor muy dispendiosa tal que condición de orden, por lo general, es suficiente para asegurar identificación y, es importante tener conciencia de condición de rango, no verificar su cumplimiento raramente resultará en un desastre.

5.3.6. Prueba de simultaneidad

Si no hay ecuaciones simultáneas o presencia de simultaneidad, MCO producen estimadores consistentes y eficientes. Si hay simultaneidad, estimadores de MCO no son ni siquiera consistentes tal que en presencia de simultaneidad, los métodos de mínimos cuadrados en dos etapas (MC2E) y variables instrumentales producirán estimadores consistentes y eficientes.

En una prueba de simultaneidad, se intenta averiguar si una regresora o endógena está correlacionada con el término de error. Si lo está, existe el problema de simultaneidad, en cuyo caso deben encontrarse alternativas a MCO, pero si no lo está, se pueden utilizar MCO.

Ejemplo ilustrativo. Considere el modelo de demanda y oferta siguiente:

Función de demanda: $$Q^{d\,(\text{Cantidad demandada})}_{t\,(\text{En tiempo})} = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \upsilon_{1t}$$

Función de oferta: $$Q^{S\,(\text{Cantidad ofertada})}_{t\,(\text{En tiempo})} = \beta_0 + \beta_1 P_t + \upsilon_{2t}$$

Con base en esto, considere el siguiente modelo con un cambio:

Función de demanda: $$Q_{t\ (\text{En tiempo})} = \alpha_0 + \alpha_1 P_{t\ (\text{Precio})} + \alpha_2 X_{t\ (\text{Ingreso o gasto})} + \upsilon_{1t}$$

Función de oferta: $$Q_{t\ (\text{En tiempo})} = \beta_0 + \beta_1 P_{t\ (\text{Precio})} + \upsilon_{2t}$$

Considere variable $X_{t\ (\text{Ingreso o gasto})}$ exógena o explicativa. La función de oferta está exactamente identificada, mientras que función de demanda no lo está. Las ecuaciones en forma reducida son:

$$P_{t\ (\text{En tiempo})} = \Pi_0 + \Pi_1 X_{t\ (\text{Ingreso o gasto})} + w_{1t\ (\text{Combinaciones lineales de perturbaciones } \upsilon_1)}$$

$$Q_{t\ (\text{En tiempo})} = \Pi_2 + \Pi_3 X_{t\ (\text{Ingreso o gasto})} + \nu_{t\ (\text{Combinaciones lineales de perturbaciones } \upsilon_2)}$$

$\Pi_{(\text{Letra mayúscula phi})}$ son coeficientes en forma reducida y son combinaciones no lineales de coeficientes estructurales:

$$\Pi_0 = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right);\ \Pi_1 = -\left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right)$$

Cada ecuación en forma reducida contiene una sola variable endógena, que es dependiente y está en función únicamente de variable exógena $X_{t\ (\text{Ingreso o gasto})}$ y perturbaciones estocásticas.

Por tanto, los parámetros de ecuaciones en forma reducida anteriores pueden ser estimados por MCO y sus estimaciones son:

$$\hat{\Pi}_0 = \overline{P}_{(\text{Precio})} - \hat{\Pi}_1 \overline{X}_{(\text{Ingreso o gasto})}; \hat{\Pi}_1 = \left(\frac{\sum p_{t\ (\text{Precio})} x_{t\ (\text{Ingreso o gasto})}}{\sum x^2_{t\ (\text{Ingreso o gasto})}}\right)$$

$$\hat{\Pi}_2 = \overline{Q}_{(\text{Cantidad})} - \hat{\Pi}_3 \overline{X}_{(\text{Ingreso o gasto})};\ \hat{\Pi}_3 = \left(\frac{\sum q_{t\ (\text{Cantidad})} x_{t\ (\text{Ingreso o gasto})}}{\sum x^2_{t\ (\text{Ingreso o gasto})}}\right)$$

Letras minúsculas $p_{t\ (\text{Precio})}, q_{t\ (\text{Cantidad})}, x_{t\ (\text{Ingreso o gasto})}$ y $x^2_{t\ (\text{Ingreso o gasto})}$ denotan desviaciones de medias muestrales, mientras que $\overline{Q}_{(\text{Cantidad})}$ y $\overline{P}_{(\text{Precio})}$ son valores de media muestral. $\hat{\Pi}_i$ son estimadores consistentes y, bajo supuestos apropiados, son insesgados, con varianza mínima o asintóticamente eficientes.

Las estimaciones pueden ser "las mejores insesgadas o asintóticamente eficientes, respectivamente, dependiendo que: i) las z = X sean exógenas y no simplemente predeterminadas; es decir, no contengan rezagados de variables endógenas o ii) la distribución de perturbaciones sea normal".

El principal objetivo es determinar coeficientes estructurales, es importante ver posibilidad de estimarlos con base en coeficientes en forma reducida. Como se señala en "modelo con un cambio", la función de oferta está exactamente identificada tal que sus parámetros pueden estimarse de manera única a partir de los coeficientes en forma reducida:

$$\beta_0 = \Pi_2 - \beta_1 \Pi_0;\ \beta_1 = \left(\frac{\Pi_3}{\Pi_1}\right)$$

Por tanto, estimaciones de estos parámetros pueden obtenerse a partir de estimaciones de coeficientes en forma reducida:

$$\hat{\beta}_0 = \hat{\Pi}_2 - \hat{\beta}_1 \hat{\Pi}_0;\ \hat{\beta}_1 = \left(\frac{\hat{\Pi}_3}{\hat{\Pi}_1}\right)$$

Estos son estimadores por MCI; aunque, los parámetros de función de demanda no pueden ser estimados así.

Su finalidad es igual a regresión lineal simple; aunque, con la diferencia que interviene más de un regresor en el modelo tal que en lugar de rectas de regresión se tendrán hiper planos de regresión. Un plano en un espacio de tres dimensiones para el caso de dos regresores. En este tema se abordan temas estimación, inferencia, adecuación de modelo y predicción.

No siempre un regresor es suficiente para explicar todas las variaciones de una variable dependiente (Y); por ejemplo: el gerente de una empresa desea incrementar las ventas tal que decide realizar gastos en publicidad y medir la

variación en sus ventas. Inicialmente, decide puntuar la publicidad en televisión, pero posteriormente decide también ponerla en la radio y los periódicos.

En la primera etapa la variable de respuesta, que es el incremento en las ventas depende de una sola variable predictora (gastos en televisión) y para realizar un análisis es suficiente empleando un modelo de regresión lineal simple. Más en la segunda etapa, si variable de respuesta depende de un sólo análisis no es suficiente la regresión lineal simple.

En general, aunque hay muchos problemas prácticos que atañen a variables predictoras únicas es mucho más frecuente que la variable repuesta dependa de un conjunto de variables predictoras o transformaciones de estas. Suponga que Y es una variable que depende funcionalmente de ciertas variables $X_2, X_3, X_4, X_5, \dots, X_k$ tal que la numeración comienza en dos pues el número uno se reserva para el término constante o intercepto.

La relación exacta entre Y y $\{X_j\}$ se desconoce, pero supone pertenece a una familia de funciones F llamada modelo. El más usado y sirve como punto de partida para hallar otros es el lineal tal que supone que F es una familia de funciones lineales afines:

$$F = \{Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots + \beta_k X_k | \beta_1, \beta_2, \beta_3, \beta_4, \dots, \beta_k \in \mathbb{R}\}$$

Se sabe que variable Y toma nombres como dependiente, endógena, respuesta o explicada mientras que variables $\{X_j\}$ toman nombres independientes, exógena, regresora, predictor o explicativa. No obstante, para determinar aproximadamente la función que liga a variable dependiente Y con variables independientes $\{X_j\}$ se realizan n observaciones de $k-$ uplas:

$$(y_i, x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})\ i = 1,2,3,4, \dots, n$$

Se conoce que y_j es valor que toma variable Y cuando $X_2 = x_{i2}, x_{i3}, x_{i4}, \dots, X_k = x_{ik}$. El modelo de regresión lineal múltiple supone que y_i es una observación de una variable aleatoria Y.

Las variables $\{X_j\}$ se estima determinísticas o no aleatorias, pues sus valores pueden ser fijados según conveniencias del experimentador. En una investigación relacionada con experimento para cada $k-1$ upla $(x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})$ se hacen varias observaciones o mediciones de variable respuesta o endógena Y.

No obstante, en econometría es común trabajar con datos provenientes de una muestra, en este caso todas las variables que intervienen en modelo pueden ser aleatorias, pero variables exógenas participan como si fueran determinísticas tal que sus valores no dependen del azar o aleatoriedad.

El modelo de regresión lineal múltiple, que liga a una variable dependiente Y con k variables independientes, exógenas, explicativas mediante la ecuación

$$y_i = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \cdots + \beta_k x_{ik} + u_i \quad i = 1,2,3,4, \dots, n$$

se llama modelo de regresión lineal múltiple con k variables independientes, exógenas, regresoras, predictoras o explicativas, $\beta_k \quad k = 1,2,3,4, \dots, j$ coeficientes de regresión y, también, parámetros desconocidos, no aleatorios, estimados con base en datos observados.

Paralelamente, el caso de una sola variable se considera que el error o perturbación $u_i \equiv e_i$ tiene esperanza 0, varianza σ^2, errores e_i por observación, no correlacionados y es una variable no observable.

Observación. El término lineal refiere a que la relación es como tal en parámetros, pero no en variables. Las técnicas de regresión lineal múltiple pueden usarse si la relación es lineal en parámetros, pero no se olvide que se busca una función, no necesariamente lineal, que relacione variables Y con $\{X_j\}$:

$$Y_{(\text{dependiente,endógena,respuesta o explicada})} = f(X_2, X_3, X_4, X_5, \dots, X_k)$$

Sin embargo, esta relación es ambigua o incierta, pues se ve acompañada por errores y depende de algunos parámetros desconocidos. Además, las hipótesis o supuestos en que se sustenta la regresión lineal múltiple son iguales que regresión lineal simple. Con el objeto de que estimadores de mínimos cuadrados sean únicos se supones que los siguientes vectores son linealmente independientes:

1. Vectores $(u_1, u_2, u_3, u_4, \dots, u_k)$ de $\mathbb{R}^n$ son linealmente independientes si:

   $$a_1u_1 + a_2u_2 + a_3u_3 + a_4u_4 + \cdots + a_ku_k = 0$$

   Implica que todos los coeficientes $a_j = 0$ o nulos. Si no, son linealmente independientes. Un conjunto infinito B de vectores es linealmente independiente ssi todo conjunto B de vectores es linealmente independiente ssi todo sub conjunto finito de B es linealmente independiente.

2. Si vectores $\{u_j\}$ son linealmente dependientes existe, en igualdad anterior, un coeficiente $a_j \neq 0$ o no nulo tal que el vector correspondiente a dicho coeficiente es combinación lineal del resto de vectores. Por ejemplo: $a_j \neq 0 \Rightarrow u_k = \left(-\frac{1}{a_k}\right)(a_1u_1 + a_2u_2 + a_3u_3 + a_4u_4 + \cdots + a_{k-1}u_{k-1})$.

3. Bases y dimensión. Sea E un espacio vectorial y B un sub conjunto de E:

   a) Si B genera a E ssi $\forall$ vector de E es una combinación lineal de vectores de B.

   b) Si B es linealmente independiente se dice que B es una base de E.

c) Si B es una base de E, el número de elementos de B se llama dimensión de B, si B es infinito se dice que E es dimensión infinita.

4. Número de filas linealmente independientes, en una matriz $A_{(n*m)}$ es igual a número de columnas linealmente independientes.

5. Sea A una matriz $A_{(n*m)}$. El número de filas o columnas linealmente independientes se llama rango de matriz y se denota como $rg(A_{(Rango)})$.

6. De bases y dimensión. Sea E un espacio vectorial y B un sub conjunto de E, es obvio que $rg(A_{(Rango)}) \leq \min\{n, m\}$. Si se logra igualdad $[rg(A_{(Rango)}) \leq \min\{n, m\}]$ se dice que matriz A es rango completo.

7. Matriz cuadrada A es simétrica ssi al cambiar filas por columnas la matriz A no cambia; es decir, ssi $A^t = A$ donde A^t es transpuesta de A.

8. Matriz cuadrada A es diagonal ssi es cuadrada y todos los elementos fuera de la diagonal principal son 0 o nulos:

$$A = diag_{(Diagonal)}\{a_1 + a_2 + a_3 + a_4 + \cdots + a_n\}$$

$$= \begin{bmatrix} a_1 & 0 & 0 & 0 & 0 & & 0 \\ 0 & a_2 & 0 & 0 & 0 & & 0 \\ 0 & 0 & a_3 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & a_4 & 0 & & 0 \\ 0 & 0 & 0 & 0 & a_5 & & 0 \\ & & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & a_n \end{bmatrix}$$

9. Matriz identidad es matriz diagonal con todos $a_i = 0$:

$$I = diag_{(Diagonal)}\{1,1,1,1, \dots ,1\}$$

Cuando requiere indicar el orden de matriz se estribe I_n en vez de I.

10. Si A matriz cuadrada $(n * n)$ es no singular ssi existe una matriz B cuadrada $(n * n)$:

$$AB = BA = I$$

Matriz B se llama inversa de A y se escribe A^{-1}

11. $(AB)^t = B^t A^t$. Si matrices (A, B) son no singulares y de igual dimensión $(AB)^{-1} = B^{-1}A^{-1}$.

12. Una matriz cuadrada $A_{(n*n)}$ es simétrica semi definida positiva, escrita como $A_{(n*n)} \geq 0$, ssi es simétrica y $\forall x, x \in \mathbb{R}^n, x^t Ax \geq 0$ se dice simétrica definida positiva, escrita $A > 0$, ssi es se mi definida positiva y $x^t Ax = 0 \Rightarrow x = 0$.

13. Sea E, F dos espacios vectoriales sobre un mismo cuerpo de escalares K. Una función $f: E \longrightarrow F$ se dice aplicación lineal ssi $\forall a \in k, \forall x, y \in E$ se tiene que $f(ax + y) = af(x) + f(y)$.

