

FUNDAMENTOS MATEMÁTICOS DE REGRESIÓN LINEAL

Parte I

Moisés Arreguín Sámano

Andrea Damaris Hernández Allauca

Miguel Ángel Gualpa Calva

Eduardo Patricio Salazar Castañeda



FUNDAMENTOS MATEMÁTICOS DE REGRESIÓN LINEAL

Parte I

© Autores

Moisés Arreguín Sámano

Docente de la Universidad Estatal de Bolívar, Bolívar, Ecuador.

Andrea Damaris Hernández- Allauca

Docente Investigador Escuela Superior Politécnica de Chimborazo- Grupo de Investigación y Transferencia de Tecnologías en Recursos Hídricos (GITRH) y Grupo de Investigación en Turismo (GITUR). Riobamba, Ecuador.

Miguel Ángel Guallpa- Calva

Docente Investigador Escuela Superior Politécnica de Chimborazo- Grupo de Investigación Conservación y Producción Sustentable de los Recursos Naturales y Entomológicos (GICOPROSURENE).

Eduardo Patricio Salazar- Castañeda

Docente Investigador Escuela Superior Politécnica de Chimborazo- Grupo de Investigación Forestal (GIFOR).



Casa Editora del Polo – CASEDELPO CIA. LTDA.
Departamento de Edición

Editado y distribuido por:

Editorial: Casa Editora del Polo
Sello Editorial: 978-9942-816
Manta, Manabí, Ecuador. 2019
Teléfono: (05) 6051775 / 0991871420
Web: www.casedelpo.com
ISBN: 978-9942-816-90-0

© Primera edición
© Mayo - 2022
Impreso en Ecuador

Revisión, Ortografía y Redacción:
Lic. Jessica Mero Vélez

Diseño de Portada:
Michael Josué Suárez-Espinar

Diagramación:
Ing. Edwin Alejandro Delgado-Veliz

Director Editorial:
Dra. Tibusay Milene Lamus-García

Todos los libros publicados por la Casa Editora del Polo, son sometidos previamente a un proceso de evaluación realizado por árbitros calificados.

Este es un libro digital y físico, destinado únicamente al uso personal y colectivo en trabajos académicos de investigación, docencia y difusión del Conocimiento, donde se debe brindar crédito de manera adecuada a los autores.

© **Reservados todos los derechos.** Queda estrictamente prohibida, sin la autorización expresa de los autores, bajo las sanciones establecidas en las leyes, la reproducción parcial o total de este contenido, por cualquier medio o procedimiento. parcial o total de este contenido, por cualquier medio o procedimiento.

Comité Científico Académico

Dr. Lucio Noriero-Escalante
Universidad Autónoma de Chapingo, México

Dra. Yorkanda Masó-Dominico
Instituto Tecnológico de la Construcción, México

Dr. Juan Pedro Machado-Castillo
Universidad de Granma, Bayamo. M.N. Cuba

Dra. Fanny Miriam Sanabria-Boudri
Universidad Nacional Enrique Guzmán y Valle, Perú

Dra. Jennifer Quintero-Medina
Universidad Privada Dr. Rafael Belloso Chacín, Venezuela

Dr. Félix Colina-Ysea
Universidad SISE. Lima, Perú

Dr. Reinaldo Velasco
Universidad Bolivariana de Venezuela, Venezuela

Dra. Lenys Piña-Ferrer
Universidad Rafael Belloso Chacín, Maracaibo, Venezuela

Dr. José Javier Nuvaez-Castillo
Universidad Cooperativa de Colombia, Santa Marta,
Colombia

Constancia de Arbitraje

La Casa Editora del Polo, hace constar que este libro proviene de una investigación realizada por los autores, siendo sometido a un arbitraje bajo el sistema de doble ciego (peer review), 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, Sexta Edición, proceso de anti plagio en línea Plagiarisma, garantizándose así la cientificidad de la obra.

Comité Editorial

Abg. Néstor D. Suárez-Montes
Casa Editora del Polo (CASEDELPO)

Dra. Juana Cecilia-Ojeda
Universidad del Zulia, Maracaibo, Venezuela

Dra. Maritza Berenguer-Gouarnaluses
Universidad Santiago de Cuba, Santiago de Cuba, Cuba

Dr. Víctor Reinaldo Jama-Zambrano
Universidad Laica Eloy Alfaro de Manabí, Ext. Chone

ÍNDICE GENERAL

ÍNDICE GENERAL.....	11
PRÓLOGO.....	13
INTRODUCCIÓN.....	15
RESUMEN.....	17
CAPÍTULO I.....	19
REGRESIÓN LINEAL SIMPLE.....	19
CAPÍTULO I.....	21
REGRESIÓN LINEAL SIMPLE.....	21
1.1. Modelo.....	22
1.2. Hipótesis o supuestos.....	25
1.2.1. Nomenclatura matricial.....	27
1.2.2. Matriz covarianza entre b0 y b1	37
1.2.3. Análisis de varianza (ANOVA).....	37
1.3. Modelos deterministas y probabilísticos.....	38
1.3.1. Modelo lineal simple.....	39
1.3.2. Estimadores de parámetros.....	41
1.3.3. Método de mínimos cuadrados (MCO).....	45
1.3.4. Propiedades de estimadores MC	54
1.3.5. Propiedades de estimadores de mínimos cuadrados ordinarios según supuestos de normalidad.....	55
.....	59
.....	59
CAPÍTULO II.....	59
CAPÍTULO II.....	61
MÁXIMA VEROSIMILITUD (MV).....	61
2.1. Estimación de máxima verosimilitud del modelo de regresión con dos variables.....	61
2.2. Método de Máxima Verosimilitud.....	62
2.3. Estimación de MCO y MV de Coeficientes de Regresión Parcial.....	65
2.4. Teorema de Gauss-Markov.....	70
2.4.1. Precisión o errores estándar de estimaciones de Mínimos Cuadrados Ordinarios.....	80
2.4.2. Grados de libertad.....	81

2.4.3.	Teorema de Gauss–Markov bajo las hipótesis H	83
2.4.4.	Estimación σ^2	83
2.5.	Desigualdad de Cauchy–Schwarz en \mathbb{R}^n	98
2.6.	Regresión sin término constante.....	106
2.6.1.	Distribución de estimadores.....	107
2.6.2.	Normalidad asintótica	109
2.6.3.	Intervalos de confianza	111
2.7.	Ejemplos:	120
2.8.	Pruebas de hipótesis.....	122
2.9.	Coefficiente de determinación.....	136
2.9.1.	Coefficiente R^2	136
2.9.2.	Coefficiente de correlación.....	142
2.10.	Varianza de regresión.....	149
2.10.1.	Inferencias de pendiente recta β_1	150
2.10.2.	Inferencias de pendiente recta β_0	151
2.10.3.	Adecuación de modelo.....	152
2.10.4.	Prueba de linealidad	153
2.10.5.	Gráficos de residuos	156
2.10.6.	Detección de normalidad.....	158
2.11.	Otros indicadores de normalidad	161
2.11.1.	¿Qué hacer ante falta de normalidad?.....	162
2.11.2.	Puntos singulares y atípicos	162
2.11.3.	Puntos palanca.....	163
2.11.4.	Puntos influyentes.....	163
2.12.	Nomenclatura matricial de ANOVA	166
2.12.1.	Análisis de varianza ANOVA	175
2.12.2.	Transformaciones de variables.....	176
2.12.3.	Usuales para linealizar relación.....	177
2.12.4.	Homocedasticidad	180
	BIBLIOGRAFÍA.....	182

PRÓLOGO

En muchos de los casos de ingeniería, agronomía, ciencias políticas, relaciones internacionales, economía, ciencias sociales, ciencias de la salud, entre otras; el problema es conocer el comportamiento de una variable aleatoria Y , predecir sus valores tal que se le puede conseguir con un estudio univariante; por ejemplo: histograma, media, varianza o mediante la construcción de un modelo probabilístico. El conocimiento, específicamente las predicciones se enriquecen si Y depende de una o más variables X conocidas o que pueden ser controladas. En este caso, es importante encontrar una función que relacione Y con las X . Asimismo, la finalidad de la regresión lineal múltiple es la misma de la regresión lineal simple, con la diferencia que interviene más de un regresor en el modelo tal que en lugar de rectas de regresión se tendrán hiperplanos de regresión. Un plano en un espacio de tres dimensiones para el caso de dos regresores.

La regresión lineal es posiblemente la técnica más usada para establecer relaciones funcionales entre variables a partir de una relación lineal de la forma $Y_{(\text{Vector aleatorio})} = X(X \text{ es una matriz compuesta por valores de las } X)\beta_{(\text{Vector de parámetros desconocidos})} + U_{(\text{Vector aleatorio})}$ mientras que Y es vector aleatorio observable y U es vector aleatorio no observable. La regresión lineal se enmarca en los modelos lineales, que a su vez se dividen en modelos lineales clásicos, que tienen la forma anterior tal que las componentes U son variables aleatorias independientes con parámetros $(0, \sigma^2)$ y modelos lineales generalizados. Sin embargo, el uso y aplicación de cada modelo es diferente; por ejemplo: el análisis de varianza se usa básicamente en diseño experimental.