14. $\forall A_{(m*n)}$ define una aplicación lineal $\mathbb{R}^n$ en $\mathbb{R}^m$:

$$A: \mathbb{R}^n \rightarrow \mathbb{R}^m : x \rightarrow Ax$$

A su vez toda aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ define una matriz $(m * n)$.

15. Si A es aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ se llama espacio imagen de A al conjunto:

$$img_{(Imagen)}(A) = \{w \in \mathbb{R}^m | w = Au, u \in \mathbb{R}^n\}$$

Se llama núcleo de A al conjunto:

$$\mathrm{Ker}(A) = \{u \in \mathbb{R}^n | Au = 0\}$$

Se puede demostrar que $\dim_{(\text{Dimensión})}\left[\mathrm{img}_{(\text{Imagen})}(A)\right] = \mathrm{rg}\left(A_{(\text{Rango})}\right)$ y $\dim_{(\text{Dimensión})}\left[\mathrm{img}_{(\text{Imagen})}(A)\right] + \dim_{(\text{Dimensión})}\left[\mathrm{Ker}_{(\text{Kernel o núcleo})}(A)\right] = n$

16. Si $A_{(n*n)}$ es matriz cuadrada $(n * n)$ simétrica definida positiva, se define el producto escalar de dos vectores u, v de $\mathbb{R}^n$ respecto de A por $\langle u, u\rangle_{(A)} = (u^t)(Au)$ se define norma de v respecto de A por $\|u\|_{(A)} = \sqrt{(u^t)(Au)}$. Si A es matriz identidad no hace falta sub índices tal que se dice que son producto escalar y norma usuales.

17. Si A es matriz cuadrada $(n * n)$ se llama traza de A a suma de elementos de su diagonal principal tal que $\mathrm{tr}_{(\text{Traza de A})}(A) = \sum_{i=1}^{n}(a_{ii})$.
18. Sea A matriz $(n * m)$, B matriz $(m * n)$ tal que $\mathrm{tr}_{(\text{Traza de AB})}(AB) = \mathrm{tr}_{(\text{Traza de BA})}(BA)$.

$$\begin{pmatrix}1\\1\\1\\1\\\vdots\\1\end{pmatrix}\begin{pmatrix}x_{12}\\x_{22}\\x_{32}\\x_{42}\\\vdots\\x_{n2}\end{pmatrix},\ldots,\begin{pmatrix}x_{1k}\\x_{2k}\\x_{3k}\\x_{4k}\\\vdots\\x_{nk}\end{pmatrix}$$

Cuando se realiza un modelo de regresión múltiple, generalmente, las unidades de medición no son las mismas para la variable dependiente y para las variables independientes; de manera que los coeficientes de regresión no se pueden comparar directamente.

Para superar esta dificultad, se emplean los coeficientes de regresión estandarizados beta. Las unidades de medición de todas las variables se

transforman a unidades de desviación estándar, dividiendo cada variable por su desviación estándar. Por ejemplo, en la ecuación de regresión se tiene:

$$\frac{\hat{y}}{s_y} = \frac{b_o}{s_y} + \left(b_1 \frac{s_{x1}}{s_y}\right) \frac{x_1}{s_{x1}} + \left(b_2 \frac{s_{x2}}{s_y}\right) \frac{x_2}{s_{x2}}$$

Los coeficientes:

$$\beta_i = b_i \left(\frac{s_{x_i}}{s_y}\right)$$

Son los coeficientes de regresión parcial estándar y su interpretación es la siguiente: si hay una variación de una desviación estándar en x_i, habrá una desviación de betas desviaciones estándar en y.

Es necesario construir estimaciones de intervalos de confianza para coeficientes de regresión $\{\beta_j\}$ y para otras cantidades de interés del modelo de regresión. El desarrollo de un procedimiento para obtener estos intervalos de confianza requiere suponer errores $\{e_i\}$ tienen una distribución normal e independiente con media cero y varianza σ^2 (Montgomery, Diseño y análisis de experimentos, 2004).

Puesto que el estimador de mínimos cuadrados $\hat{\beta}$ es una combinación lineal de observaciones tal que $\hat{\beta}$ tiene una distribución normal con vector medio β y matriz de covarianza $\sigma^2(X'X)^{-1}$. Cada uno de los estadísticos:

$$\left(\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 * C_{jj}}}\right) \quad j = 0,1,2,3,4, \ldots, k$$

Se distribuye con t con $n - p$ grados de libertad, donde C_{jj} es elemento $jj - ésimo$ de matriz $(X'X)^{-1}$ y $\hat{\sigma}^2$ es estimación de varianza de error obtenida de ecuación $\hat{\sigma}^2 = \left(\frac{SS_E \text{ (Suma de cuadrados del error}}{n-p}\right)$. Por tanto, un intervalo de confianza $100 * (1 - \alpha)$ para coeficiente de regresión $\beta_j, j = 0,1,2,3,4, \dots, k$ es (Montgomery, 2004):

$$\hat{\beta}_j - t_{(\alpha/2;n-p)} * \sqrt{\hat{\sigma}^2 * C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{(\alpha/2;n-p)} * \sqrt{\hat{\sigma}^2 * C_{jj}}$$

Este intervalo de confianza también puede escribirse como $\left(se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 * C_{jj}}\right)$:

$$\hat{\beta}_j - t_{(\alpha/2;n-p)} * se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{(\alpha/2;n-p)} * se(\hat{\beta}_j)$$

Puede obtenerse un intervalo de confianza para respuesta media en un punto particular; por ejemplo: $x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}$. Se define el vector:

$$x_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ x_{03} \\ x_{04} \\ \vdots \\ x_{0k} \end{bmatrix}$$

La respuesta media en este punto es:

$$\mu_{(y|x_0)} = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \beta_3 x_{03} + \beta_4 x_{04} + \cdots + \beta_k x_{0k} = x_0' \beta$$

La respuesta media estimada en este punto es:

$$\hat{y} * (x_0) = x_0' \hat{\beta}$$

Este estimador es insesgado, pues $E[\hat{y} * (x_0)] = E[x_0' \hat{\beta}] = x_0' \beta = \mu_{(y|x_0)}$ y varianza $\hat{y} * (x_0)$ es:

$$\mathrm{Var}[\hat{y} * (x_0)] = \sigma^2 * x_0^{'} * x_0 * \left(X^{'}X\right)^{-1}$$

Por tanto, un intervalo de confianza $100 * (1 - \alpha)$ par respuesta media en punto $x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}$ es:

$$\hat{y} * (x_0) - t_{(\alpha/2;n-p)} * \sqrt{x_0^{'} * x_0 * (X^{'}X)^{-1}} \leq \mu_{(y|x_0)} \leq \hat{y} * (x_0) + t_{(\alpha/2;n-p)} * \sqrt{x_0^{'} * x_0 * (X^{'}X)^{-1}}$$

Es posible usar un modelo de regresión para predecir observaciones futuras de respuesta que corresponden a valores particulares de regresores; por ejemplo: $x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}$. Si $x_0^{'} = [1, x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}]$ tal que una estimación puntual de observación futura y_0 en punto $x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}$ se estima con ecuación:

$$\hat{y} * (x_0) = x_0^{'}\hat{\beta}$$

Un intervalo de predicción $100 * (1 - \alpha)$ para esta observación futura es:

$$\hat{y} * (x_0) - t_{(\alpha/2;n-p)} * \sqrt{\hat{\sigma}^2 * \left(1 + x_0^{'} * x_0 * (X^{'}X)^{-1}\right)} \leq y_0 \leq \hat{y} * (x_0) + t_{(\alpha/2;n-p)} * \sqrt{\hat{\sigma}^2 * \left(1 + x_0^{'} * x_0 * (X^{'}X)^{-1}\right)}$$

Cuando se predicen nuevas observaciones y estima respuesta media en un punto dado $x_{01}, x_{02}, x_{03}, x_{04}, \dots, x_{0k}$ es necesario cuidar no hacer extrapolación fuera de región que contiene observaciones originales. Es muy probable que un modelo se ajuste bien en región de datos originales tal que deje de hacerlo fuera de esa región.

Con supuestos de modelo de regresión clásico tal que al tomar la esperanza condicional de Y en ambos lados de modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ se obtiene:

$$E[Y_i | X_{2i}, X_{3i}] = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Expresado en términos de esta ecuación, se obtiene la media condicional o valor esperado de Y condicionado a valores dados o fijos de variables X_2 y X_3. Por consiguiente, como en caso de dos variables, el análisis de regresión múltiple es análisis de regresión condicional sobre valores fijos de variables explicativas y lo que se obtiene es valor promedio o media de Y, o respuesta media de Y a valores dados de regresoras X.

Los coeficientes de regresión β_2, mide el cambio en valor de media de Y, E(Y) por unidad de cambio en X_2 con X_3 constante o proporciona efecto "directo" o "neto" que tiene una unidad de cambio X_2 sobre valor medio de Y neto de cualquier efecto que X_3 pueda ejercer en media Y y β_3, mide el cambio en valor de media de Y, E(Y) por unidad de cambio en X_3 con X_2 constante o proporciona efecto "directo" o "neto" que tiene una unidad de cambio, se conocen como coeficientes de regresión parcial o coeficientes parciales de pendiente. β_2 y β_3 son derivadas parciales de $E[Y|X_2, X_3]$ respecto a X_2 y X_3.

En caso de dos variables $r^2 = \left(\frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}\right) = \left(\frac{SC_{E\ (Error)}}{SC_{T\ (Total)}}\right)$ mide bondad de ajuste de ecuación de regresión; es decir, da proporción o porcentaje de variación total en variable dependiente Y explicada por variable única explicativa X. Esta notación r^2 se extiende fácilmente a modelos de regresión con más de dos variables. En modelo de tres variables se busca conocer proporción de variación en Y explicada por variables X_2 y X_3 conjuntamente. La medida que da esta información se conoce como coeficiente de determinación múltiple y se denota

por R^2, que asemeja a r^2. Para obtener R^2 se puede seguir el procedimiento para obtener r^2 tal que:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i = \hat{Y}_i + \hat{u}_i$$

$\hat{Y}_i$ es valor estimado de Y_i a partir de la línea de regresión ajustada y es un estimador de la verdadera $E[Y_i|X_{2i}, X_{3i}]$. Al reemplazar letras mayúsculas por minúsculas para indicar desviaciones de sus medias, la ecuación anterior se escribe:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i = \hat{y}_i + \hat{u}_i$$

Elevando al cuadrado esta ecuación en ambos lados y sumándole valores muestrales se obtiene:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

En esta ecuación se afirma que suma de cuadrados total (SCT) es igual a suma de cuadrados explicada (SCE) más suma de cuadrados de Residuos (SCR). Sustituyendo el equivalente a $\sum \hat{u}_i^2$ dado en ecuación $\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$ se obtiene:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{y}_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

Reordenando esta ecuación se tiene:

$$SCE = \sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$$

$$R^2 = \left(\frac{SC_{E\ (Explicada\ por\ regresión)}}{SC_{T\ (Total)}}\right) = \left(\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}\right) = \left(1 - \frac{RSS}{TSS}\right)$$
$$= \left(1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}\right) = \left(1 - \frac{(n-3) * \hat{\sigma}^2}{(n-1) * S_y^2}\right)$$

R^2, al igual que r^2, se encuentra entre 0 y 1. Si es 1, la línea de regresión ajustada explica 100% de variación en Y. Si es 0, el modelo no explica nada de variación en Y. Por lo general, R^2 se encuentra entre estos valores extremos. Se dice que el ajuste del modelo es "mejor" entre más cerca esté R^2 de 1.

Al igual que en el modelo de regresión lineal simple, en el modelo múltiple el R^2 mide la proporción de la variación en la variable dependiente explicada por la variación conjunta de todas las variables independientes o explicativas, una vez descontado el efecto de la media. Es también un buen indicador de la bondad de ajuste del modelo. Varía entre cero y uno.

Toma el valor cero si la regresión es una línea horizontal; es decir, todas las betas son cero, excepto el término constante. En este caso el valor predicho de y es siempre la media de y, pues las desviaciones de **x** con respecto a su media no se traducen en diferentes predicciones para y.

Por lo tanto, x no tiene poder explicativo sobre y. El otro extremo, cuando $R^2 = 1$ que ocurre si los valores de x y de y están en el mismo hierplano (en una línea recta en la regresión simple de dos variables) por lo que los residuos son cero. Si todos los valroes de y están en una línea vertical, R^2 no tiene sentido y no puede calcularse.

Para el cálculo de R^2 se utiliza la misma expresión que en el caso de regresión simple:

R^2 = SCR/(SCR + SCE), o

$R^2 = b'X'y - nY / b'X'y - nY + y'y - b'X'y = b'X'y – nY/ y'y - nY$

Alternativamente, $R^2 = b^2 S_{xx}/S_{yy}$

Existen algunos problemas al utilizar R^2 para medir la bondad de ajuste. Uno de ellos hace referencia al número de grados de libertad utilizados en la estimación de los parámetros. R^2 nunca decrecerá cuando se añade otra variable a la ecuación de regresión por lo que se puede estar tentado a explotar este resultado mediante la continua adición de variables en el modelo; R^2 continuará creciendo hasta su límite de 1.

En vista de ello, a veces se presenta el R^2 ajustado (por grados de libertad), que se calcula como sigue:

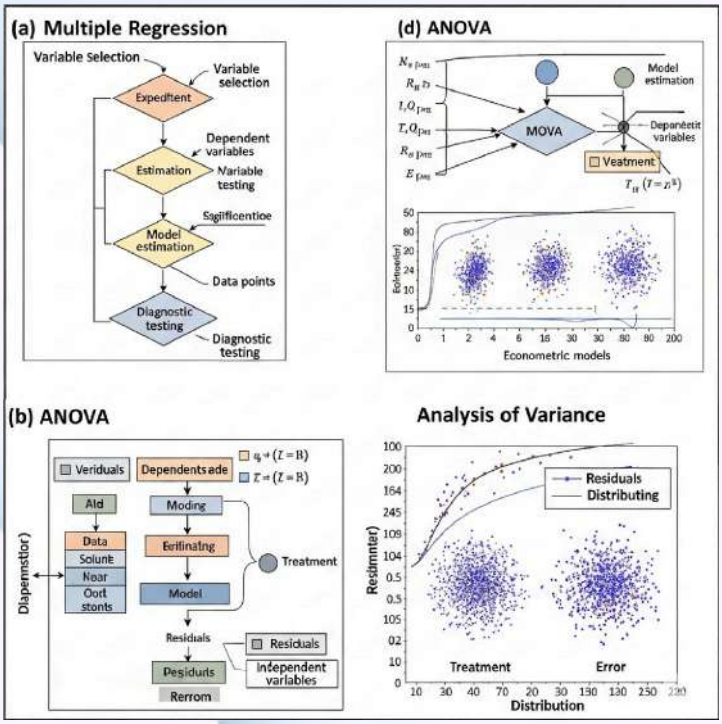
$$R^2 = 1 – (n-1)/(n-k)(1- R^2)$$

El principal punto para tener en cuenta de esta medida es que podría disminuir cuando se añade una variable al conjunto de variables independientes. De hecho, R^2 podría ser incluso negativo.