INTRODUCCIÓN

La estadística es una rama de la matemática que usa un gran conjunto de datos provenientes de una muestra poblacional, recursos naturales e industriales con el fin de hacer inferencias basadas en cálculo de probabilidades; es decir, es el arte y la ciencia que reúne datos, analizar, presenta e interpreta mediante el uso de modelos matemáticos, entendidos como la abstracción simplificada de una realidad más compleja en que siempre existirá una cierta discrepancia entre lo observado y lo previsto por estos (Capa, 2015). Tiene uso en múltiples disciplinas científicas, como ingeniería forestal, agronomía, agro negocios, economía, procesamiento de alimentos, ciencias de la salud, marketing, elecciones políticas, estudios de mercado, control de calidad, inventarios, padrón de cobro de impuestos, muestreo, entre otras.

Para un investigador que recién inicia, el procesamiento de recuento, enumeración, conteo o cuantificación de grandes volúmenes productivos, cosechas agrícolas, listas de productos, censos poblacionales u otros grandes conjuntos de datos están íntimamente relacionados, mediante técnicas estadísticas, con actividades soportadas de organización de bases, tabulación de datos, presentación gráfica, estimación de cantidades “representativas”, procesos inductivos, procesos deductivos, uso indiscriminado de software, posible pago de licencias, manejo complejo, material bibliográfico o tutoriales abstrusos, procesamiento automático de datos, criterios interpretativos incomprensibles y transcripción mecánica de resultados (Levine, Krehbiel, & Berenson, 2006). No obstante, el problema es conocer el comportamiento de una variable aleatoria Y , predecir sus valores tal que se le puede conseguir con un estudio univariante; por ejemplo: histograma, media, varianza o mediante la construcción de un modelo probabilístico. El conocimiento, específicamente las predicciones se enriquecen si Y depende de una o más variables X conocidas o que pueden ser controlas.

Con base en esto, el libro “Regresión Lineal” presenta y desarrolla ejercicios teóricos-prácticos, interpretación con criterios informales o nocionales y formales respecto a establecer relaciones funcionales entre variables a partir de una relación lineal de la forma $Y_{(\text{Vector aleatorio})} = X_{(X \text{ es una matriz compuesta por valores de las } X)}\beta_{(\text{Vector de parámetros desconocidos})} + U_{(\text{Vector aleatorio})}$. Sus capítulos abordan conocimiento progresivo mediante regresión lineal simple (presenta hipótesis o supuestos, nomenclatura matricial, modelos deterministas y probabilísticos, modelo lineal simple, método de mínimos cuadrados, propiedades de mínimos cuadrados, propiedades de estimadores de mínimos cuadrados ordinarios según supuestos de normalidad), máxima verosimilitud (teorema de Gauss–Markov, precisión o errores estándar de estimadores de mínimos cuadrados ordinarios, grados de libertad, hipótesis de normalidad, coeficiente de determinación, adecuación de modelo, nomenclatura matricial de ANOVA, predicción, ejemplos, ejercicios), regresión lineal múltiple (modelo, estimación puntual, inferencia paramétrica, coeficientes de regresión lineal múltiple, coeficientes de determinación, regresión polinomial, regresión con variables cualitativas, ejemplos, ejercicios), introducción a modelos econométricos (Cobb–Douglas, función de consumo, estructuras de consumo, curva de Phillips) e introducción a series de tiempo (introducción, componentes, atenuación, comparación de métodos, ejercicios).

Finalmente, este libro pretende contribuir a superar deficiencias de estudiantes, investigadores y personas involucradas en manejo de bases de datos a través de la sensibilidad y experiencia de profesionales del área, pues su objetivo es aportar conocimiento en manejo de datos estadísticos de manera confiable, práctica y de uso común en diferentes disciplinas mediante el uso de ofimática actual, diversa, práctica, con licencia pagada o versión libre (Microsoft Excel, R Studio – y Python – Jupyter), con diferentes niveles de rigurosidad científica y mediante ejercicios prácticos que incluyan herramientas informáticas de uso simple hasta programación avanzada para que el usuario tengan certeza sobre salidas o resultados antes de aceptarlos inmediatamente.

RESUMEN

El análisis de regresión lineal es una técnica estadística usada para estudiar la relación entre variables. Se adapta a una amplia variedad de situaciones y se enmarca en modelos lineales. Su objetivo es hallar una relación que describa o resuma la relación entre dos o más variables; por ejemplo: ¿cómo varía el promedio anual del maíz, según producción a nivel nacional? o ¿cómo varía el consumo de gasolina de un auto, según peso y potencia de su motor? En contexto de la investigación de mercados puede utilizarse para determinar en cuál de diferentes medios de comunicación puede resultar más eficaz invertir o predecir el número de ventas de un determinado producto. En caso de dos, regresión simple, o más variables, regresión múltiple, el análisis de regresión lineal puede usarse para explorar y cuantificar la relación entre una variable llamada dependiente, endógena y una o más variables independientes, exógenas, así como desarrollar una ecuación lineal predictiva. Asocia una serie de procedimientos de diagnóstico, análisis de residuos y puntos de influencia respecto a estabilidad e idoneidad de análisis. Probablemente, es la técnica más usada para establecer relaciones funcionales entre variables a partir de una relación lineal de forma $Y = X\beta + U$.

Actualmente, es una herramienta metodológica cuantitativa y cualitativa que permite la asignación óptima de recursos escasos, apoyar eficientemente el proceso de toma de decisiones respecto a maximizar ganancias, utilidades, satisfacción de clientes o minimizar costos, distancias y tiempos. Hace uso de modelos matemáticos, que procesan datos, representan el problema real, establecen hipótesis que es una representación precisa de la realidad tal que sus conclusiones sean válidas para resolver un problema práctico real; por ejemplo: cuánto producir con base en variables estocásticas-no estocásticas, dónde producir, a qué precio ofertar, a qué precio comprar materia prima, medios y rutas optimas de transporte.

Con base en esto, Regresión Lineal con Ofimática presenta terminología básica de regresión lineal, demostraciones matemáticas superiores, modelado matemático, soluciones factibles, cálculos iterativos de maximización, minimización con método algebraico con método de ecuaciones simultáneas, programación lineal, algoritmos matemáticos, entre otras. Con información del Censo Nacional Agropecuario (CNA, 2000), Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC, 2002 a 2020) del Instituto Nacional de Estadística y Censos (INEC) del Ecuador a nivel Nacional, Regional, Provincial y Cantonal de cultivos permanentes (banano, cacao, café, caña de azúcar para biocombustible, caña de azúcar tallo fresco, naranja, palma africana, plátano y tomate de árbol), transitorios (arroz, cebada, fréjol, maíz duro seco, maíz suave choclo, papa, trigo y yuca), producción animal (asnal, caballar, caprino, mular, pollos-gallinas en campo, pollos-gallinas en plantales avícolas, porcino y ovino), pastos (cultivados y naturales) y montes-bosques mediante uso de software con licencia (Microsoft Excel), open office (Python-Jupyter, R y R-Studio) e interactúa amplia y profundamente con Estadística para ingeniería, Diseño experimental para ingeniería e Investigación operativa para ingeniería, que paralelamente usan software con licencia y open office mencionados para desarrollo de ejercicios teóricos-prácticos complementarios.

CAPÍTULO I

REGRESIÓN LINEAL SIMPLE

CAPÍTULO I

REGRESIÓN LINEAL SIMPLE

Muchas de las aplicaciones estadísticas requieren la estimación de las relaciones existentes entre dos o más variables; por ejemplo, puede ser necesario responder a las preguntas ¿cómo varía el promedio anual del maíz, según la producción a nivel nacional? o ¿cómo varía el consumo de gasolina de un auto, según su peso y la potencia del motor? El interés se centra, entonces, en determinar una ecuación que relacione una variable dada con una o más variables que contienen información sobre la primera.

A estos problemas dedicaremos los dos siguientes capítulos; antes revisemos algo de historia de esta parte de la estadística. No se conoce, con exactitud, quién y cuándo empezó a tratar de expresar algebraicamente las relaciones entre dos o más variables, de las cuales solo se dispone de un conjunto de observaciones; pero en los escritos de Leonardo da Vinci, cuando él trata de las proporciones del cuerpo humano, se encuentran expresiones aritméticas que relacionan las medidas de diversas partes del cuerpo.

Un intento, que está bien documentado, data de 1755, cuando Boscovich y Christopher Maire estaban; encargados de medir la longitud del arco de meridiano que pasa por Roma. Boscovich concibió un método para encontrar un modelo que relacione los datos correspondientes a dos variables, mediante el empleo de las llamadas «regresiones elementales».

Esta técnica fue mejorada por su autor en 1760 y puesta en una forma más estructurada por Laplace, unos años más tarde. Incluso, en 1805 Legendre publicó una obra de astronomía, en la que describió el método de los mínimos cuadrados y lo aplicó al ajuste de datos observacionales. También, hay una serie de artículos presentados por Johann Carl Friedrich Gauss a la Sociedad Real de Gotinga en los que describe el método de mínimos cuadrados. Luego, en 1885, Sir Francis Galton

presentó en la revista Nature el desarrollo completo de esta técnica, aplicada a lo que él denominó modelos de regresión.

A partir de esta fecha se mejoró y se completó la técnica, haciendo que ella sea la de mayor empleo en el ajuste de conjuntos de datos. Actualmente, la construcción de modelos lineales es la base de todas las técnicas estadísticas de naturaleza predictiva y su aplicación se ha extendido a prácticamente todas las ciencias.

1.1. Modelo

Los modelos de regresión lineal simple tienen por objeto hallar una relación funcional entre dos variables, representadas por X e Y . Se supone que Y es función de X , en un sentido físico se dice que X causa Y . Entonces, Y es variable explicada, respuesta, variable dependiente o endógena mientras que X es variable explicativa, independiente o exógena. A variable X se llama regresor o predictor. El análisis de regresión puede tener por objetivos:

A. Descripción.

El objetivo indicado anteriormente es hallar una relación que describe o resume la relación entre dos variables, pues es el objetivo básico del análisis de regresión.

B. Estimación de coeficientes.

Se puede conocer la relación teórica entre variables, pero esta relación depende de uno o más parámetros desconocidos y dependen de datos, en este caso es importante su estimación. El objetivo del investigador es estudiar si los datos permiten confirmar la teoría.

C. Predicción.

La finalidad de construir un modelo y estimar sus parámetros puede ser la predicción; es decir, dado un valor de variable X es conocer el valor que podría tener variable Y . Estas predicciones son muy importantes en Economía como en

Agroindustria, pero hay que tener cuidado con extrapolaciones mientras que la hipótesis básica de predicciones es que el modelo no cambia fuera del rango observado.

D. Control.

Modelos de regresión pueden servir para controlar y monitorear un sistema.

E. Selección de variables.

Se puede disponer de un “conjunto grande” de variables independientes, pero se requiere reducir a un “conjunto pequeño” que explique adecuadamente la variable dependiente tal que su objetivo corresponde, en realidad, a regresión lineal múltiple.

Con base en lo anterior, los objetivos anteriores dependen de la relación de una variable Y en función de otra variable X. La regresión lineal y sus modelos lineales suponen que la relación es lineal afín:

$$Y = \beta_1 + \beta_2 X$$

Donde:

Y: Variable endógena

β_1 : Coeficiente del intercepto

β_2 : Coeficiente de pendiente

X: Variable exógena

Aunque esta relación está sujeta a errores experimentales, medición, muestras y, básicamente, aislamiento de variables X e Y.

Por ejemplo: el consumo depende del ingreso del consumidor; aunque, no sólo depende del ingreso, sino de otras variables exógenas, como hábitos de consumo, inflación, precios, gustos, preferencias, nivel de publicidad, ahorro, promociones en productos, productos sustitutos, productos complementarios, etcétera.

Entonces, la relación entre consumo e ingreso no es exacta; por lo que se agrega un “término error”. P

or lo tanto, los modelos lineales suponen que el error es aditivo y depende de la pareja (y_i, x_i) tal que $y_i = \beta_1 + \beta_2 X_i + u_i$ $i = 1, 2, 3, 4, \dots, n$ tal que u_i es error. Esto, supone que se han realizado n observaciones de pareja (y_i, x_i) , en diseño de experimentos se fijan valores x_i y se miden respuestas y_i .

Los parámetros β_1 y β_2 son desconocidos y deben ser estimados con base en n observaciones. El término error representa:

A. Su presencia se debe a no inclusión de otras variables en modelo.

No se incluyen pues su efecto sobre variable respuesta Y es muy pequeño y no sistemático o, simplemente, no se dispone de información suficiente. Pueden existir variables que son muy difíciles de medir y, prácticamente, son imponderables, como gustos y preferencias que afectan al consumo y le confieren una aleatoriedad propia.

B. El modelo podría ser no lineal en X .

Por ejemplo: cuadrático ($y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$), cúbico ($y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$), logarítmico ($\ln(X)$), exponencial ($\exp(X)$), etcétera tal que si se usa el término del error $u_i = \beta_3 X_i^2 + e_i$.

C. Errores de medición.

Si, por ejemplo, en una encuesta se pregunta a un conjunto de personas (muestra) no recuerdan exactamente o mienten cuando se le pregunta respecto al nivel de consumo. Asimismo, puede suceder que respondan sobre el consumo más reciente y, en consecuencia, no refleja el promedio de consumos en el año.

En conclusión, el experimentador tendrá en cuenta el error aleatorio o diferencia entre lo que observa y espera observar según el modelo, más aún que

variable Y también es aleatoria, una forma de modelar su distribución es mediante error.

1.2. Hipótesis o supuestos.

y_i son variables aleatorias observables mientras que valores x_i se supone son conocidos y no aleatorios tal que errores o perturbaciones u_i son variables aleatorias no observables. La hipótesis básica que sirve como punto inicial es que variables están ligadas por relación lineal $y_i = \beta_1 + \beta_2 X_i + u_i$ $i = 1, 2, 3, 4, \dots, n$ y coeficientes β_1, β_2 no son aleatorios. Los errores u_i satisfacen las hipótesis:

H₁: Para $\forall i$ $E(u_i) = 0$.

H₂: Para $\forall i$ $Var(u_i) = \sigma^2$ llamada hipótesis de homocedasticidad de errores.

H₃: Para $\forall i, \text{ todo } i' \neq i$ $Cov(u_i, u_{i'}) = 0$. \nexists correlación de errores. Cuando datos son cronológicos se habla de correlación serial o auto correlación en vez de correlación. La aditividad del error, la no aleatoriedad de variable independiente y parámetros, permiten que estas hipótesis se pueden escribir en función de momentos y_i .

H₁' Para $\forall i$ $E(Y_i|x_i) = \beta_1 + \beta_2 x_i$. Para enfatizar que esperanzas dependen de valores x_i se escribe $E(Y_i|x_i)$ en vez de $E(Y_i)$.

H₂' Para $\forall i$ $Var(Y_i|x_i) = \sigma^2$. La varianza es constante y no depende de x_i .

H₃' Para $\forall i, \text{ todo } i' \neq i$ $Cov(Y_i, Y_{i'}|x_i, x_{i'}) = 0$. Las covarianzas son nulas y no dependen de $x_i, x_{i'}$.

Las hipótesis H y H' son equivalentes, se implican mutuamente. Su cumplimiento depende de condiciones de observación o condiciones experimentales, en general, se cumplirán en casos en:

- Estudios transversales, si se usa un muestreo aleatorio adecuado; por ejemplo: muestreo aleatorio simple (MAS).
- Diseños experimentales (DE).

La hipótesis H_1' significa que la media de y_i sólo depende de β_1, β_2 y x_i . No depende de variables no incluidas en modelo. Supone que el efecto de estas sobre la variable respuesta es muy pequeño y nulo en promedio.

La hipótesis H_2' afirma que la varianza no depende de x_i mientras que hipótesis H_3' sostiene que no existe correlación entre observaciones de y_i de forma independiente de valores (x_i, x_i') .

En general, estas hipótesis serán falsas para modelos econométricos; es decir, las variables son de tipo económico y observadas periódicamente o si el modelo está mal especificado.

Las tres hipótesis anteriores permiten calcular estimadores puntuales de parámetros β_1, β_2 y σ^2 pero no permiten realizar inferencias con métodos paramétricos usuales, salvo recurrir a resultados asintóticos. Para construir intervalos de confianza y contrastar hipótesis, de nivel exacto y no asintótico, se asume la normalidad de errores tal que supone:

$$N_1 \text{ Para } \forall i \ u_i \rightarrow N(0, \sigma^2)$$

$$N_2 \text{ Variables aleatorias } \{u_1, u_2, u_3, u_4, \dots, u_n\}$$

Con base en esto, la hipótesis N_1 implica hipótesis H_1 y H_2 mientras que hipótesis N_2 implica H_3 , en realidad estas dos últimas, bajo hipótesis de normalidad, son equivalentes. Por un lado, al primer grupo de hipótesis refiere a como las hipótesis H y, al segundo, que añade la normalidad de errores, llamadas hipótesis N . También, supone no existen valores atípicos.

Por otro lado, cuando variables independientes son aleatorias se recurre a leyes condicionales, así para hipótesis sobre la variable dependiente se tiene:

1. $E(Y_i|X = x_i) = \beta_1 + \beta_2 x_i$. La notación $E(Y_i|X = x_i)$ indica esperanza condicional de Y_i dado que $X = x_i$.
2. $\text{Var}(Y_i|x_i) = \sigma^2$.
3. $(Y_i|X_2 = x_i) \rightarrow N(\beta_1 + \beta_2 x_i, \sigma^2)$.

4. $\{Y_1, Y_2, Y_3, Y_4, \dots, Y_n\}$ son condicionalmente independientes.

Además, se asume que error y regresor o regresores tienen distribuciones independientes.