EDITORIAL ANDES COGNITIO

CAPÍTULO VI

REGRESIÓN MÚLTIPLE Y ANÁLISIS DE VARIANZA (ANOVA) EN MODELOS ECONÓMETRICOS



CAPÍTULO VI

REGRESIÓN MÚLTIPLE Y ANÁLISIS DE VARIANZA (ANOVA) EN MODELOS ECONOMÉTRICOS

6. Modelo de regresión múltiple

En el modelo de regresión múltiple es natural buscar una extensión de la noción de correlación simple para tomar en cuenta el hecho de que otras variables se mantienen constante cuando analizamos un coeficiente de regresión en particular.

En otras palabras, queremos saber si la variable dependiente y una variable independiente cualesquiera están relacionados después de descontar el efecto de las otras variables independientes del modelo.

Para efectuar esta tarea, considérese el modelo:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$$

6.1. Coeficiente de correlación

El coeficiente de correlación parcial entre Y y X_1 debe ser definido de tal forma que mida el efecto de X_1 sobre Y que no es explicado por la otra variable en el modelo. Más claramente, el coeficiente de correlación parcial es calculado eliminando el efecto lineal de X2 sobre Y (así como el efecto lineal de X2 sobre X1) y entonces correr la regresión apropiada.

Los pasos que se siguen son los siguientes:

1. Se corre el modelo de Y sobre X2 y se obtienen los valores ajustados

$$Yi = b0 + b_2 X_{i2}$$

2. En seguida se corre el modelo de X_1 sobre X_2 para obtener la ecuación

$$X_{i1} = g_0 + g_2 X_{i2}$$

3. Se remueve la influencia de X_2 tanto en Y como en X_1. Se hace

$$Y^* = Y - Y \quad \text{y} \quad X^*_1 = X_1 - X_1$$

4. La correlación parcial entre X_1 y Y es entonces la correlación simple entre Y^* y X^*_1.

Para ver porqué la regresión de Y^* sobre X^*_1 se proporciona el coeficiente de correlación parcial deseado, nótese que Y^* y X^*_1 no están correlacionadas con X_2 por construcción, dado que si tomamos por ejemplo Y^* y X_2 sabemos que no están correlacionadas por el hecho de que Y^* representa el residuo de la regresión de Y sobre X_2 y se ha visto anteriormente que los residuos no están correlacionados con las variables explicativas.

Entonces, la regresión de Y^* sobre X^*_1 relaciona la parte de Y, que no está correlacionada con X_2, con la parte de X_1 que tampoco está correlacionada con X_2. Generalmente se denotan estas correlaciones de la siguiente manera:

$\mathbf{r}_{YX_1 * X_2}$: que es la correlación parcial de Y y X_1 (controlando X_2)

$\mathbf{r}_{YX_1}$: correlación simple entre Y y X_1

$\mathbf{r}_{X_1X_2}$: correlación simple entre X_1 y X_2

Con esta definición de correlación parcial, no es difícil (aunque algo tedioso) derivar la relación entre la correlación parcial y la correlación simple entre variables de interés. Estableceremos, sin probarlo, el siguiente resultado (el lector puede hacer el ejercicio para demostrarlo):

$$\mathbf{r}_{YX_1*X_2} = (\mathbf{r}_{YX_1} - \mathbf{r}_{YX_2}\mathbf{r}_{X_1X_2})/\mathbf{r}(1 - \mathbf{r}^2_{X_1X_2})\mathbf{r}(1 - \mathbf{r}^2_{YX_2})$$

$$\mathbf{r}_{YX_2*X_1} = (\mathbf{r}_{YX_2} - \mathbf{r}_{YX_1}\mathbf{r}_{X_1X_2})/\mathbf{r}(1 - \mathbf{r}^2_{X_1X_2})\mathbf{r}(1 - \mathbf{r}^2_{YX_1})$$

A partir de estas ecuaciones se pueden hacer las siguientes afirmaciones con relación al coeficiente de correlación parcial:

1. Tal como en el caso de las correlaciones simples, las correlaciones parciales toman valores entre -1 y +1.

2. Una correlación parcial igual a cero entre Y y X_1 indica que no existe relación lineal entre estas variables, después que el efecto lineal de X_2 sobre ambas ha sido removido. En este caso se dirá que X_1 no tiene un efecto directo sobre Y en el modelo. De hecho, los coeficientes de correlación parcial son utilizados para determinar la importancia relativa de diferentes variables en el modelo múltiple.

3. Es también útil determinar la relación entre la correlación parcial y el coeficiente de determinación múltiple R^2. En el modelo de dos variables no es difícil demostrar de que es posible interpretar R^2 como el cuadrado de la correlación simple entre la variable dependiente y la independiente.

También es posible interpretar la correlación parcial entre Y y X_1 como una medida de la raíz cuadrada del porcentaje de la varianza en Y que no puede ser atribuible a X_2 pero que es atribuible a la parte de X_1 que no está correlacionada a X_2.

A partir de este hecho, es posible factorizar el coeficiente de correlación múltiple en distintos componentes, cada uno de los cuales sirve para explicar una porción de la suma de cuadrados de la regresión en el modelo. Por lo tanto, es posible derivar la siguiente relación entre el coeficiente de correlación múltiple y el coeficiente de correlación parcial:

$$r^2_{YX_1*X_2} = (R^2 - r^2_{YX_2})/(1 - r^2_{YX_2})$$

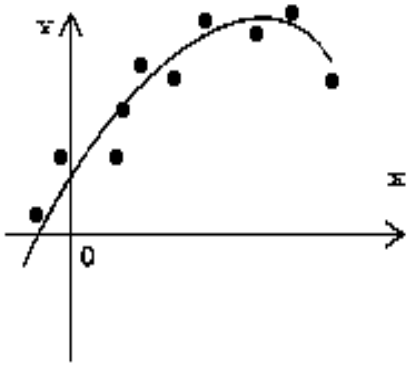
$$\mathbf{1 - R^2} = (1 - \mathbf{r}^2_{YX_2})(1 - \mathbf{r}^2_{YX_1*X_2})$$

Estas relaciones no son fáciles de interpretar. Tomando raíz cuadrada en ambos lados de la primera ecuación, se puede afirmar que el coeficiente de correlación parcial puede ser determinado sacando raíz cuadrada del porcentaje de la varianza en Y explicado por X_1 (habiendo ajustado ambas variables para eliminar el efecto de X_2). Un uso importante de la correlación parcial es en el procedimiento de "regresión stepwise" (paso sabio).

6.2. Procedimiento

El primer paso para escoger un modelo que describa los datos es la realización de una dispersión de observaciones. La relación sugerida por datos es la que permite escoge el modelo que los describa adecuadamente. Cuando los datos presentan un esquema de comportamiento curvilíneo puede ser necesario un modelo de tipo polinomial para datos:

Figura 6.1. Modelo de segundo grado para datos



Fuente: (Gujarati & Porter, 2010)

Para transformar un modelo polinomial en uno de regresión multiplique se escoge un modelo de segundo grado en una variable:

$$y_i = \beta_1 + \beta_2 x_1 + \beta_2 x_2 + e_i$$

Si se hace situaciones de variables $x_1 = x^1$ y $x_2 = x^2$ la ecuación sería:

$$y_i = \beta_1 + \beta_2 x^1 + \beta_2 x^2 + e_i$$

Es un modelo de regresión múltiple en dos variables. Se está en posibilidad de aplicar la teoría descrita. En general, si se tiene un modelo polinomial con una variable explicativa

$$y_i = \beta_1 + \beta_2 x^1 + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \cdots + \beta_k x^k + e_i$$

Se puede convertir en uno de regresión múltiple mediante la transformación $x_i + x^i$:

$$y_i = \beta_1 + \beta_2 x_1 + \beta_2 x^2 + \beta_3 x_3 + \beta_4 x^4 + \cdots + \beta_k x_k + e_i$$

Otros modelos polinomiales que incluyen más de un variable, que pueden transformarse a una línea múltiple, son los polinomios en variables, como el de segundo grado en dos variables:

$$y_i = \beta_1 + \beta_2 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + e_i$$

Cuando se ajusta un modelo polinomial es preciso escoger el polinomio de menor grado posible tal que se deberán realizar reiteradas pruebas de hipótesis, en las que se fijaran aquellas variables que se han de incluir y excluir en modelo final.

En los modelos tratados se empleó variables independientes de naturaleza cuantitativa; es decir, se expresan numéricamente y son el resultado de mediciones instrumentales. Si se desea incorporar en modelo una variable cualitativa será necesario introducir variables indicadoras, ficticias, que permiten diferenciar los distintos niveles que toma tal variable; por ejemplo, una variable X que indiquen la estación del año puede ser definida:

$$X = \begin{cases} 0: \text{Es invierno} \\ 1: \text{Es verano} \end{cases}$$

En general, una variable cualitativa con t niveles se representa mediante $t - 1$ variable indicadoras, se asignan valores de 0 y 1 llamadas dicotómicas, binomiales o dummy.

Después que se ha estimado un modelo de regresión se debe, obligatoriamente, estudiar si ese modelo retenido es adecuado. También, se debe probar si supuestos con base en que se sustenta son aceptables a vista de datos.

De manera más precisa, ¿será verdad que errores son independientes, normalmente distribuidos con media cero y varianza constante, será lineal el modelo? No se puede dar una respuesta exacta, pero residuos darán pautas para decidir si no se acepta o no se rechaza estas hipótesis.

En primer término, se estudia una prueba de linealidad bajo supuesto que existen observaciones repetidas y, después, se analiza residuos.

Si para cada upla $(x_{i2}, x_{i3}, x_{i4}, x_{i5}, \dots, x_{ik})$ existen varias observaciones de variable dependiente Y, se puede construir una prueba de linealidad. Suponga para cada upla $(x_{i2}, x_{i3}, x_{i4}, x_{i5}, \dots, x_{ik})$ se hacen n_i mediciones u observaciones de variable respuesta, así que se tiene el modelo:

$$y_{ir} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \cdots + \beta_k x_{ik} + u_{ir}\ i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$$

Suponga $\sum_{i=1}^{m} n_i = n, m > k$ y rango de matriz de diseño es k. Las hipótesis respecto a errores son habituales, independientes y normalmente distribuidos con parámetros $(0, \sigma^2)$. Se hallarán estimadores mínimos cuadrados ordinarios (MCO). Sea:

$$f(b_1, b_2, b_3, b_4, \dots, b_k) = \sum_{i=1}^{m} \sum_{r=1}^{n_i} \left(y_{ir} - b_1 - \sum_{j=2}^{k} (b_j x_{ij}) \right)^2$$

$$\frac{\partial(f)}{\partial(b_1)} (b_1, b_2, b_3, b_4, \dots, b_k) = -2 \sum_{i=1}^{m} \sum_{r=1}^{n_i} \left(y_{ir} - b_1 - \sum_{j=2}^{k} (b_j x_{ij}) \right)$$

$$\frac{\partial(f)}{\partial(b_1)} (b_1, b_2, b_3, b_4, \dots, b_k) = -2 \sum_{i=1}^{m} \sum_{r=1}^{n_i} \left(y_{ir} - b_1 - \sum_{j=2}^{k} (b_j x_{ij}) \right) x_{il}\ \forall l$$

Si se iguala a 0 se tiene el sistema normal de ecuaciones:

$$\begin{cases} \sum_{i=1}^{m}\left[(n_i)\left(\bar{y}_{i.}-\hat{\beta}_1-\sum_{j=2}^{k}(\hat{\beta}_j x_{ij})\right)\right]=0 \text{ tal que } \bar{y}_{i.}=\left[\left(\frac{1}{n_i}\right)\left(\sum_j(y_{ij})\right)\right] \\ \sum_{i=1}^{m}\left[(n_i)\left(\bar{y}_{i.}-\hat{\beta}_1-\sum_{j=2}^{k}(\hat{\beta}_j x_{ij})\right)\right]x_{il}=0\ \forall l \end{cases}$$

Es equivalente a ecuación matricial:

$$(X^t R X)\hat{\beta}=X^t R\bar{Y} \text{ tal que } R=\text{diag}_{(\text{Diagonal})}\{n_1,n_2,n_3,n_4,n_5,\dots,n_k\}$$

$$\bar{Y}=\begin{bmatrix}\bar{y}_{1.}\\ \bar{y}_{2.}\\ \bar{y}_{3.}\\ y_{4.}\\ \vdots\\ \bar{y}_{k.}\end{bmatrix}, X=\begin{bmatrix}1 & x_{12} & x_{13} & x_{14} & \cdots & x_{1k}\\ 1 & x_{22} & x_{23} & x_{24} & \cdots & x_{2k}\\ 1 & x_{32} & x_{33} & x_{34} & \cdots & x_{3k}\\ 1 & x_{42} & x_{43} & x_{44} & \cdots & x_{4k}\\ & \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{m2} & x_{m3} & x_{m4} & \cdots & x_{mk}\end{bmatrix}$$

Por hipótesis la matriz X es de rango $k<m$ tal que la matriz $X^t R X$ es de rango completo, tal que:

$$\hat{\beta}=(X^t R X)^{-1}(X^t R\bar{Y})$$

$$\hat{y}_{ir}=\hat{y}_i=(1,x_{i2},x_{i3},x_{i4},x_{i5},\dots,x_{ik})\hat{\beta} \text{ tal que } r=1,2,3,4,5,\dots,n_i$$

Prueba. La repetición de mediciones es de suma utilidad para validar el modelo. Para probar el ajuste del modelo $y_{ir}=\beta_1+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}+\beta_5 x_{i5}+\cdots+\beta_k x_{ik}+u_{ir}\ i=1,2,3,4,5\dots,m; r=1,2,3,4,5,\dots,n_i$ a datos, se compara con modelo:

$$y_{ir}=\mu_i+\varepsilon_{ir}\ i=1,2,3,4,5\dots,m; r=1,2,3,4,5,\dots,n_i$$

Tal que errores $\{\varepsilon_{ir}\}$ son independientes normalmente distribuidos con parámetros $(0,\sigma^2)$. En modelo $y_{ir}=\mu_i+\varepsilon_{ir}\ i=1,2,3,4,5\dots,m; r=$

$1,2,3,4,5,\dots,n_i$ supone que medias μ_i dependen de valores $(x_{i2}, x_{i3}, x_{i4}, x_{i5}, \dots, x_{ik})$, pero no obedecen a estructura alguna en particular.