1.2.1. Nomenclatura matricial

Según (Grossman & Flores, 2012), $V_{(\text{Espacio Vectorial})} = \mathbb{R}^n =$

$$\left\{ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{matrix} : X_j \in \mathbb{R} \text{ para } j = 1, 2, 3, 4 \dots n \right\}$$

tal que cada vector \mathbb{R}^n es una matriz $n \times 1$,

siendo \mathbb{R}^n e símbolo que denota al conjunto de todos los vectores de dimensión n :

$\left\{ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{matrix} \right\}$ donde cada a_i es un número real (\mathbb{R}). Según la definición de matrices, una

matriz A de $m \times n$ es un arreglo rectangular de mn números dispuestos en m

renglones y n columnas tal que $A = \left\{ \begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \dots & a_{2j} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \dots & a_{3j} & \dots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \dots & a_{4j} & \dots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & a_{i4} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} & \dots & a_{mj} & \dots & a_{mn} \end{matrix} \right\}$,

$\mathbf{x} + \mathbf{y}$ es una matriz de $n \times 1$ si \mathbf{x} e \mathbf{y} son matrices $n \times 1$. Haciendo $\mathbf{0} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$ y $-\mathbf{x} =$

$\left\{ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{matrix} \right\}$, 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 $m \times n$ y sean $\alpha - \beta$ dos escalares 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 sumas

(Ley de Composición Interna en V) y multiplicación

(Ley de Composición Externa en V) por un escalar, que satisfacen 8 axiomas:

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

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

Se llama espacio vectorial V sobre \mathbb{R} a la terna $(V +, *) \Rightarrow V = \{\vec{\mu}, \vec{\nu}, \vec{w}, \dots\}$ donde sus elementos son vectores y se cumplen los 8 axiomas anteriores. Entonces, un sub espacio vectorial $(H +, *) \subset (V +, *)$ tal que H es un sub espacio vectorial de V si H

es un subconjunto no vacío 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 V .

Por teorema, un subconjunto no vacío H de un espacio vectorial V es un subespacio de V si se cumplen las dos reglas de cerradura:

I) Si $\vec{u} \in H$ y $\vec{v} \in H \Rightarrow \vec{u} + \vec{v} \in H$ y II) Si $\vec{u} \in H$ y $\alpha\vec{u} \in H \Rightarrow \forall \alpha_{(\text{Escalar})} \in \mathbb{R}$ tal que la caracterización de un subespacio vectorial está dado por que $H \subset V \Leftrightarrow \left. \begin{matrix} \vec{u}, \vec{v} \in H \\ \alpha, \beta \in \mathbb{R} \end{matrix} \right\} \Rightarrow (\alpha\vec{u} + \beta\vec{v}) \in H$; 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 $(0, 0)$.

Además, combinación lineal se demuestra a partir que sean $v_1, v_2, v_3, v_4, \dots, v_n$ vectores en un espacio vectorial V tal que cualquier vector de la forma $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$ se denomina **combinación lineal de $v_1, v_2, v_3, v_4, \dots, v_n$** , 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 $v_1, v_2, v_3, v_4, \dots, v_n$ 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, $\forall v \in V_{(\text{Espacio Vectorial})} \exists$ escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n \Rightarrow v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$.

Además, un espacio generado por un conjunto de vectores se genera a partir que sea $v_1, v_2, v_3, v_4, \dots, v_k$ k vectores de un espacio vectorial V .

El espacio generado por $\{v_1, v_2, v_3, v_4, \dots, v_k\}$ es el conjunto de combinaciones lineales $v_1, v_2, v_3, v_4, \dots, v_k$; es decir, $\text{gen}\{v_1, v_2, v_3, v_4, \dots, v_k\} = \{v | v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_k v_k\}$.

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 $X \in V$ y $Y \in V$, entonces $X + Y \in V$ (**cerradura bajo suma**)
- ii) Para $\forall x, y, z \in V$, $(x + y) + z = x + (y + z)$ (**ley asociativa de suma de vectores**)
- iii) \exists un vector $0 \in V \Rightarrow \forall X \in V, X + 0 = 0 + X = X$ (**0 es llamado vector cero o idéntico aditivo**)
- iv) Si $X \in V$, existe un vector $-X \in V$ tal que $X + (-X) = 0$ (**-x se llama inverso aditivo de X**)
- v) Si X y Y están en V , entonces $X + Y = Y + X$ (**Ley conmutativa de suma de vectores**)
- vi) Si $X \in V$ y α es una escalar, entonces $\alpha X \in V$ (**cerradura bajo multiplicación por un escalar**)
- vii) Si X y Y están en V y α es una escalar, entonces $\alpha(X + Y) = \alpha X + \alpha Y$ (**primera ley distributiva**)
- viii) Si $X \in V$ y $\alpha - \beta$ son escalares, entonces $(\alpha + \beta)X = \alpha X + \beta X$ (**segunda ley distributiva**)
- ix) Si $X \in V$ y $\alpha - \beta$ son escalares, entonces $\alpha(\beta X) = (\alpha\beta)X$ (**ley asociativa de multiplicación por escalares**)
- x) Para \forall vector $X \in V, 1X = X$

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

Demostración:

- i) $X = (X_1, Y_1)$ y $Y = (X_2, Y_2)$ están en $V \Rightarrow Y_1 = mX_1, Y_2 = mX_2$ y $X + Y = (X_1, Y_1) + (X_2, Y_2) = (X_1, mX_1) + (X_2, mX_2) = (X_1 + X_2, mX_1 + mX_2) = (X_1 + X_2, m(X_1 + X_2)) \in V \therefore$ se cumple este axioma.
- ii) $(X, Y) \in V \Rightarrow Y = mX - (X, Y) = -(X, mX) = (-X, m(-X))$ tal que $-(X, Y) \in V$ y $(x, mX) + (-X, m(-X)) = (X - X, m(X - X)) = (0, 0)$.

Por lo tanto, \forall vector en V es vector en \mathbb{R}^2 y \mathbb{R}^2 es V , como se muestra enseguida:

$$V_{(\text{Espacio Vectorial})} = \mathbb{R}^n = \left\{ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{matrix} : X_j \in \mathbb{R} \text{ para } j = 1, 2, 3, 4 \dots n \right\} \text{ tal que cada vector } \mathbb{R}^n$$

es una matriz $n \times 1$, siendo \mathbb{R}^n e símbolo que denota al conjunto de todos los

vectores de dimensión n : $\left\{ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{matrix} \right\}$ 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 $V \Rightarrow (X_1, Y_1, Z_1) + (X_2, Y_2, Z_2) = (X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2) \in V$ 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 $V = \{(X, Y): Y = 2X + 1, X \in V\}$, 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 $X \in V$ y $Y \in V$, entonces $X + Y \notin V$).

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 $Y_2 = 2X_2 + 1$, de manera que $Y_1 + Y_2 = (2X_1 + 1) + (2X_2 + 1) = 2X_1 + 2X_2 + 2$.

Por lo tanto, $(X_1 + X_2, Y_1 + Y_2) \notin V \Leftrightarrow (X_1, Y_1) \in V$ y $(X_2, Y_2) \in V$. 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 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 $p(X) = a_n X^n + a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + a_{n-3} X^{n-3} + a_{n-4} X^{n-4} + \dots + a_1 X + a_0 \forall a_i \in \mathbb{R}$.

La suma de $p(X) + q(X)$ está definida usualmente, si $q(X) = b_n X^n + b_{n-1} X^{n-1} + b_{n-2} X^{n-2} + b_{n-3} X^{n-3} + b_{n-4} X^{n-4} + \dots + b_1 X + b_0 \forall b_j \in \mathbb{R} \Rightarrow \therefore p(X) + q(X) = (a_n + b_n) X^n + (a_{n-1} + b_{n-1}) X^{n-1} + (a_{n-2} + b_{n-2}) X^{n-2} + (a_{n-3} + b_{n-3}) X^{n-3} + (a_{n-4} + b_{n-4}) X^{n-4} + \dots + (a_1 + b_1) X + (a_0 + b_0)$ 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 $\mathbf{0} = 0X^n + 0X^{n-1} + 0X^{n-2} + 0X^{n-3} + 0X^{n-4} + \dots + 0X + 0 \Rightarrow 0 \in P_n$ 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 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 $m \times n$ 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 $m \times n$.

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 3×3 . 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 $m \times n \Rightarrow AB$ 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 puede 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 $m \times n$, $B = (b_{ij})$ matriz de $m \times p$ y $C = (c_{ij})$ matriz de $p \times q$, 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 $n \times q$. 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 usa 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 $0\mathbf{X} + (-0\mathbf{X}) = 0\mathbf{X} + [0\mathbf{X} + (-0\mathbf{X})] = \mathbf{0} = 0\mathbf{X} + \mathbf{0} = 0\mathbf{X}$

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.

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), $(-1)\mathbf{X} \in H$ 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.