Se dice que modelo $y_{ir} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \cdots + \beta_k x_{ik} + u_{ir}$ $i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$ es adecuado a nivel de significancia α si hipótesis $H_0: \mu = X\beta$ no es rechazada a nivel α.

Teorema. Bajo hipótesis de normalidad, no se acepta modelo $y_{ir} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \cdots + \beta_k x_{ik} + u_{ir}$ $i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$ a favor de modelo $y_{ir} = \mu_i + \varepsilon_{ir}$ $i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$ a nivel de significancia α si y solo si:

$$\left(\frac{n-m}{m-k}\right) * \left(\frac{\sum_{i=1}^{m}[(n_i)(y_{i.} - y_{i.})^2]}{\sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \bar{y}_{i.})^2}\right) \geq F_{(m-k;n-m;\alpha)}$$

Demostración. Se debe estimar cociente:

$$F_{(y)} = \left[\left(\frac{n-m}{m-k}\right)\left(\frac{SRC_{(0:\text{ Suma de Residuos al Cuadrado de modelo})} - SRC_{(\text{Suma de Residuos al Cuadrado})}}{SRC_{(\text{Suma de Residuos al Cuadrado})}}\right)\right]$$

Tal que $SRC_{(0:\text{ Suma de Residuos al Cuadrado de modelo})}$ $y_{ir} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \cdots + \beta_k x_{ik} + u_{ir}$ $i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$:

$$SRC_{(0:\text{ Suma de Residuos al Cuadrado de modelo})} = \sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \hat{y}_{i.})^2$$

Estimador de mínimos cuadrados de μ_i en modelo $y_{ir} = \mu_i + \varepsilon_{ir}$ $i = 1,2,3,4,5 \dots, m; r = 1,2,3,4,5, \dots, n_i$ es:

$$\tilde{\mu}_i = \bar{y}_{i.}\ i = 1,2,3,4,5 \dots, k;\ SRC_{(\text{Suma de Residuos al Cuadrado})} = \sum_{i=1}^{m}\sum_{r=1}^{n_i}(Y_{ir} - \bar{y}_{i.})^2$$

Vectores $\{Y_{ir} - \bar{y}_{i.}\}$ y $\{\bar{y}_{i.} - \hat{y}_{i.}\}$ son ortogonales:

$$\sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \bar{y}_{i.})(\bar{y}_{i.} - \hat{y}_{i.}) = \left[\sum_{i=1}^{m}(\bar{y}_{i.} - \hat{y}_{i.})\sum_{r=1}^{n_i}(y_{ir} - \bar{y}_{i.})\right] = 0$$

Por teorema de Pitágoras :

$$\|Y_{ir} - \overline{Y}_{i.}\|^2 + \left\|\overline{Y}_{i.} - \widehat{Y}_{i.}\right\|^2 = \left\|Y_{ir} - \widehat{Y}_{i.}\right\|^2$$

Tal que:

$$SRC_{(0:\ \text{Suma de Residuos al Cuadrado de modelo})} - SRC_{(\text{Suma de Residuos al Cuadrado})} = \sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \hat{y}_{i.})^2 - \sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \bar{y}_{i.})^2 = \sum_{i=1}^{m}[(n_i)(\bar{y}_{i.} - \hat{y}_{i.})^2]$$

Definición. Suma $\sum_{i=1}^{m}[(n_i)(\bar{y}_{i.} - \hat{y}_{i.})^2]$ es error por mala adecuación. La suma $\sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir} - \bar{y}_{i.})^2$ es llamado error puro. Resultados se escriben en ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher que puede ser insertada en:

Análisis de Varianza o ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher					
F. de V$_{\text{(Factor de}}$ **o VF**$_{\text{(Variation facto}}$	**Gl**$_{\text{(Grados de L}}$ **o Df**$_{\text{(Degrees of}}$	**SC**$_{\text{(Suma de Cua}}$ **o Sum Sq**$_{\text{(Sum}}$	**CM**$_{\text{(Cuadrado M}}$ **o Mean Sq**$_{\text{(Mea}}$	**F**$_{\text{(calculada)}}$ **o F value**$_{\text{(Calc}}$	**Pr(>F) o ρ − Value**
Error total	$(n-k)$	$\sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir}-\hat{y}_{i.})^2$			
Error por mala adecuación	$(m-k)_{\text{(gl numer}}$	$\sum_{i=1}^{m}[(n_i)(\bar{y}_{i.}-\hat{y}_{i.})^2]$	$MMA/gl_{\text{(nume}}$	$\frac{MMA_{\text{Adecuac}}}{MEP_{\text{Error}}}$	$\frac{gl_{\text{(numera}}}{gl_{\text{(denomir}}}$
Error puro	$(n-m)_{\text{(gl denor}}$	$\sum_{i=1}^{m}\sum_{r=1}^{n_i}(y_{ir}-\bar{y}_{i.})^2$	$MEP/gl_{\text{(denom}}$		

Fuente: Elaboración propia con base en modificaciones de (González B. G., Métodos estadísticos y principios de diseño experimental, 2010) y (Hurtado & Gómez, 2010)

Una vez construido el modelo de regresión se comprueba si hipótesis de linealidad, normalidad e independencia se cumplen. La matriz de covarianzas de errores es:

$$\text{Cov}(e) = \sigma^2(I - V) \text{ tal que } V = XX^t(X^tX)^{-1} \Rightarrow \text{Cov}(e_i) = \sigma^2(I - V_{ii})$$

Tal que $V_{(ii)}$ mide distancia entre X_i y $\overline{X}$. Para comparar los residuos suele ser más cómodo cambiarlos de escala, estandarizándolos o estudentizandolos. Los residuos estandarizados se definen:

$$r_i = \left[\frac{e_{(i)}}{s\sqrt{1} - V_{(ii)}}\right]$$

Los residuos estudentizados se estiman mediante:

$$t_i = \left[\frac{e_{(i)}}{s_{(i)}\sqrt{1} - V_{(ii)}}\right]$$

Siguen una ley t con $(n - k - 2)$ grados de libertad tal que $s_{(i)}$ son residuos de regresión cuando se excluye la i -ésima observación.

Se espera que datos correspondientes a observaciones se encuentren distribuidos en una región más o menos cercana, pero puede suceder que una o varias observaciones estén alejadas del resto de los datos. Estas observaciones pueden fluir mucho en modelo final. Se puede disponer de 100 observaciones y construir con ellos un modelo con propiedades debidas únicamente a dos puntos. Conocer si este tipo de puntos influye perjudicialmente en modelo permite mejorarlo.

6.3. Valores atípicos

La primera forma para determinar si un valor es atípico es mediante los residuos estudentizados. Se comparan valores de t_i con valores críticos de una ley t con $(n - k - 2)$ grados de libertad. Otra forma de conocer cuáles son puntos distantes es mediante distancia:

$$D_i^2 = \sum_{j=1}^{k} \left(\frac{b_j\,(x_{ij}\bar{x}_j}{\sqrt{MCE}}\right)^2 \quad i = 1,2,3,4,\ldots,n$$

El procedimiento consiste en ajustar el modelo, estimar $D_i^2, i = 1,2,3,4,\ldots,n$, y ordenarlas en forma ascendente de acuerdo con el D_i^2. Los puntos con alto D_i^2 son inusuales. Existen otros tipos de distancia que permiten la detección de valores atípicos y puntos influyentes a cada uno con distintas propiedades, pero todas siguen igual principio para identificar tales observaciones.

6.3.1. Identificación

Para la detección de la autocorrelación se emplea el estadístico de Durbin-Watson:

$$r_h = \frac{\sum e_i e_{i-h}}{\sum e_i^2}$$

Para la realización de pruebas estadísticas sobre la autocorrelación existen tablas, pero dar la siguiente regla empírica para detectar la autocorrelación de orden 1. Se construye:

$$d = 2(1 - r_1)$$

Si su valor es próximo a 2 no existirá autocorrelación, pero si d tiene un valor entre 0 y 2 será positiva y si el valor d esta entre 2 y 4 indica que correlación será negativa. En los programas informáticos se suele calcular su valor y el nivel de simplificación de la prueba. Alternativamente, se aplica el estadístico Box – Ljung para detectar la autocorrelación.

6.3.2. Tratamiento

Para explicar la evolución de variables que tiene un comportamiento en que aparece autocorrelación es conveniente usar métodos de análisis de series de tiempo, que permiten abordar de mantenimiento global el problema de construcción de modelos para estas variables. También, se pueden utilizar métodos especiales de regresión, como los mínimos cuadrados generalizados o modelos generalizados.

Por la dificultad que entraña la realización manual de cálculos; especialmente en determinación de potenciales problemas que modelo pudiera presentar. Se hace

mediante el empleo de programas estadísticos especializados, que facilitan su cálculo, correspondiendo al usuario la interpretación correcta de los resultados. Deberá notar que temas tratados sólo cubren la parte central de análisis de regresión.

6.4. Multicolinealidad

El término multicolinealidad se atribuye a Ragnar Frisch (, que asigna una relación lineal "perfecta" o exacta entre algunas o todas las variables explicativas de un modelo de regresión. En regresión con k variables que incluye las variables explicativas $(X_1, X_2, X_3, X_4, \dots, X_k)$ tal que existe una relación lineal exacta si se satisface la siguiente condición:

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_k X_k$$

Multicolinealidad incluye el caso de multicolinealidad perfecta, como ecuación anterior y, también, caso en que hay X variables intercorrelacionadas, pero no en forma perfecta:

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_k X_k + \nu_{i\ (\text{Término de error estocástico})} = 0$$

Para apreciar la diferencia entre multicolinealidades perfecta y menos que perfecta suponga; por ejemplo: $\lambda_2 \neq 0$:

$$X_{2i} = -\left[X_{1i}\left(\frac{\lambda_1}{\lambda_2}\right)\right] - \left[X_{3i}\left(\frac{\lambda_3}{\lambda_2}\right)\right] - \left[X_{4i}\left(\frac{\lambda_4}{\lambda_2}\right)\right] - \left[X_{5i}\left(\frac{\lambda_{15}}{\lambda_2}\right)\right] - \cdots - \left[X_{ki}\left(\frac{\lambda_k}{\lambda_2}\right)\right]$$

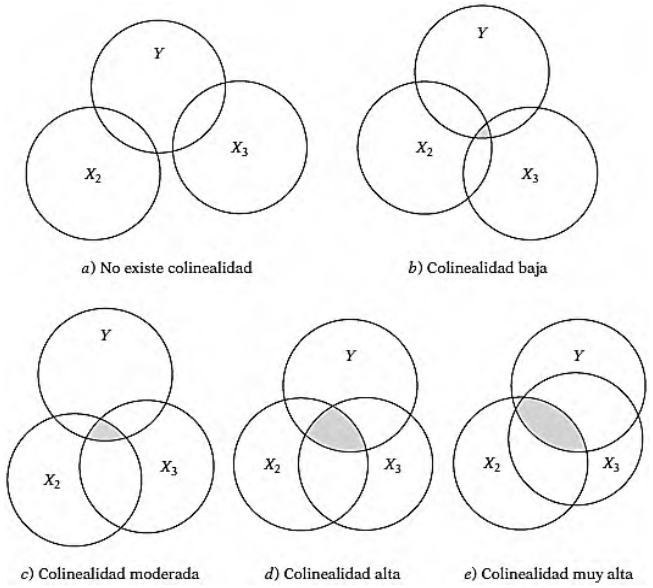
Muestra la forma como X_2 está exactamente relacionada de manera lineal con otras variables, o cómo se deriva de una combinación lineal de otras variables X. El coeficiente de correlación entre variable X_2 y combinación lineal del lado derecho de ecuación pasada está obligado a ser igual a uno.

Paralelamente, si $\lambda_2 \neq 0$, ecuación $\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_k X_k + \nu_{i\ (\text{Término de error estocástico})} = 0$ se escribe:

$$X_{2i} = -\left[X_{1i}\left(\frac{\lambda_1}{\lambda_2}\right)\right] - \left[X_{3i}\left(\frac{\lambda_3}{\lambda_2}\right)\right] - \left[X_{4i}\left(\frac{\lambda_4}{\lambda_2}\right)\right] - \left[X_{5i}\left(\frac{\lambda_{15}}{\lambda_2}\right)\right] - \cdots$$
$$- \left[X_{ki}\left(\frac{\lambda_k}{\lambda_2}\right) - \nu_{i\,(\text{Término de error estocástico})}\left(\frac{1}{\lambda_2}\right)\right]$$

Esto demuestra que X_2 no es combinación lineal exacta de otras X, pues está determinada por $\nu_{i\,(\text{Término de error estocástico})}$. El método algebraico para problema de multicolinealidad se expresa concisamente mediante un diagrama de Ballentine:

Figura 6.2. Diagrama de Ballentine



Fuente: (Gujarati & Porter, 2010)

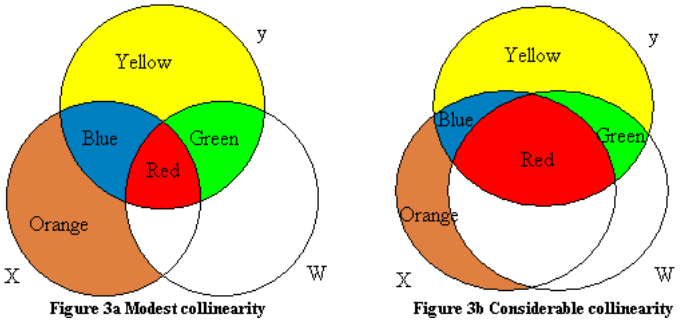
Los círculos Y, X_2 y X_3 indican variaciones en $Y_{(\text{variable dependiente})}$ en X_2 y X_3, variables explicativas. El grado de colinealidad se mide por magnitud de intersección, área sombreada, de círculos X_2 y X_3.

En figuras 6.2a), no hay intersección entre X_2 y X_3 tal que no hay colinealidad; de 6.2 b) a 6.2e), el grado de colinealidad va de "bajo" a "alto", pues entre mayor

sea la intersección entre X_2 y X_3 -entre mayor sea el área sombreada- mayor será el grado de colinealidad. Si X_2 y X_3 estuvieran superpuestos completamente o si X_2 estuviera por completo dentro de X_3, o viceversa, la colinealidad sería perfecta.

En otras figuras:

Figura 6.3. Diagrama de representación



Fuente: Elaboración propia con base en modificaciones de (Gujarati & Porter, 2010)

Multicolinealidad, como se define, refiere sólo a relaciones lineales entre las variables X tal que no aplica a las relaciones no lineales entre ellas:

$$Y_{i\ (\text{Costo de producción total})} = \beta_1 + \beta_2 X_{i\ (\text{Producción})} + \beta_3 X^2_{i\ (\text{Producción al cuadrado})} + \beta_4 X^3_{i\ (\text{Producción al cubo})} + u_{i\ (\text{Término de error estocástico})}$$

En econometría se presenta una fuerte correlación entre variables explicativas del modelo. La correlación ha de ser fuerte, pues siempre existirá correlación entre dos variables explicativas en un modelo tal que la no correlación de dos variables es un proceso idílico únicamente se podría hallar en condiciones de laboratorio. Modelos como estos, no violan el supuesto de no multicolinealidad.

Sin embargo, en aplicaciones concretas, el coeficiente de correlación medido de forma convencional demostrará que $X_{i\ (Producción)}$, $X^2_{i\ (Producción\ al\ cuadrado)}$ y $X^3_{i\ (Producción\ al\ cubo)}$ están altamente correlacionadas, que dificultará estimar parámetros de ecuación anterior con mayor precisión o errores estándar pequeños.

El modelo clásico de regresión lineal supone no hay multicolinealidad entre X_i, pues si la multicolinealidad es perfecta en sentido de $\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_k X_k$ tal que coeficientes de regresión de variables X son indeterminados y errores estándar infinitos.

Si multicolinealidad es menos que perfecta, como $\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + \cdots + \lambda_k X_k + \nu_{i\ (Término\ de\ error\ estocástico)} = 0$, coeficientes de regresión; aunque sean determinados, poseen grandes errores estándar en relación con coeficientes mismos, que significa que coeficientes no pueden ser estimados con gran precisión o exactitud.

La multicolinealidad puede ser por factores:

a) **Método de recolección de información**. Por ejemplo, la obtención de muestras en un intervalo limitado de valores tomados por regresoras en población.
b) **Restricciones en modelo o población objeto de muestreo**. Por ejemplo, en regresión del consumo de electricidad sobre ingreso (X_2) y tamaño de viviendas (X_3) hay una restricción física en población, pues familias con ingresos más altos suelen habitar viviendas más grandes que familias con ingresos más bajos.
c) **Especificación del modelo**. Por ejemplo, adición de términos polinomiales a un modelo de regresión, específicamente cuando rango de la variable X es pequeño.
d) **Modelo sobredeterminado**. Esto sucede cuando el modelo tiene más variables explicativas que número de observaciones. Esto puede suceder

en investigación médica, en ocasiones hay un número reducido de pacientes sobre quienes se reúne información respecto a gran número de variables.

e) **Regresoras de modelo compartan una tendencia común**. Datos de series de tiempo aumenten o disminuyan en el tiempo tal que en regresión del gasto de consumo sobre el ingreso, riqueza y población regresoras, ingreso, riqueza y población quizás todas crezcan con el tiempo a una tasa aproximadamente igual; por lo que, presentaría colinealidad entre dichas variables.

Si se satisfacen supuestos del modelo clásico, estimadores de MCO de coeficientes de regresión son MELI o MEI si se añade el supuesto de normalidad. Puede demostrarse que, aunque la multicolinealidad sea muy alta, como en el caso de casi multicolinealidad, los estimadores de MCO conservarán la propiedad MELI. En los casos de casi o alta multicolinealidad es probable que se presenten consecuencias:

a) Estimadores de MCO son MELI, presentan varianzas y covarianzas grandes que dificultan estimación precisa.

Para ver varianzas y covarianzas grandes, para modelo $y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i$, varianzas, covarianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$ están dadas por:

$$\mathrm{Var}(\hat{\beta}_2) = \left[\frac{\sigma^2}{\sum\left[(x_{2i}^2)\left(1 - r_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}^2\right)\right]}\right]$$

$$\mathrm{Var}(\hat{\beta}_3) = \left[\frac{\sigma^2}{\sum\left[(x_{3i}^2)\left(1 - r_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}^2\right)\right]}\right]$$

$$\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \left[\frac{(-r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)})(\sigma^2)}{\left(\sqrt{\sum[(x^2_{2i})(x^2_{3i})]}\right)\left(1 - r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}\right)}\right]$$

Con base en estas ecuaciones, se desprende que, a medida que r_{23} tiende a 1 tal que a medida que aumenta la colinealidad lo hacen varianzas de estimadores y, en el límite, cuando $r_{23} = 1$, son infinitas. En caso de última ecuación, a medida que r_{23} aumenta a 1, la covarianza de dos estimadores aumenta en valor absoluto tal que $\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \mathrm{Cov}(\hat{\beta}_3, \hat{\beta}_2)$. La velocidad con que se incrementan varianzas y covarianzas se mide mediante (FIV):

$$\mathrm{FIV}_{(\text{Factor inflacionario})} = \left[\frac{1}{\left(1 - r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}\right)}\right]$$

$\mathrm{FIV}_{(\text{Factor inflacionario})}$ muestra la forma como la varianza de un estimador se infla por presencia de la multicolinealidad, pues a medida que $r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}$ se acerca a 2, $\mathrm{FIV}_{(\text{Factor inflacionario})}$ se acerca a infinito tal que medida que grado de colinealidad aumenta varianza de un estimador también y, en límite, se vuelve infinita.

Si no existe colinealidad entre X_2 y X_3, $\mathrm{FIV}_{(\text{Factor inflacionario})}$ será 1 tal que con esta definición, $\mathrm{Var}(\hat{\beta}_2) = \left[\frac{\sigma^2}{\sum\left[(x^2_{2i})\left(1 - r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}\right)\right]}\right]$ y $\mathrm{Var}(\hat{\beta}_3) = \left[\frac{\sigma^2}{\sum\left[(x^2_{3i})\left(1 - r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}\right)\right]}\right]$, será $\mathrm{FIV}_{(\text{Factor inflacionario})} = 1$:

$$\text{Var}(\hat{\beta}_2) = \left[\left(\frac{\sigma^2}{\sum[(x_{2i}^2)]}\right)(\text{FIV}_{(\text{Factor inflacionario})})\right]$$

$$\text{Var}(\hat{\beta}_3) = \left[\left(\frac{\sigma^2}{\sum[(x_{3i}^2)]}\right)(\text{FIV}_{(\text{Factor inflacionario})})\right]$$

Esto demuestra que varianzas $\hat{\beta}_2$ y $\hat{\beta}_3$ son directamente proporcionales a $\text{FIV}_{(\text{Factor inflacionario})}$. Si se extiende a modelo con k variables:

$$\text{Var}(\hat{\beta}_j) = \left[\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)\left(\frac{1}{1-R_j^2}\right)\right]$$

b) Con base en consecuencia anterior, intervalos de confianza tienden a ser mucho más amplios, propicia una aceptación más fácil de "hipótesis nula cero"; es decir, el verdadero coeficiente poblacional es cero.

Con base en errores estándar grandes, los intervalos de confianza para los parámetros poblacionales relevantes tienden a ser mayores:

Efecto de incrementar la colinealidad sobre intervalo de confianza a 95% para β_2: $\hat{\beta}_2 \pm 1.96\ ee(\hat{\beta}_2)$	
Valor $r_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}^2$	**Intervalos de confianza a 95% para β_2**
0.00	$\hat{\beta}_2 \pm 1.96\sqrt{\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)}$
0.50	$\hat{\beta}_2 \pm 1.96\sqrt{(1.33)}\sqrt{\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)}$
0.95	$\hat{\beta}_2 \pm 1.96\sqrt{(10.26)}\sqrt{\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)}$

Efecto de incrementar la colinealidad sobre intervalo de confianza a 95% para β_2: $\hat{\beta}_2 \pm 1.96\ ee(\hat{\beta}_2)$	
Valor $r^2_{23(\text{Coeficiente de correlación entre } x_2 - x_3)}$	**Intervalos de confianza a 95% para β_2**
0.995	$\hat{\beta}_2 \pm 1.96\sqrt{(100.00)}\sqrt{\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)}$
0.999	$\hat{\beta}_2 \pm 1.96\sqrt{(500.00)}\sqrt{\left(\frac{\sigma^2}{\sum[(x_j^2)]}\right)}$

Fuente: Elaboración propia con base en modificaciones de (Gujarati & Porter, 2010)

En casos de alta multicolinealidad, datos muestrales pueden ser compatibles con un diverso conjunto de hipótesis. De ahí que aumente la probabilidad de aceptar hipótesis falsa; es decir, error tipo II.

c) Con base en consecuencia primera, razón t de uno o más coeficientes tiende a ser estadísticamente no significativa.

Para probar hipótesis nula que, por ejemplo $H_0: \hat{\beta}_2 = 0$ se usa razón t tal que $t_{(\text{Calculado})} = \left[\frac{\hat{\beta}_2}{ee(\hat{\beta}_2)}\right]$ vs $t_{(\text{Tablas})}$; aunque, en casos de alta colinealidad, errores estándar estimados aumentan drásticamente, que disminuye los valores t. Por consiguiente, no se rechaza cada vez con mayor facilidad H_0, que el verdadero valor poblacional relevante es cero.

d) Aún con razón t de uno o más coeficientes sea estadísticamente no significativa, R^2, la medida global de bondad de ajuste puede ser muy alta.

Si modelo de regresión lineal con k variables es:

$$Y_{i\ (\text{Costo de producción total})} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \cdots + \beta_k X_{ki} + u_{i\ (\text{Término de error estocástico})}$$

En casos de alta colinealidad es posible encontrar que uno o más coeficientes parciales de pendiente son, de manera individual, no significativos estadísticamente con base en la prueba t. Aun así, R^2 puede ser tan alto, quizás superior a 0.9, que, con base en razón F, es posible rechazar convincentemente hipótesis $H_0: \beta_2 = \beta_3 = \beta_4 = \cdots = \beta_k = 0$. Esta es una de las señales de multicolinealidad, pues valores t son no significativos pero tiene valor R^2 global alto y razón F significativo.

e) Estimadores de MCO y sus errores estándar son sensibles a pequeños cambios en datos.

Si multicolinealidad no es perfecta, es posible estimación de coeficientes de regresión; sin embargo, sus estimaciones y errores estándar se tornan muy sensibles al más ligero cambio de datos. En presencia de una alta colinealidad, no se pueden estimar los coeficientes de regresión individuales en forma precisa, pero sus combinaciones lineales se estiman con mayor exactitud.

6.5. Correlación en series de tiempo

Es "correlación entre miembros de series de observaciones ordenadas en tiempo, datos de series de tiempo, o en espacio, datos de corte transversal". Autocorrelación o dependencia secuencial es una herramienta estadística usada frecuentemente en procesado de señales. La función de autocorrelación se define como la correlación cruzada de señal consigo misma. La función de autocorrelación resulta de gran utilidad para encontrar patrones repetitivos dentro de una señal, como periodicidad de señal enmascarada bajo el ruido o para identificar la frecuencia fundamental de una señal que no contiene dicha componente, pero aparecen numerosas frecuencias armónicas de esta. Los procesos de raíz unitaria, autorregresivos, tendencia estacionaria y Modelos de Medias Móviles son ejemplos de procesos con autocorrelación. En contexto de

regresión, el modelo clásico de regresión lineal supone que no existe tal autocorrelación en perturbaciones $u_{i\ (\text{Término de error estocástico})}$:

$$\text{Cov}\left(u_{i\ (\text{Término de error estocástico})}, u_{j\ (\text{Término de error estocástico})} \middle| x_i, x_j\right)$$
$$= E\left(u_{i\ (\text{Término de error estocástico})}, u_{i\ (\text{Término de error estocástico})}\right) = 0\ i \neq j$$

El modelo clásico supone que el término de perturbación relacionado con una observación cualquiera no recibe influencia del término de perturbación relacionado con cualquier otra observación.

Por ejemplo:

Si existe información productiva trimestral de series de tiempo que es inferior en este ciclo, no hay razón para esperar que sea baja en siguiente periodo. En forma similar, si se trata con información de corte transversal que implica la regresión del gasto de consumo familiar sobre el ingreso familiar, no se esperaría que efecto de incremento en ingreso de una familia sobre su gasto de consumo incida en gasto de consumo de otra.

Además, si se trata con información de corte transversal que implica la regresión del gasto de consumo familiar sobre ingreso familiar, no esperaría que efecto de un incremento en ingreso de una familia sobre su gasto de consumo incida en gasto de consumo de otra.

No obstante, si existe tal dependencia, hay autocorrelación:

$$E\left(u_{i\ (\text{Término de error estocástico})}, u_{j\ (\text{Término de error estocástico})}\right) \neq 0\ i \neq j$$

Tal que la interrupción ocasionada por una huelga este trimestre puede afectar muy fácilmente la producción del siguiente trimestre o incrementos del gasto de consumo de una familia pueden muy bien inducir a otra familia a aumentar su gasto de consumo para no quedar rezagada.