Ecuación Paramétrica o Vectorial de Recta. 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:

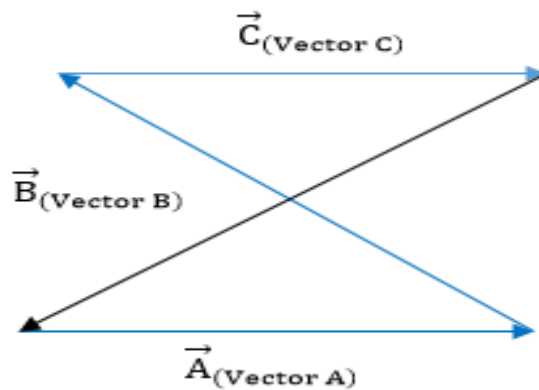


Figura. 1. 1. Ecuación paramétrica o vectorial de recta

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

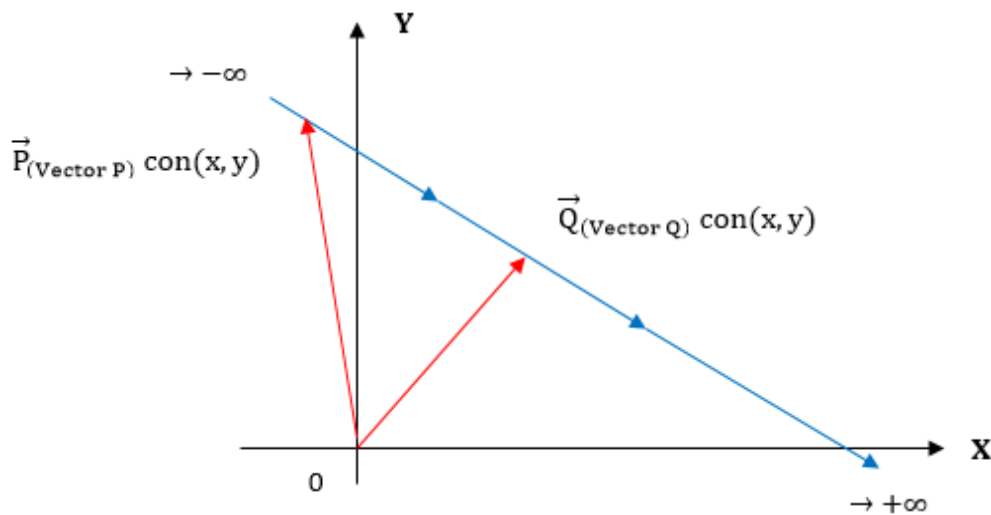


Figura. 1. 2. Suma de vectores por el método del polígono

Sin embargo, esta sumatoria de vectores puede transformarse en la forma de Regresión Lineal Simple ($Y = b + mx$) y su nomenclatura matemática sería

$$\vec{r}_{(\text{Vector Posicion})} = \vec{P}_{(\text{Vector P})} + t_{(n \text{ veces la multiplicación de vector } \overline{PQ})} \overline{PQ}_{(\text{Vector PQ})} \Rightarrow$$

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

1.2.2. Matriz covarianza entre b_0 y b_1

Es importante recordar:

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{Var}(b_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

Además, la covarianza entre b_0 y b_1 es:

$$\text{Cov}(b_0, b_1) = \text{Cov}[(\bar{y} - b_1 \bar{x}), b_1] = \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Entonces, la matriz de varianza-covarianza del vector b es:

$$\text{Var}(b) = \begin{pmatrix} \text{Var}(b_0) & \text{Cov}(b_0, b_1) \\ \text{Cov}(b_0, b_1) & \text{Var}(b_1) \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix}$$

Si se extrae fuera de la matriz el factor σ^2 :

$$\text{Var}(b) = \sigma^2 (X^t X)^{-1}$$

La varianza σ^2 se estima por mediante:

$$s^2 = \frac{Y^t Y - b^t X^t Y}{n - 2}$$

1.2.3. Análisis de varianza (ANOVA)

Con empleo de formulación matricial de componentes de regresión se puede obtener de cuadros:

$$\text{SCE} = Y^t Y - b^t X^t Y,$$

$$SCR = b^t X^t Y - \frac{(\sum_{i=1}^n y_i)^2}{n},$$

$$SC_{yy} = Y^t Y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

Tal que el coeficiente de determinación, en formulación matricial, es:

$$r^2 = \frac{SCR}{SC_{yy}} = \frac{b^t X^t Y - \frac{(\sum_{i=1}^n y_i)^2}{n}}{Y^t Y - \frac{(\sum_{i=1}^n y_i)^2}{n}}$$

Aunque, se puede reconstruir la tabla de análisis de la varianza en su forma matricial:

Tabla 1.1. Reconstrucción de tabla ANOVA en su forma matricial

Tabla de análisis de la varianza			
Fuente de variación	Grados de libertad(g.l.)	Suma de cuadrados (SC) medio (MC)	Cuadrado
Regresión	1	$SCR = b^t X^t Y - \frac{1}{n} (\sum_{i=1}^n y_i)^2$	MCR
Error o Residual	$n - 2$	$SCE = Y^t Y - b^t X^t Y$	$s^2 = \frac{SCE}{n-2}$
Total Corregido	$n - 1$	$SC_{yy} = Y^t Y - \frac{1}{n} (\sum_{i=1}^n y_i)^2$	

1.3. Modelos deterministas y probabilísticos

Existen aplicaciones en las que se dispone de un modelo que presenta una relación exacta entre las variables de interés; por ejemplo, la ley que describe el tiempo que tarda en caer un objeto desde una Altura dada o la fórmula que nos indica el interés ganado por un capital, datos la tasa de interés y el periodo de la inversión. Tales modelos se denominan deterministas.

Sin embargo, en la vida diaria, rara vez se presentan fenómenos que reproducen con exactitud una ley ya sea porque existen errores en la medición o porque hay

otras que no son consideradas, por su escasa influencia, pero que son suficientes para que el modelo propuesto no sea exacto.

En consecuencia, un modelo en el que una o más variables es de naturaleza aleatoria se denomina probabilístico y a la determinación y examen de la calidad del modelo encontrado se llama análisis de regresión

Algunas de las más importantes aplicaciones del análisis de regresión son:

1. Descripción cuantitativa de las relaciones entre una variable dada y un conjunto de variables.
2. Interpolación entre valores de la función.
3. Predicción y pronóstico de datos.

El interés será determinar una ecuación que relacione una variable dada con otra variable de respuesta, bajo el supuesto que ellas se vinculan mediante una ecuación lineal de primer grado, caso particular conocido como regresión lineal simple.

1.3.1. Modelo lineal simple

Su ecuación de recta es:

$$y = \beta_0 + \beta_1 x$$

Donde:

y: Variable endógena, explicada o respuesta

β_0 : Interpretación de la recta con el eje y o intercepto

β_1 : Pendiente de la recta

x: Variable exógena, explicativa o predictora

ϵ : Error aleatorio

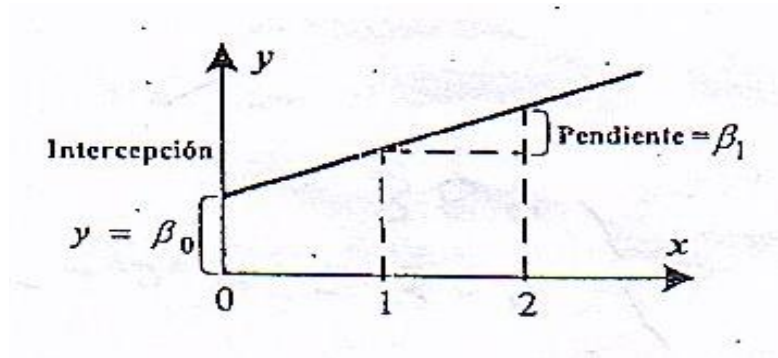


Figura. 1. 3. Ecuación del modelo lineal simple

Este modelo es determinista porque no considera el error y los valores e “y” se obtienen, de manera exacta, al sustituir los valores de “x” en la ecuación de la recta. Sin embargo, cuando se desea incorporar al modelo determinista el efecto aleatorio de las variables se le añade un componente que corresponde al error y el modelo queda como:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Para recoger el efecto aleatorio del error se plantean las siguientes hipótesis respecto a ε (Error aleatorio):

- 1) Se distribuye normalmente con media cero y varianza σ^2 : $\varepsilon \sim N(0, \sigma^2)$.
- 2) Errores correspondientes a observaciones distintas son independientes entre sí: $E(\varepsilon_i, \varepsilon_j) = 0$.

Tabla 1.2. Ejemplos de modelos de regresión

y	x	ε
Presupuestos de gastos de un hogar	Número de miembros del hogar	Efecto del nivel socioeconómico, tenencia de la vivienda, servicios que dispone, etc.
Precio de un departamento	Área de construcción	Efecto de la zona de ubicación, tipo de acabados, piso en el que se encuentra, etc.

y	x	ϵ
Precio de un libro	Número de páginas del libro	Efecto del tipo de papel, la encuadernación, número de ilustraciones, etc.

En análisis de regresión es necesario tener en cuenta los siguientes pasos:

- 1) Tener una visión clara de objetivos del estudio con el fin de determinar cuál he de ser la varia respuesta y que variables puedan incluirse como variables independientes.
- 2) Recopilar los datos correspondientes a las variables identificadas como dependiente e independiente.
- 3) Postular un modelo al que se supone se ajustan los datos, en este caso se presume que es el lineal simple.
- 4) Determinar la ecuación de regresión; es decir, estimar los coeficientes del modelo propuesto.
- 5) Comprobar estadísticamente la adecuación del modelo. Incluye la realización de pruebas estadísticas sobre los parámetros, la ejecución de transformaciones de las variables para tener un mejor ajuste o retirar variables de una ecuación si su aporte no es significativo en la ejecución de predicción.
- 6) Cuando la ecuación sea satisfactoria, usar el modelo para realizar predicciones o estimaciones.

Una vez que se han cumplido los tres primeros pasos, su objetivo será estimar coeficientes del modelo y comprobar la adecuación del modelo.

1.3.2. Estimadores de parámetros

Se ocupa de deducir estimadores puntuales de parámetros $\beta_1, \beta_2, \sigma^2$ tal que se deducirá sus principales propiedades usando únicamente hipótesis H. **Teorema.** La recta de regresión para por el punto (\bar{x}, \bar{y}) .