La correlación serial sucede por varias razones:

6.5.1. Incercia o pasividad

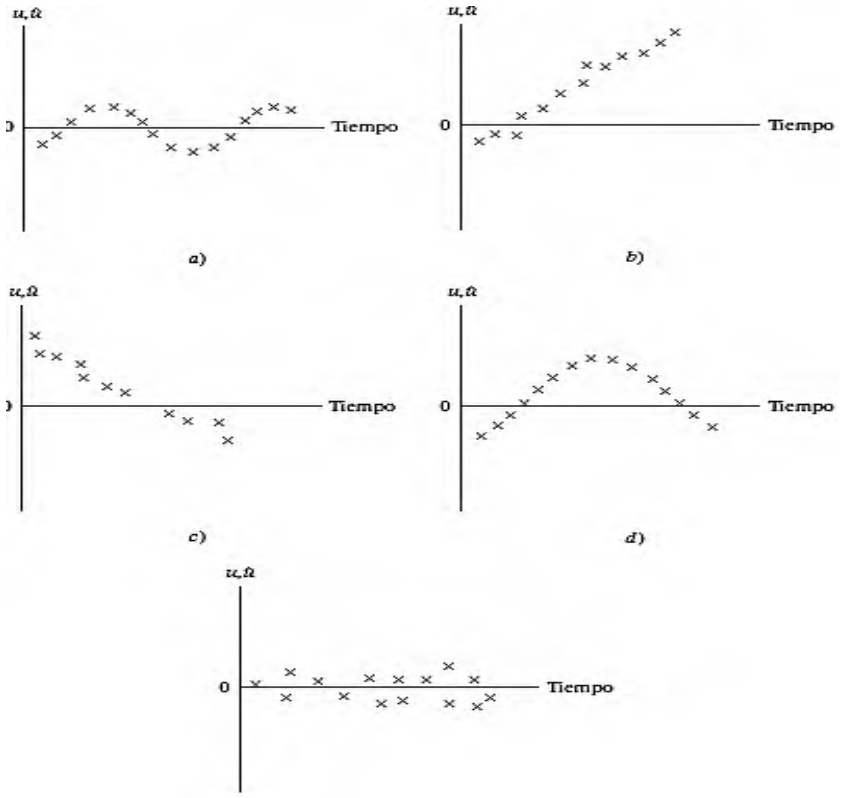
Las series de tiempo como PNB, índices de precios, producción, empleo y desempleo presentan ciclos económicos. A partir del fondo de la recesión, cuando se inicia la recuperación económica, la mayoría de estas series empieza a moverse hacia arriba. En este movimiento ascendente, el valor de una serie en un punto del tiempo es mayor que su valor anterior. Se genera un "impulso" en ellas y continuará hasta que suceda otra cosa.

Por ejemplo: un aumento en tasa de interés, impuestos o ambos para reducirlo. Por consiguiente, es probable que en regresiones que consideran datos de series de tiempo, las observaciones sucesivas sean interdependientes.

6.5.2. Sesgo de especificación: caso de variables excluidas

Con frecuencia el investigador empieza con un modelo de regresión razonable que puede no ser "perfecto". Después del análisis de regresión, el investigador haría el examen post mortem para ver si los resultados coinciden con las expectativas a priori. Por ejemplo, el investigador graficaría residuos $\hat{u}_{i\ (\text{Término de error estocástico})}$ obtenidos de regresión ajustada y observaría patrones (6.4a a 6.4d):

Figura 6.4. Gráfica de residuos



Fuente: (Gujarati & Porter, 2010)

Estos residuos, representaciones de u_i (Término de error estocástico), pueden sugerir la inclusión de algunas variables originalmente candidatas pero que no se incluyeron en el modelo por diversas razones. Es el caso del sesgo de especificación ocasionado por variables excluidas. Con frecuencia, la inclusión de tales variables elimina patrón de correlación observado entre los residuales.

Por ejemplo:

$$Y_{i\ (\text{Cantidad de res demandada})} = \beta_1 + \beta_2 X_{2t\ (\text{Precio de carne de res y t tiempo})} + \beta_3 X_{3t\ (\text{Ingreso de consumidor y t tiempo})} + \beta_4 X_{4t\ (\text{Precio del cerdo y t tiempo})} + u_{i\ (\text{Término de error estocástico})}$$

No obstante, se efectúa por alguna razón:

$$Y_{i\ (\text{Cantidad de res demandada})} = \beta_1 + \beta_2 X_{2t\ (\text{Precio de carne de res y t tiempo})} + \beta_3 X_{3t\ (\text{Ingreso de consumidor y t tiempo})} + \nu_{t\ (\text{Término de error estocástico})}$$

Si el primer modelo es "correcto", el "verdadero" o relación auténtica es el segundo modelo tal que equivale a permitir $\nu_{t\ (\text{Término de error estocástico})} = \beta_4 X_{4t\ (\text{Precio del cerdo y t tiempo})}$; por lo tanto, en la medida en que el precio del cerdo afecte el consumo de carne de res, el término de error o de perturbación $\nu_{t\ (\text{Término de error estocástico})}$ reflejará un patrón sistemático que crea una falsa autocorrelación.

Una prueba sencilla de esto sería llevar a cabo primer modelo y, también, segundo modelo tal que se vea si la autocorrelación observada en segundo modelo, de existir, desaparece cuando se efectúa el primero.

6.5.3. Sesgo de especificación: forma funcional incorrecta

Suponga que modelo "verdadero" o correcto en un estudio de costo-producción es:

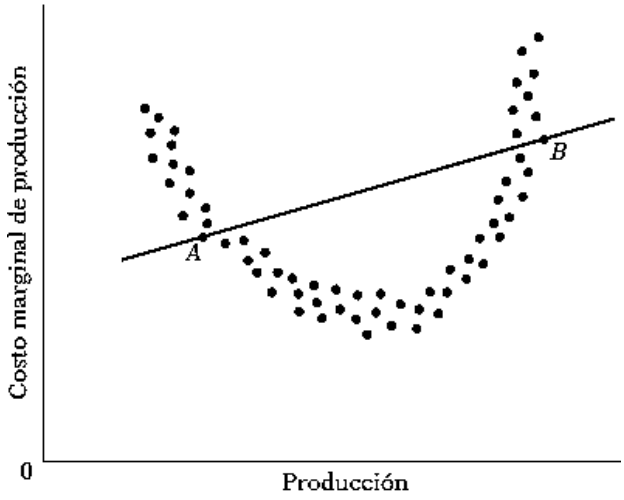
$$Y_{i\ (\text{Costo marginal})} = \beta_1 + \beta_2 X_{i\ (\text{Producción})} + \beta_3 X^2_{i\ (\text{Producción al cuadrado})} + u_{i\ (\text{Término de error estocástico})}$$

Este modelo ajustado es:

$$Y_{i\ (\text{Costo marginal})} = \alpha_1 + \alpha_2 X_{i\ (\text{Producción})} + u_{i\ (\text{Término de error estocástico})}$$

Tal que la curva de costo marginal que corresponde al "verdadero" modelo con curva de costo lineal "incorrecta" es:

Figura 6.5. Curva de costo marginal



Fuente: (Gujarati & Porter, 2010)

Entre puntos A y B de curva de costo marginal lineal sobreestimará consistentemente el costo marginal verdadero, mientras que más allá de estos puntos, lo subestimará consistentemente. Este resultado es de esperarse porque el término de perturbación $v_{i\ (\text{Término de error estocástico})} = u_{i\ (\text{Término de error estocástico})}$ tal que capta el efecto sistemático del término producción sobre el costo marginal. $v_{i\ (\text{Término de error estocástico})}$ reflejará autocorrelación por el uso de una forma funcional incorrecta.

6.5.4. Fenómenos de telaraña

La oferta de muchos productos agrícolas refleja el llamado fenómeno de telaraña, en donde la oferta reacciona al precio con un rezago de un periodo, pues debido

a instrumentación de decisiones de oferta tarda algún tiempo, llamado periodo de gestación. Por tanto, en siembra de cultivos al principio de año, agricultores reciben influencia del precio prevaleciente el año pasado tal que su función de oferta es:

$$Y_{t\ (\text{Oferta t})} = \beta_1 + \beta_2 P_{(t-1)\ (\text{Producción rezagada})} + u_{t\ (\text{Término de error estocástico})}$$

Suponga que a final de periodo t, el precio $P_{(t)\ (\text{Producción actual})}$ es inferior a $P_{(t-1)\ (\text{Producción rezagada})}$. En consecuencia, es muy probable que en periodo $(t+1)$ agricultores decidan producir menos de lo producido en periodo t. En esta situación no esperaría que perturbaciones $u_{i\ (\text{Término de error estocástico})}$ sean aleatorias, pues si agricultores producen excedentes en el año t, es probable que reduzcan su producción en $(t+1)$ y, sucesivamente, para generar un patrón de telaraña.

6.5.5. Rezagos

En una regresión de series de tiempo del gasto de consumo sobre ingreso no es raro hallar que gasto de consumo en periodo actual dependa, entre otras cosas, del gasto de consumo del periodo anterior:

$$Y_{t\ (\text{Consumo t})} = \beta_1 + \beta_2 P_{(t)\ (\text{Ingreso t})} + \beta_3 C_{(t-1)\ (\text{Consumo rezagado})} + u_{t\ (\text{Término de error estocástico})}$$

Esta función es autorregresiva, pues es una variable explicativa es el valor rezagado de la variable dependiente. Los consumidores no cambian sus hábitos de consumo fácilmente por razones psicológicas, tecnológicas o institucionales. Si se ignora el término $C_{(t-1)\ (\text{Consumo rezagado})}$, $u_{t\ (\text{Término de error estocástico})}$ resultante reflejará un patrón sistemático debido a influencia del consumo rezagado en consumo actual.

6.5.6. "Manipulación" de datos

En análisis empírico con frecuencia se "manipulan" datos simples. Por ejemplo: en regresiones de series de tiempo con datos trimestrales, por lo general provienen de datos mensuales a los que se agregan simplemente observaciones de tres meses y se divide entre 3. Este procedimiento de promediar cifras suaviza en cierto grado datos al eliminar fluctuaciones en datos mensuales.

Por consiguiente, la gráfica referente a datos trimestrales aparece mucho más suave que la que contiene datos mensuales tal que este suavizamiento puede, por sí mismo, inducir un patrón sistemático en perturbaciones, lo que agrega autocorrelación. Otra fuente de manipulación es interpolación o extrapolación de datos. Estas técnicas de "manejo" podrían imponer sobre datos un patrón sistemático que quizá no estaría presente en datos originales

6.5.7. Transformación de datos

Considere:

$$Y_{t\ (\text{Gasto de onsumo t})} = \beta_1 + \beta_2 X_{(t)\ (\text{Ingreso t})} + u_{t\ (\text{Término de error estocástico})}$$

Esta ecuación es válida para cada periodo y, también, es para el periodo anterior $(t - 1)$. Así, ecuación forma de nivel es:

$$Y_{t-1\ (\text{Consumo rezagado t}-1)} = \beta_1 + \beta_2 X_{(t-1)\ (\text{Ingreso rezagado t}-1)} + u_{t-1\ (\text{Término de error estocástico rezagado t}-1)}$$

En este caso están rezagados un periodo. Si restan las ecuaciones anteriores se obtiene ecuación forma en primeras diferencias:

$$\Delta_{\text{(Operador de primeras diferencias)}} Y_{t\ \text{(Incremento de consumo t)}}$$
$$= \beta_2 \Delta_{\text{(Operador de primeras diferencias)}} X_{\text{(t) (Incremento de ingreso t)}}$$
$$+ \Delta_{\text{(Operador de primeras diferencias)}} u_{t\ \text{(Incremento de término de error estocástico)}}$$

Tal que, el modelo dinámico de regresión, modelo con regresadas rezagadas, es:

$$\Delta_{\text{(Operador de primeras diferencias)}} Y_{t\ \text{(Incremento de consumo t)}}$$
$$= \beta_2 \Delta_{\text{(Operador de primeras diferencias)}} X_{\text{(t) (Incremento de ingreso t)}}$$
$$+ \Delta_{\text{(Operador de primeras diferencias)}} \nu_{t\ \text{(Incremento de término de error estocástico)}}$$

Si $Y_{t\ \text{(Gasto de onsumo t)}} = \beta_1 + \beta_2 X_{\text{(t) (Ingreso t)}} + u_{t\ \text{(Término de error estocástico)}}$ satisface supuestos usuales de MCO, específicamente inexistencia de autocorrelación, se prueba probar que término de error $\nu_{t\ \text{(Incremento de término de error estocástico)}}$ en última ecuación está autocorrelacionado.

6.5.8. No estacionalidad

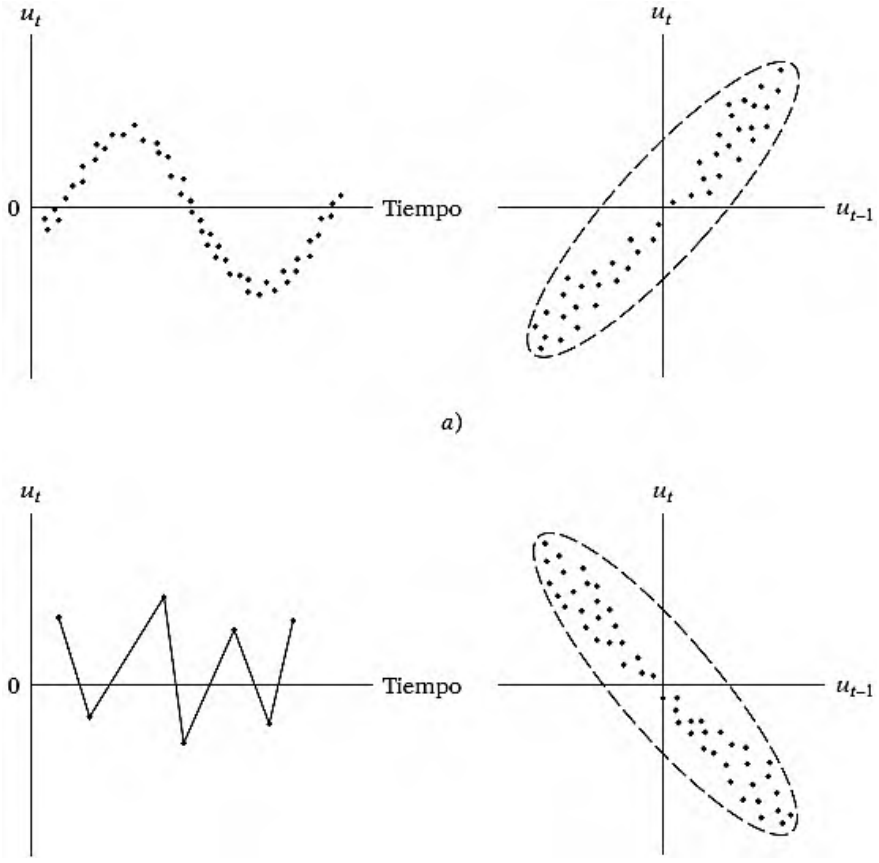
Al trabajar con datos de series de tiempo, quizá habría que averiguar si una determinada serie de tiempo es estacionaria. De manera informal, una serie de tiempo es estacionaria si sus características, como media, varianza y covarianza, son invariantes respecto del tiempo; es decir, no cambian en relación con el tiempo. Si no es así, tenemos una serie de tiempo no estacionaria.

en un modelo de regresión como $Y_{i\ \text{(Costo marginal)}} = \alpha_1 + \alpha_2 X_{i\ \text{(Producción)}} + u_{i\ \text{(Término de error estocástico)}}$ es muy probable que Y, X sean no estacionarias y, por consiguiente, que el error u también sea no estacionario. En ese caso, el término de error mostrará autocorrelación.

En conclusión, existen varias razones por las que término de error en un modelo de regresión pueda estar autocorrelacionado. Autocorrelación puede ser positiva,6.6a o negativa; aunque, por lo general, la mayoría de las series de tiempo

económicas muestra autocorrelación positiva, pues casi todas se desplazan hacia arriba o hacia abajo en extensos periodos y no exhiben un movimiento ascendente y descendente constante, 6.6b:

Figura 6.6. Gráfica de autocorrelación



Fuente: (Gujarati & Porter, 2010)

BIBLIOGRAFÍA

Adams, J. A., Benitez, M. D., Guidek, R. C., & Domínguez, G. A. (2016). *Enseñanza de Programación Lineal y Juegos de Empresa*. Asunción, Paraguay: Universidad Autónoma de Paraguay.

Adhikari, R., & Dandekar, R. (2012). Optimization models for control of parasitic diseases. *Mathematical and Computer Modelling, 56(1-2)*, 233-250.

Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). *Network Flows. Theory, Algorithms and Applications.* Upper Saddle River. New Jersey, USA.: Prentice-Hall, Inc.

Alarcón, S. (1998). La programación estocástica discreta como instrumento de planificación de plantaciones de olivos en Castilla-La Mancha. *Economía Agraría N0. 183*, 173-200.

Alidaee, A., & Aneja, Y. P. (2018). *Non-predetermined goal programming. In Operations research: A model-based approach.* New Jersey, U.S.A.: Springer, Cham.

Ancco, P. R., Chalco, C. D., Clavijo, T. P., Delgado, H. P., Huarac, S. Y., & Ugarte, F. D. (2020). Métodos para el procesamiento de imágenes digitales. *Universidad Nacional de San Antonio Abad del Cusco, Cusco Perú*, 1-8.

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2006). *Estadística para administración y economía.* Col. Cruz Manca, Santa Fe, México, D. F.: Cengage Learning Editores, S. A. de C. V.

Arévalo, J. J., Castro, A., & Villa, É. (2002). Un análisis del ciclo económico en competencia imperfecta. *Revista de Economía Institucional, vol. 4, núm. 7, segundo semestre*, 11-39 Pp.

Arriaza, G. A., Fernández, P. F., López, S. M., Muñoz, M. M., Pérez, P. S., & Sánchez, N. A. (2008). *Estadística Básica con R y R-Commander.* Cádiz, España: Servicio de Publicaciones de la Universidad de Cádiz.

Atucha, A. J., & Gualdoni, P. (2018). *El funcionamiento de los mercados.* Mar de Plata, Argentina: Facultad de Ciencias Económicas y Sociales, Universidad Nacional de Mar del Plata.

Avdoshin, S. M., & Pesotskaya, E. Y. (2011). Software risk management. *National Research University Higher School of Economics* , 1-6 Pp.

Baquela, E. G., & Redchuk, A. (2013). *Optimización Matemática con R. Volumen I: Introducción al modelado y resolución de problemas.* Madrid, España: Bubok Publishing S. L.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2009). *Nonlinear programming: Theory and applications.* Hoboken, New Jersey, U.S.A.: John Wiley & Sons, Inc, Publication.

Bertsimas, D., & Tsitsiklis, J. N. (1997). *Introduction to Linear Optimization.* Belmont, Mass. U.S.A.: Athena Scientific, Belmont, Massachusetts.

Bertsimas, D., Teo, C., & Vohra, R. (1999). On dependent randomized rounding algorithms. *ELSEVIER*, 105 –114.

Bhatt, A., Das, V., & Sundaram, V. (2011). Economic analysis of malaria control programs: A review of the literature. *Malaria Journal, 10(1)*, 202.

Birge, J. R., & Louveaux, F. (2011). *Introduction to Stochastic Programming.* New York, USA: Springer Science+Business Media, .

Brambila, P. J. (2011). *Bioeconomía: Instrumentos para su análisis económico.* Texcoco, estado de México. México: Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA) y Colegio de Postgraduados (COLPOS).

Broz, D., Mac Donagh, P., Arce, J., & Yapura, P. (2017). La Investigación Opetrativa, la Ingeniería Forestal y los Problemas Sectoriales: Ante la Necesidad de un Cambio de Paradigma. *Yvyraretá*, 64-72.

Burbano, R., Espinosa, A. M., & Horna, L. (2018). *Economía para matemáticos.* Quito, Ecuador: Escuela Politécnica Nacional (EPN).

Burkard, R., Dell'Amico, M., & Martello, S. (2012). *Operations research: Theory and practice.* New Jerse, U.S.A.: Springer Science & Business Media.

Bustos, F. E. (2023). *La importancia del teorema de suficiencia de Kuhn-Tucker en la tarea de decisiones organizacionales.* Carácas, Venezuela: Escuela Superior de Cómputo de Venezuela.

Cabrera, G. E., & Díaz, G. E. (2021). *Manual de uso de Jupyter Notebook para aplicaciones docentes.* Madrid, España: Universidad Complutense de Madrid.

Camps, P. R., Casillas, S. L., Costal, C. D., Gibert, G. M., Martín, E. C., & Pérez, M. O. (2005). *Software libre. Bases de datos.* Av. Tibidado, 39-43, Barcelona España: Fundació per a la Universitat Oberta de Catalunya. Eureca Media, SL.

Capa, S. H. (2015). *Probabilidades y estadística para una gestión científica de la información.* Quito, Ecuador: Escuela Politécnica Nacional .

Carballo, R. E., & Peñate, P. E. (2014). *Benchmarking de Herramientas Forenses Móviles.* Managua: COMPDES, UNAN, Managua.

Carro, P. R. (2014). *Investigación de operaciones en administración. 1ra Edición.* Mar de Plata, Provincia de Buenos Aires, Argentina: Facultad de Ciencias Económicas y Sociales. Universidad Nacional de Mar de Plata.

Cestero, E. V., & Caballero, A. M. (2017). *Data Science y Redes Complejas. Métodos y aplicaciones.* Tomás Bretón, 21-28045, Madrid, España: Editorial Centro de Estudios Ramón Areces, S. A.

Challenger, P. I., Díaz, R. Y., & Becerra, G. R. (Abril-Junio, 2014). El lenguaje de programación Python. *Ciencias Holguín, vol. XX, núm. 2.*, 1-13.

Dantzig, G. B. (1963). *Linear Programming and Extensions.* Chichester, West Sussex, U.S.A.: Princeton University Press.

Dantzig, G. B., & Wolfe, P. (1954). The decomposition principle for linear programming. *Operations Research, 2(1)*, 76-81.

Daza, P. J. (2006). *Estadística Aplicada con Excel.* Lima, Perú: Grupo Editorial Megabyte s.a.c.

De los Reyes García, M. M., & Romero, C. J. (2004). *Investigación de operaciones I.* UAM Unidad Azcapotzalco. Av. San Pablo 180. Col.

Reynosa Tamaulipas. Delegación Azcapotzalco. México, D. F.: División de Ciencias Básicas e Ingeniería. Departamento de Sistemas. Universidad Autónoma Metropolitana-Azcapotzalco.

De Mendiburu, F. (2007). *Análisis estadístico con R.* Av. Túpac Amaru 210, Rímac 15333, Lima, Perú: Sección de extensión universitaria y proyección social. Centro de estudiantes de facultad de ingeniería económica y ciencias sociales. Universidad Nacional de Ingeniería.

De Mendiburu, F. (2017). *Tutorial de agricolae (Versión 1.2-8).* La Molina-Perú: Departamento Academico de Estadísstica e Informática de la Facultad de Economía y Planificación. Universidad Nacional Agraria La Molina-Perú.

Delgado, Q. S. (2023). *Aprende Python.* Ciudad de México, México: Mundi Prensa.

Deng, Z. X., & Wang, X. (2022). A survey on non-predetermined goal programming. *European Journal of Operational Research, 297(2)*, 1115-1133.

Devia, J., Azacon, J., López, N., Villarroel, Z., Carvajal, A., Rodríguez, B., . . . Suárez, F. (2012). *Software TORA.* Maturín, Venezuela: Núcleo de Monagas. Universidad de Oriente de Venezuela.

Di Perro, M. (2017). What is the Blocchaing? *SChool of Computing at DePaul University, Chicago, IL, USA*, 92-95.

Díaz, C. M., & Gómez, M. P. (2011). *Comportamiento de volatilidad del tipo de México 1970 a 2010.* Michoacán, México: Universidad Michoacana de San Nicolás de Hidalgo.

DiLorenzo, T. J. (1992). El Mito del Monopolio Natural. *The Review of Austrian Economics Vol. 9, No. 2* , 1-15 Pp.

Eiselt, H. A., & Sandblom, C. L. (2009). *Integer programming: Theory and practice.* New Jersey, U.S.A.: Springer.

Engler, P. A. (2009). *Estrategías del manejo del riesgo.* Quilamapu, Chillán, Chile: TRAMA Impresores S. A.

Galindo, D. T. (2006). *Estadística. Métodos y Aplicaciones.* Quito, Ecuador: Prociencia Editores.

Garcés, R. A. (2020). *Optimización convexa. Aplicaciones en operación y dinámica de sistemas de potencia.* Pereira, Colombia: Universidad Tecnológica de Pereira, Pereira, Colombia.

García, M. A., & García, S. (2015). Investigación de operaciones en el control de enfermedades parasitarias. *Revista de Investigación Académica, 26*, 1-12.

García, M. M., & Romero, C. J. (2004). *Investigación de operaciones 1. 1ra parte.* Delegación Azcapotzalco, D. F.: Universidad Autónoma Metropolitana (UAM-Azcapotzalco).

Gen, M., & Cheng, R. (2000). *Genetic algorithms and engineering optimization.* New York, U.S.A.: Wiley-Interscience Publication.

Gómez, G. M., Danglot, B. C., & Vega, F. L. (2003). Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. *Revista Mexicana de Pediatría*, 10.

Gómez, P. M. (2006). *Introducción a la microecconomía.* Barcelona, España: Universidad de Barcelona.

González, B. G. (1985). *Métodos Estadísticos y Principios de Diseño Experimental.* Quito, Ecuador: Facultad de Ciencias Agrícolas. Universidad Central del Ecuador.

González, B. G. (2010). *Métodos estadísticos y principios de diseño experimental.* Quito, Ecuador: Universidad Central del Ecuador.

González, J. A., García, S. J., Chalita, T. L., Matus, G. J., Cruz, G. B., Sangerman, J. D., . . . Fortis, H. M. (2012). Modelo de equilibrio espacial para determinar costos de transporte en la distribución de durazno en México. *Revista Mexicana de Ciencias Agrícolas Vol.3 Núm.4* , 701-712 Pp.

Grossman, S. S., & Flores, G. J. (2012). *Álgebra Lineal.* México D. F.: Colonia Santa Fe, Delegación Álvaro Obregón, México D. F.

Guerrero, S. H. (2009). *Programación Lineal Aplicada.* Bogotá, Colombia: Ecoe Ediciones.

Gujarati, D. N., & Porter, D. C. (2010). *Econometría.* Delegación Álvaro Obregón, México, D. F.: McGraw-Hill/Interamericana Editorres, S.A. de C.V.

Hernández, N., Soto, F., & Caballero, A. (2009). Modelos de simulación de cultivos, características y usos. *Cultivos Tropicales, vol. 30, núm. 1*, 73-82 Pp.

Hernandez, S. R., Fernandez, C. C., & Baptista, L. P. (2006). *Metodologia de la investigaci6n.* Colonia Desarrollo Santa Fe, Delegación Álvaro Obregón, México, D. F. : McGraw Hill/Interamericana Editores, S.A. de C. V.

Herrera, A. L. (2009). *Perspecctivas para la carna de bovino en México aplicando un sistema de trazabilidad.* Campus Montecillo, Texcoco, estado de México: Colegio de Postgraduados. Tesis de maestría.

Hillier, F. S., & Lieberman, G. J. (2010). *Introducción a la investigación de operaciones.* Delegación Álvaro Obregón, México D. F.: McGraw-Hill Companies, Inc.

Hillier, F. S., & Lieberman, G. J. (2015). *Operations Reseach.* New York, U.S.A.: McGraw-Hill Education.

Hung, M. T., & Hung, C. C. (2019). A review of non-predetermined goal programming: Methods, applications, and challenges. *European Journal of Operational Research, 274(3)*, 909-924.

Hurtado, M. J., & Gómez, F. R. (2010). *Diseño Experimental.* Valencia, España: Universidad de Valencia.

Iusem, A. N. (2003). On the convergence properties of the projectedgradient method for convex optimization. *Computational and Applied Mathematics*, 37–52.

Kaiser, H. M., & Messer, K. D. (2011). *Mathematical Programming for Agricultural, Enviromental, and Resource Economics.* Rosewood Drive, Danvers, United States of America: John Wiley & Sons, Inc.