Se puede comprender mejor la fórmula $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\sigma^2 \bar{x}}{S_{xx}}$. En la primera gráfica de la figura siguiente, \bar{x} es negativa pues cuando aumenta la pendiente de la recta, el punto de corte al eje Y aumenta proporcionalmente. En la gráfica derecha \bar{x} es positiva tal que cuando aumenta la pendiente de la recta, el punto de corte al eje Y baja.

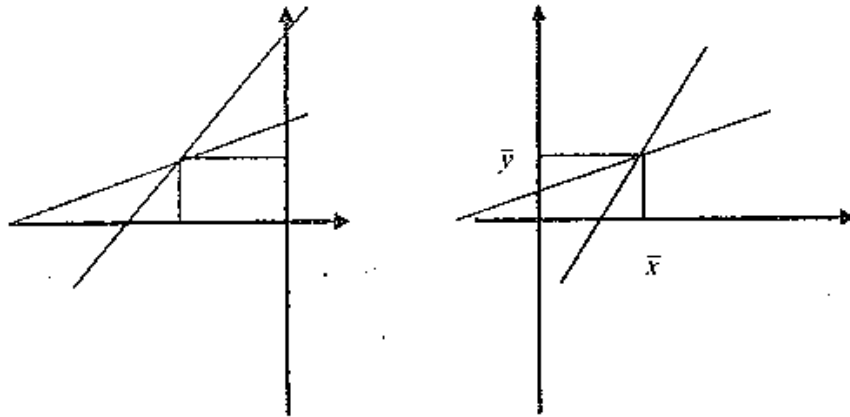


Figura. 1. 4. Comportamiento de las gráficas al aumentar la pendiente de la recta

Definición. Las diferencias $\hat{u}_i = y_i - \hat{y}_i$ se llaman residuos. Estos dan cuenta de errores u_i no observables.

Observación. De ecuaciones $\begin{cases} \bar{y} - b_1 - b_2 \bar{x} = 0 \\ \sum y_i x_i - n \bar{x} b_1 - b_2 \sum x_i^2 = 0 \end{cases}$ se deduce:

$$\sum x_i \hat{u}_i = 0, \sum \hat{u}_i = 0$$

Tal que, la igualdad $\sum \hat{u}_i = 0$ puede ser falsa en un modelo que no incluye el término constante (intercepto) β_1 .

Lema. En toda regresión lineal $\sum_{i=1}^n \hat{u}_i \hat{y}_i = 0$.

Notación. $SRC = \sum_{i=1}^n \hat{u}_i^2$, $SEC = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ y $STC = \sum_{i=1}^n (y_i - \bar{y})^2$.

Donde:

SRC: Suma de Residuos al Cuadrado

SEC: Suma de Estimaciones al Cuadrado

STC: Suma Total de Cuadrado

Teorema. Si β_1 está en modelo o si la regresión no incluye el término constante, pero $\bar{y} = 0$ tal que $STC_{(Suma\ Total\ de\ Cuadrados)} = SEC_{(Suma\ de\ Estimaciones\ al\ Cuadrado)} + SRC_{(Suma\ de\ Residuos\ al\ Cuadrado)}$.

Demostración:

$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 = SRC + SEC + \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$ tal que $\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^n \hat{u}_i \hat{y}_i - \bar{y} \sum_{i=1}^n \hat{u}_i = 0$ tal que la igualdad a cero se tiene por ecuaciones $\begin{cases} \bar{y} - b_1 - b_2 \bar{x} = 0 \\ \sum y_i x_i - n \bar{x} b_1 - b_2 \sum x_i^2 = 0 \end{cases}$ se deduce $\sum x_i \hat{u}_i = 0, \sum \hat{u}_i = 0$ y lema $\sum \hat{u}_i \hat{y}_i = 0$.

Teorema. Bajo hipótesis H y si el término constante β_1 está en modelo tal que:

$$S_R^2 = \frac{SRC}{n - 2}$$

Es estimador insesgado de σ^2 .

Demostración:

$y_i - \hat{y}_i = (y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x})$. Por $\hat{\beta}_2 = \frac{S_{xy}}{S_{xx}}, \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ tal que $y_i - \hat{y}_i = \beta_2(x_i - \bar{x}) + (u_i - \bar{u}) - \hat{\beta}_2(x_i - \bar{x}) = (u_i - \bar{u}) - (\hat{\beta}_2 - \beta_2)(x_i - \bar{x})$.

En consecuencia, $SRC = \sum_{i=1}^n (u_i - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2) \sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x}) + S_{xx}(\hat{\beta}_2 - \beta_2)^2$.

Las esperanzas de términos primero y tercero se calculan de manera sencilla. Viendo el segundo:

$$\begin{aligned} \hat{\beta}_2 &= \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})[\beta_2(x_i - \bar{x}) + (u_i - \bar{u})] \\ &= \beta_2 + \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}) \end{aligned}$$

De donde:

$$\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}) = S_{xx}(\hat{\beta}_2 - \beta_2)$$

Reemplazando el segundo término:

$$SRC = \sum_{i=1}^n (u_i - \bar{u})^2 - S_{xx}(\hat{\beta}_2 - \beta_2)^2$$

$$SRC = E\left(\sum_{i=1}^n (u_i - \bar{u})^2\right) - S_{xx}\text{Var}(\hat{\beta}_2)^2$$

La primera esperanza es conocida en estadística básica:

$$E(SRC) = (n - 1)\sigma^2 - \sigma^2 = (n - 2)\sigma^2$$

Definición. La raíz cuadrada de estimación de varianza se llama error estándar o error típico. El error estándar de $\hat{\beta}_1(\hat{\beta}_2)$ es:

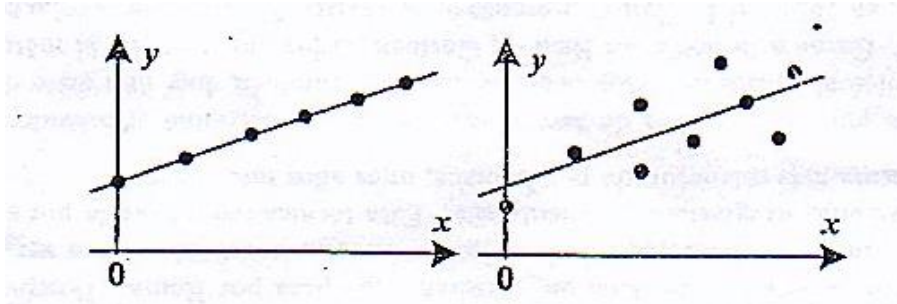
$$ee(\hat{\beta}_1) = \sqrt{S_R^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}, ee(\hat{\beta}_2) = \sqrt{\left(\frac{S_R^2}{S_{xx}}\right)}$$

Tal que las fórmulas que continúan son útiles para estimar sumas de cuadrados.

Teorema. $STC = S_{yy}$. Si el término constante β_1 está en modelo o si datos están centrados:

$$SRC = S_{yy} - \left(\frac{S_{xy}^2}{S_{xx}}\right) \text{ i } SEC = \left(\frac{S_{xy}^2}{S_{xx}}\right)$$

Demostración. Por $y_i - \hat{y}_i = (y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x})$ tal que $SRC = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [(y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x})]^2 = S_{yy} - 2\hat{\beta}_2 S_{xy} + \hat{\beta}_2^2 S_{xx} = S_{yy} - \left(\frac{S_{xy}^2}{S_{xx}}\right)$. La segunda igualdad es consecuencia de $STC = SEC + SRC$.



1.3.3. Método de mínimos cuadrados (MCO)

El método de MCO es el más común en el análisis de regresión, específicamente por ser mucho más intuitivo y matemáticamente más sencillo que el método de máxima verosimilitud. Además, por lo general los dos métodos proporcionan resultados similares.

El método de mínimos cuadrados ordinarios se atribuye al matemático alemán Carl Friedrich Gauss (1777-1855), considerado el matemático más grande del siglo XIX, además de uno de los tres matemáticos más importantes de todos los tiempos (Arquímedes y Newton son los otros dos).