Kmenta, J. (1971). *Elements of econometrics.* Collier Macmillan Canada, Inc.: Macmillan Publishing Ccompany.

Larrañaga, L. J., Zulueta, G. E., Elizagarate, U. F., & Alzola, B. J. (2020). Algoritmos Meméticos en problemas de Investigación Operativa. *Escuela Universitaria de Ingeniería de Vitoria-Gasteiz. Universidad del País Vasco Euskal Herriko Unibertsitatea (UPV-EHU).*, 1-19.

Leek, J. T., & Peng, R. D. (2015). *Statistics: P values are just the tip of the iceberg.* Maryland, USA.: School of Public Health in Baltimore.

Li, H., Zhang, J., & Zhang, X. (2019). Randomized rounding for general integer programs. *Operations Research Letters, 47(2)*, 164-169.

Li, W., & Yang, L. (2022). A novel non-predetermined goal programming model for renewable energy planning. *Energy, 220*, 122772.

Lind, D. A., Marchal, W. G., & Mason, R. D. (2004). *Estadística para Administración y Economía.* Col. del Valle, México D. F.: Alfaomega Grupo Editor S. A. de C. V.

Lind, D. A., Marchal, W. G., & Mason, R. D. (2006). *Estadística para Administración y Economía.* Col. del Valle, México D. F.: Alfa Omega Grupo Editor S. A. de C. V.

López, B. E., & González, R. B. (2014). *Diseño y Análisis de Experimentos. Fundamentos y Aplicaciones en Agronomía.* Ciudad Universitaria zona 12. Ciudad de Guatemala: Universidad de San Carlos de Guatemala, Facultad de Agronomía.

Loubet, B. (2020). *Excel: Herramienta Solver.* Centro Universitario, Ciudad de Mendoza. Provincia de Mendoza, Argentina.: Facultad de ciencias económicas. Universidad Nacional de Cuyo.

lpízar, M. J. (12 de Octubre de 2023). *Universidad Latina*. Obtenido de La organización industrial: https://cisprocr.com/cispro/system/files/Clase%209%20Organizaci%C3%B3n%20Industrial.pdf

Luenberger, D. G. (1984). *Linear and Nonlinear Programming* . New JErsey, U.S.A.: Springer.

Machado, D. E., & Coto, F. H. (2017). Sistema de adquisición de datos con Pyhton y Arduino. *Revista, Ciencia, Ingeniería y Desarrollo Tec Lerdo*, 1-6.

Madrid, C. C. (2020). Filosofía de la ciencia del cambio climático: Modelos, problemas e incertidumbres. *Revista Colombiana de Filosofía de la Ciencia, vol. 20, núm. 41*, 201-234 Pp.

Martínez, Á. (26/10/2023 de Octubre26 de 2023). *Optimización global.* Obtenido de Optimización global: http://www.dma.uvigo.es/~aurea/Optimizaci%C3%9Bn_global.pdf?fbclid=IwAR1z2kTahmpEcDLkOYNhhObwwxDVviGT62FDMF6uGU1byE4MU39bTK_6L68

Martínez, A. L. (2013). *Modelo de Programación Cuadrática y Ratios Financieros para minimizar el riesgo de las inversiones en la Bolsa de Valores de Lima.* Lima, Perú: Facultad de Ciencias Matemáticas, E.A.P. de Investigación Operativa. Universidad Nacional Mayor de San Marcos.

Mehrotra, S. (2022). Non-predetermined goal programming: A review of recent advances. *Journal of the Operational Research Society, 73(5)*, 1018-1032.

Mendiburu, F. (2007). *Análisis Estadístico con "R".* Lima, Perú: Sección de Extensión Universitaria y Proyección Social. Centro de Estudios de la Facultad de Ingeniería Económica y Ciencias Sociales. Universidad Nacional de Ingeniería.

Mendiburu, F. (2015). *Agricolae tutorial (Version 1.2-3).* La Molina, Perú: Departamento Academico de Estadística e Informática de la Facultad de Economíaa y Planificación. Universidad Agraria La Molina.

Mijangos, L. J. (2019). *Investigación de Operaciones II.* Tuxtla Guitiérrez, Chiapas, México: Intituto Tecnológico de Tuxtla Guitiérrez.

Moghaddam, S. A., Arabshahi, A., Yazdani, D., & Dehshibi, M. M. (2012). A novel Hybrid Algorithm for Optimization in Multimodal Dynamic Enviroments. *ResearchGate*, 143-148.

Montgomery, D. C. (2004). *Diseño y análisis de experimentos.* México, D. F.: Limusa S. A. de C. V.

Montgomery, D. C. (2012). *Design and analysis of experiments.* Arizona, USA: John Wiley & Sons, Inc.

Muñoz, C. R., Ochoa, H. M., & Morales, G. M. (2011). *Investigación de operaciones.* Delegación Álvaro Obregón, México D. F.: McGraw-Hill/Interamericana Editores, S.A. DE C.V.

Murillo, F. E. (2016). Riesgo agropecuario. *Revista de la Carrera de Ingeniería Agronómica – UMSA*, 103-127 Pp.

Navidi, W. (2006). *Estadística para ingenieros y científicos.* Colonia Desarrollo Santa Fe, Delegación Álvaro Obregón, México,D. F.: McGrawHill McGraw-Hill/Interamericana Editores, S. A. de C. V.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization.* New Jersey, U.S.A.: Springer Science & Business Media.

Nolasco, V. J. (2018). *Python. Apliacaciones prácticas.* Calle Jarama, 3A, Poligono Industrial Igarsa. Paracuellos de Jarama, Madrid, España: Ra-Ma Editorial.

Oliveira, W., Rocha, R. S., & Katz, N. (2010). Esquistossomose mansônica: situação atual e perspectivas de controle. *Revista da Sociedade Brasileira de Medicina Tropical, 43(2)*, 123-130.

O'Neill, P., & Thomas, J. (2010). Mathematical modelling of infectious diseases. *Nature Reviews Microbiology, 8(11)*, 803-814.

Padrón, C. E. (1996). *Diseños Experimentales: con aplicaciones a la agricultura y la ganadería.* Ciudad de México, México: Trillas. Mex.

Parra, R. F. (2016). *Curso de Estadística con R.* . Santander, Cantabría: Instituto Cántabro de Estadística.

Pérez, A. E. (2017). *Relajación Lagrangiana, métodos heurísticos y metaheurísticos en algunos modelos de Localización e Inventarios.* Valladolid, España: Máster en investigación en Matemáticas, Universidad de Valladolid.

Prawda, W. J. (2004). *Métodos y modelos de investigación de operaciones Vol. I: Modelos deterministicos.* Ciudad de México, México: Limusa. Noriega Editores.

Quintana, S. F. (octubre 2016/marzo 2017). Dinámica, escalas y dimensiones del cambio climático. *Tla-Melaua, revista de Ciencias Sociales. Facultad de Derecho y Ciencias Sociales, año 10, núm. 41*, 180-200 Pp.

Ravindran, A., & Garg, D. P. (2018). *Meta-heuristics for optimization.* New Jersey, U.S.A.: Springer.

Rodríguez, C. E. (11 de Octubre de 2023). *La competencia imperfecta.* Obtenido de Biblioteca digital de la Universidad Católica Argentina: http://bibliotecadigital.uca.edu.ar/repositorio/contribuciones/competencia-imperfecta-carlos-rodriguez.pdf

Rodríguez, V. A. (2007). Cambio climático, agua y agricultura. *Desarrollo Rural Sostenible, Nº 1, II Etapa*, 13-23 Pp.

Rojas, C. L., & Rojas, C. L. (2000). *Exploración al diseño experimental.* Granada, España: Ciencia e Ingeniería Neogranadina, 9, 51–59. Universidad Militar Nueva Granada.

Romero, C., & Ventura, S. (2013). *Metaheuristics: Applications to optimization and machine learning.* NewJersey, U.S.A.: Springer.

Rosales, A. J. (2001). *Análisis y diseño en el espacio de estado utilizando Matlab.* quito, Ecuador: Escuela de Ingeniería, Escuela Politécnica Nacional .

Rosell, E. K. (2022). *Optimización con programación dinámica.* Barccelons, España: Facultad de Matemáticas e Informática, Universidad de Barcelona.

Rossum, V. G. (2009). *Tutorial Python versión 2.5.2.* Ciudad de México, México: Mundi Prensa.

Salas, F. V. (2009). Modelos de negocio y nueva economía industrial. *Universia Business Review | Tercer Trimestre*, 1-24 Pp.

Sampayo, M. S. (2019). *Apuntes de economía del medio ambiente.* Iztacala, D. F., México: Universidad Nacional Autónoma de México (UNAM).

Sarasa, C. A. (2017). *Gestión de la información web usando Python.* Rambla del Poblenou, 156. Barcelona, España.: Editorial UOC (Oberta UOC Publishing, SL).

SEMARNAT, & CONAFOR. (2019). *Estrategia Nacional de Sanidad Forestal 2019-2024.* Ciudad de México, México.: Secretaría de Medio Ambiente y Recursos Naturales; Comisión Nacional Forestal.

Szigeti, F., Cardillo, J., Hennet, J. C., & Calvet, J. L. (2014). El Método de Relajación Aplicado a Optimización de Sistemas Discretos. *ResearchGate*, 1-9.

Taha, H. A. (2004). *Investigación de operaciones 7ma Edición.* Atlacomulco No. 500-5° piso. Col. Industrial Atoto. Naucalpan de Juárez, Edo. de México: Pearson Educación de México, S.A. de C.V.

Taha, H. A. (2012). *Investigación de Operaciones.* Fayetteville, Arkansas, USA: PEARSON EDUCACIÓN DE MÉXICO, S.A. de C.V.

Taha, H. A. (2012). *Investigación de Operaciones.* Fayetteville, Arkansas, USA: PEARSON EDUCACIÓN DE MÉXICO, S.A. de C.V.

Toledo, T. R. (2009). *El riesgo en la agricultura.* Quilamapu, Chillán, Chile: TRAMA Impresores S. A.

Triola, M. F. (2004). *Estadística.* Col. Industrial Atoto, Naucalpan de Juárez, Edo. de México: Pearson Educación de México, S. A. de C. V.

Trívez, B. F. (2004). Economía Espacial: una disciplina en auge. *Estudios de Economía Aplicada, Vol. 22-3*, 409-429 Pp.

Urquía, M. A., & Martín, V. C. (2016). *Métodos de simulación y modelado.* Madrid, España: Universidad Nacional de Educación a Distancia.

Valencia, N. E. (2018). *Investigación Operativa. Programación lineal, problemas resueltos con soluciones detalladas.* San Juan Bautista de Ambato, Provincia Tungurahua, Ecuador: Universidad Técnica de Ambato (UTA).

Vanegas, C. J. (2018). *Metodología para la planeación de requerimientos de materiales-MRP.* Bogotá, Colombia: Facultad de Ingeniería.

Especialización en Gerencia Integral de Proyectos. Universidad Militar Bueva Granada.

Vásquez, A. V. (2014). *Diseños Experimentales con SAS.* Cajamarca, Perú: CONCYTEC FONDECYT.

Vega, G. C. (05 de Mayo de 2021). *IESIP*. Obtenido de iesip.edu.ve: https://virtual.iesip.net/mod/page/view.php?id=5933

Velásquez, S., & Velásquez, R. (2012). Modelado con variables aleatorias en simulink utilizando simulación Montecarlo. *Universidad, Ciencia y Tecnología, Vol. 16, No. 64*, 203-211 Pp.

Vercher, G. E. (2015). Soft Computing approaches to portfolio selection. *Boletín de Estadística e Investigación Operativa. Vol. 31. No. 1*, 23-46.

Vicéns, O. J., Herrarte, S. A., & Medina, M. E. (2005). *Análisis de la varianza (ANOVA).* Distrito Federal, México: Trillas, S. A.

Villota, C. E. (2010). *Control moderno y óptimo (MT 227C).* Lima, Perú: Departamento Académico de Ingeniería Aplicada, Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería.

Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2010). *Estadística matemática con aplicaciones.* Col. Cruz Manca, Santa Fé. México, D.F.: Cengage Learning Editores, S.A.

Wasserstein, R. L., & Lazar, N. A. (2016). *Advance online publication The American Statistician* .

World, H. O. (2016). *Global technical strategy for malaria 2016-2030.* Geneva, Switzerland: Geneva, Switzerland: World Health Organization.

Zill, D. G., & Wright, W. S. (2011). *Cálculo. Trascendentes tempranas.* Delegación Álvaro Obregón, México, D. F.: McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V.

EDITORIAL ANDES COGNITIO

Moisés Arreguín Sámano
Profesor e investigador de la Universidad Estatal de Bolívar (UEB), Ecuador.

Ángel Leyva- Ovalle
Docente Investigador Universidad Autónoma Chapingo (UACh-DiCiFo), México. Ingeniero forestal con orientación en economía y ordenación por la Universidad Autónoma Chapingo (UACh), con experiencia en la División de Ciencias Forestales (DiCiFo), específicamente en el Departamento de Productos Forestales.

Johanna Dueñas Durán
Docente Investigador de la Universidad Estatal de Bolívar, ingeniera geóloga graduada en la Universidad Central del Ecuador, máster en gestión y prevención de Riesgos por el IAEN.

Daniel Santiago Paredes Gaibor
Ingeniero Ambiental, Titulado por la Universidad UTE (Ecuador, 2019). Posee dos maestrías: un Máster en Prevención de Riesgos Laborales con énfasis en Seguridad en el Trabajo por la Universidad Politécnica de Valencia (España, 2022) y un Máster en Gestión de Riesgos con mención en Manejo de la Respuesta a Desastres por la Universidad Internacional SEK (Ecuador, 2024).
A lo largo de su trayectoria profesional, ha desempeñado roles clave en empresas como Empresa Pública Metropolitana de Movilidad y Obras Públicas, Gourmet Food Service GFS S.A. y Trade Alliance Corporation, con un enfoque en la seguridad ocupacional y la gestión ambiental. Actualmente, es docente a tiempo completo en la Universidad Estatal de Bolívar, donde imparte clases en la carrera de Ingeniería en Riesgos de Desastres, contribuyendo al desarrollo académico y profesional en este campo.

ISBN: 978-9942-7408-1-6