Para estimar los coeficientes de la ecuación de regresión se empleará el método de los mínimos cuadrados, consistente en minimizar la suma de los cuadrados de los errores; esto es, que si se nota a la ecuación de predicción por:

$$\hat{y} = b_0 + b_1x$$

Donde b_0 y b_1 son estimadores de β_0 y β_1 , respectivamente; ellos deben ser tales que la suma de los cuadrados de las diferencias entre los valores observados de la variable respuesta y su estimación por la ecuación de regresión sea mínima. Si se dispone de n pares de observaciones de las variables independientes y dependientes $(x_1; y_1)$, $(x_2; y_2)$, $(x_n; y_n)$ y si son los valores de las predicciones de y :

$$\hat{y} = b_0 + b_1x_i$$

Los residuos de la predicción (errores) se calculan por:

$$y_1 = y_1 - \hat{y}_i$$

Se busca valores de b_0 y b_1 que minimicen la suma de los cuadrados de la también llamada suma de los cuadrados de los residuos:

$$\begin{aligned} \text{SCE} &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2 \end{aligned}$$

Derivando SCE con respecto a b_0 y b_1 , e igualando el resultado a cero se obtiene la ecuación:

$$\begin{aligned} \frac{\partial(\text{SCE})}{\partial b_0} &= -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0, \\ \frac{\partial(\text{SCE})}{\partial b_1} &= -2 \sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) = 0, \end{aligned}$$

Su solución es:

$$\begin{aligned} b_1 \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} &= \frac{SC_{xy}}{SC_{xx}} \\ b_0 &= \bar{y} - b_1 \bar{x}, \end{aligned}$$

Donde $\frac{\sum_{i=1}^n x_i}{n}$ y $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ Son los promedios de los valores de las variables independientes y dependientes. Una vez obtenidos los valores de b_0 y b_1 se los sustituye en la ecuación, de esta manera queda esta la recta de predicción por mínimos cuadrados:

$$\hat{y} = b_0 + b_1 x,$$

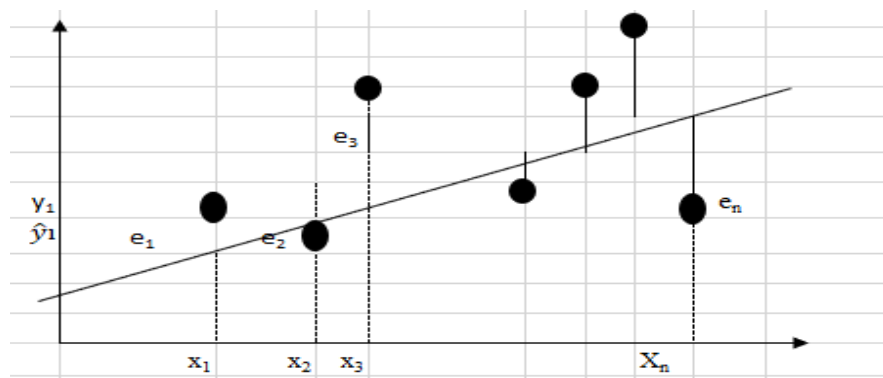


Figura. 1. 6. Comportamiento de las gráficas con base en los errores de nube de puntos

Observación. En la estimación de los parámetros se debe tener presente la incorporación de un redondeo en el cálculo de SC_{XY} y de SC_{XX} ; se recomienda el empleo de un número sufriente en cifras significativas al realizar los cálculos de forma manual.

Ejemplo:

Tabla 1.3. En un estudio para determinar la relación entre el peso de los automóviles y su consumo de combustible se escogió una muestra de 10 carros, con resultados

Variable	Valores numéricos									
Consumo (1/100 Km)	8.00	16.00	6.00	7.00	7.00	9.00	11.00	12.00	18.00	20.00
Peso(kg)	739.00	1187.00	655.00	729.00	888.00	797.00	963.00	802.00	1551.00	1650.00

A partir de los siguientes supuestos el método de mínimos cuadrados presenta propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión, partiendo de la idea que el modelo de Gauss, modelo clásico o estándar de regresión lineal (MCRL) es el cimiento de la mayor parte de la teoría econométrica y plantea siete supuestos clásicos, en sentido que Gauss lo empleó por primera vez en 1821 y desde esta fecha sirve como norma o patrón con que compara modelos de regresión que no satisfacen los supuestos Gaussianos:

I. **Modelo de Regresión es Lineal en los Parámetros.** Aunque, la variable regresada, dependiente o explicada (Y) y la regresora, independiente o explicativa (X) pueden o no ser lineal, inclusive puede incluir más variables explicativas, como se muestra enseguida:

$$Y_i = \beta_1 + \beta_1 X_i + u_i$$

II. **Valores fijos de X, valores de X independientes del término error.** Los valores que toma la variable regresora, independiente o explicativa (X) pueden considerarse fijos en muestras repetidas (regresora fija, no aleatoria, Modelos de Mínimos Clásicos de Regresión Lineal –MCRL– o

Regresora Fija) o haber sido muestreados junto con la variable regresada, dependiente o explicada (Y) (regresora estocástica, aleatoria o Modelos Noeclásico de Regresión Lineal –MNRL– o Regresora Estocástica). En el segundo caso, se supone que la(s) variable(s) X y término error son independientes; es decir, $cov(X_i, u_i) = 0$.

III. El valor medio de perturbación $u_i = 0$ (no hay error de especificación en modelo de regresión elegido). Dado el valor de X_i , la media o valor esperado del término de perturbación aleatoria u_i es cero. Simbólicamente es $E(u_i|X_i) = 0 \Rightarrow E(Y_i|X_i) = \beta_1 + \beta_1 X_i$ o, si X no es estocástica, equivale a $E(u_i) = 0$, pues si la media condicional de una variable aleatoria, dada otra variable aleatoria, es cero, la covarianza entre las dos variables es cero y, por tanto, las dos variables no están correlacionadas o X_i y u_i no están correlacionadas.

Cuando la Función de Regresión Poblacional (FRP) se expresa en una ecuación, se supone que X y u, representando la influencia de todas las variables omitidas, ejercen influencias independientes y aditivas, en Y, pero si X y u están correlacionadas, no es posible evaluar los efectos de cada una sobre Y tal que si X y u tienen correlación positiva, X aumenta cuando u aumenta y, viceversa, disminuye cuando u disminuye. Asimismo, si X y u tienen correlación negativa, X se incrementa cuando u se reduce, disminuye cuando u aumenta. Geométricamente, éste supuesto se representa:

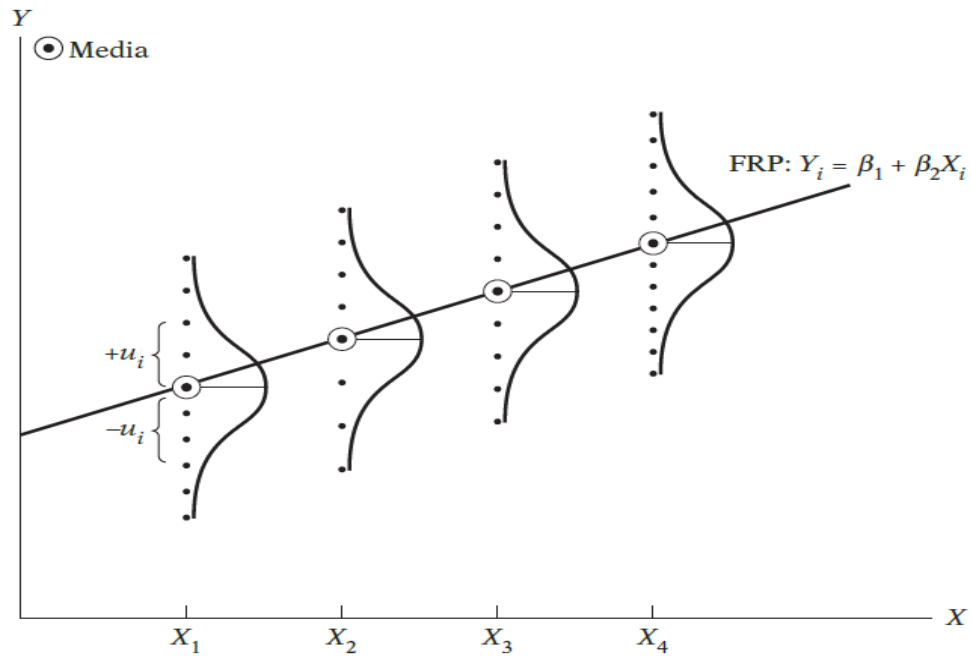


Figura. 1. 7. Representación geométrica del supuesto

IV. **Homocedasticidad o varianza constante de u_i** (del griego skedanime que significa dispersar o esparcir tal que homo es igual y cedasticidad significa dispersión o, en otras palabras, igual varianza). La varianza del término error o de perturbación es la misma sin importan el valor de X. Simbólicamente, se tiene:

$$\begin{aligned} \text{Var}(u_i) &= E[u_i - (Y_i|X_i)]^2 = E(u_i^2|X_i) \text{ por supuesto 3} \\ &= E(u_i^2) \text{ si } X_i \text{ son variables no estocásticas} = \sigma^2 \end{aligned}$$

Esta ecuación establece que la varianza de u_i para cada X_i , varianza condicional de u_i , es algún número positivo constante igual a σ^2 . Por lo tanto, esta ecuación representa el supuesto de homocedasticidad. En términos llanos, la variación alrededor de la línea de regresión (relación promedio entre X y Y) es la misma para todos los valores de X tal que no aumente ni disminuye conforme varía X:

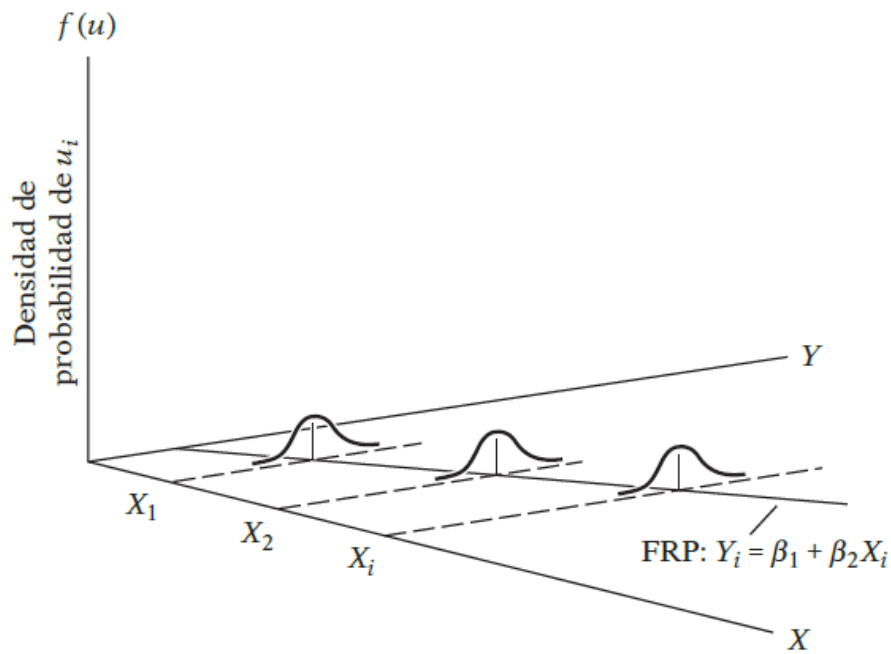


Figura. 1. 8. Variación alrededor de la línea de regresión (relación promedio entre X y Y) es la misma para todos los valores de X

Caso contrario, si se considera la siguiente figura como varianza condicional de la población Y varia con X se conoce apropiadamente como **heterocedasticidad** o dispersión desigual o varianza desigual, escrita como $E(u_i^2|X_i) = \sigma_i^2$ e indica con subíndice i que la varianza de la población Y no es constante:

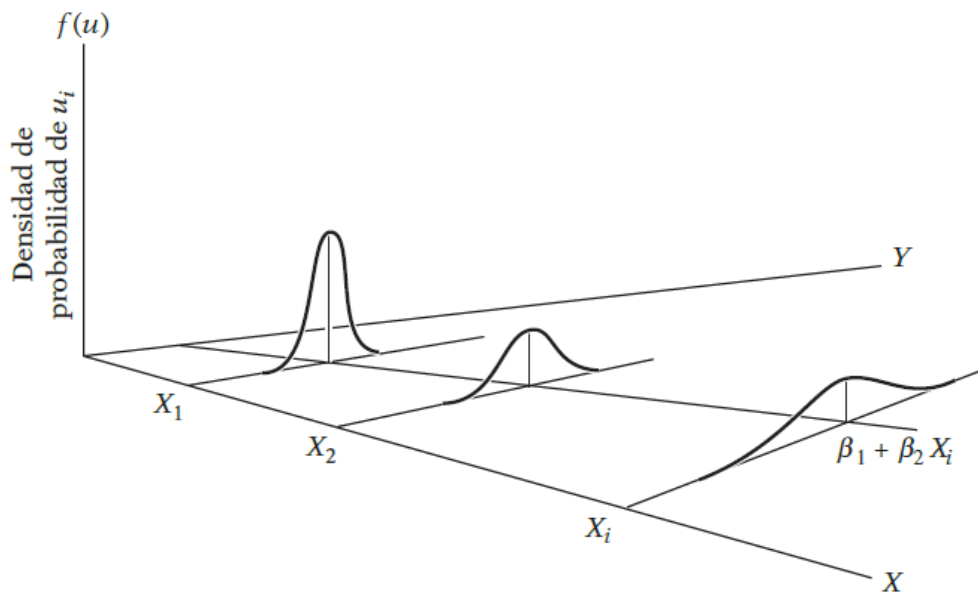


Figura. 1. 9. Varianza condicional de la población Y varia con X

V. **No hay autocorrelación entre perturbaciones** (perturbaciones u_i y u_j no están correlacionadas, supuesto de no correlación serial o no autocorrelación). Dados dos valores cualesquiera de X , X_i y X_j ($i \neq j$), la correlación entre dos u_i y $u_j = 0 \forall (i \neq j)$. En otras palabras, estas observaciones se muestran independientemente y simbólicamente es:

$$\text{cov}(u_i, u_j | X_i, X_j) = 0 \Rightarrow \text{cov}(u_i, u_j) = 0 \text{ si } X \text{ no es estocástica}$$

Esto significa que, dado X_i , las desviaciones de dos valores cualesquiera de Y de sus valores promedio no muestran patrones como en figura a en que las u están **correlacionadas positivamente**, pues a una u positiva sigue una u igual o viceversa, mientras que en figura b las u están **correlacionadas negativamente**, pues a una u positiva sigue una u negativa y viceversa:

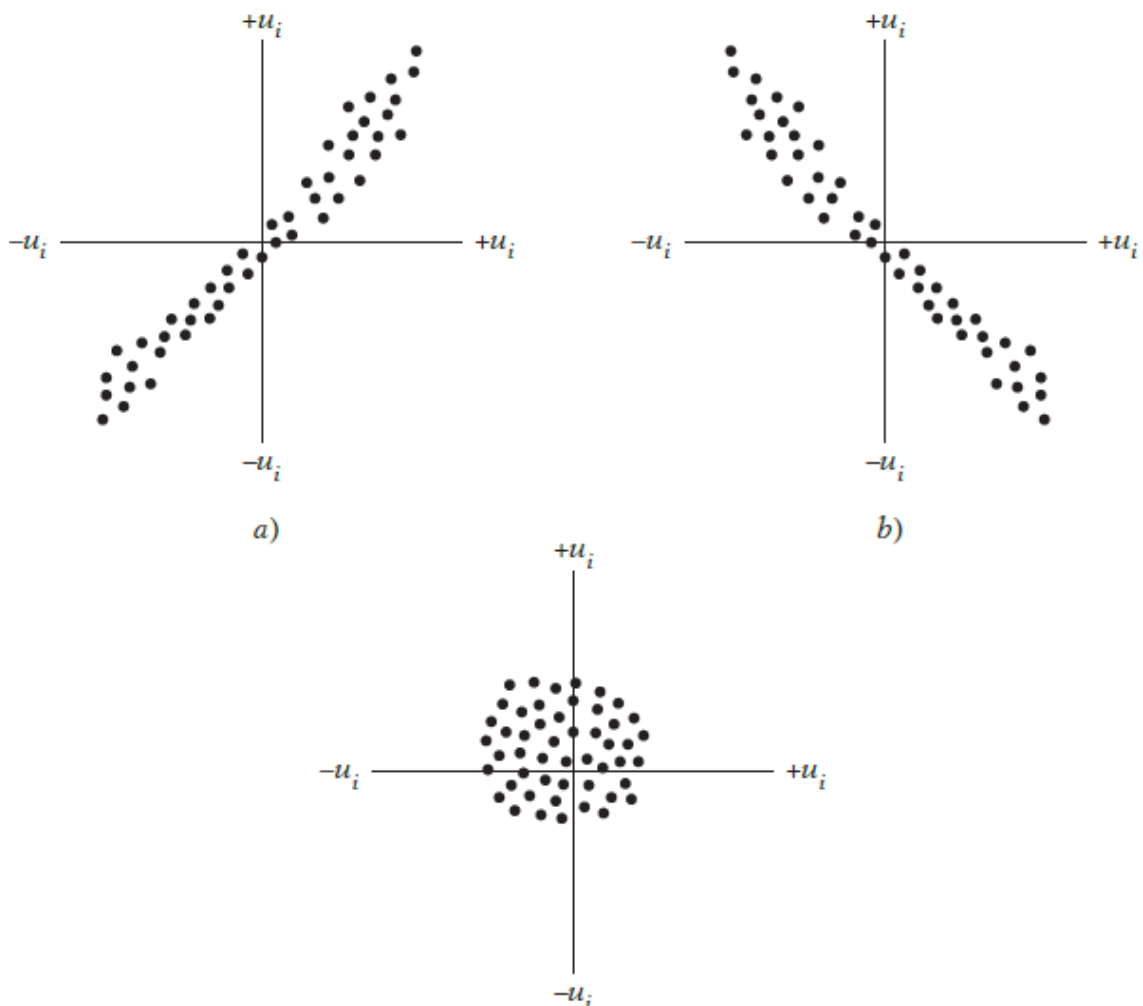


Figura. 1. 10. Patrones de las desviaciones de dos valores cualesquiera de Y respecto a sus valores promedio

Si las perturbaciones (desviaciones) siguen patrones sistemáticos, como figuras a y b, hay correlación serial o autocorrelación mientras que figura c muestra que no hay un patrón sistemático para las u, indica cero correlación.

VI. Número de observaciones n debe ser mayor que número de parámetros por estimar. Sucesivamente, el número de observaciones será $n \geq$ número de variables explicativas.

VII. Naturaleza de variables X (variables deben variar). No todos los valores X en una muestra determinada deben ser iguales. Técnicamente, $\text{Var}(X)$ debe ser un número positivo. Además, no puede haber **valores atípicos** de variable X; es decir, valores muy grandes en relación con el resto de las observaciones.

Esto tiene base según ecuación $\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$, pues si todos los valores X son idénticos, $X_i = \bar{X}$ y el denominador de esta ecuación es cero (indeterminación) imposibilita la estimación de $\hat{\beta}_2$ y, por consiguiente, de $\hat{\beta}_1$. Por lo tanto, la variación tanto en Y como X es esencial para utilizar el análisis de regresión como herramienta de investigación.

Entonces, ¡las variables deben variar! El requisito que no existan valores atípicos de X es para evitar que resultados de regresión estén dominados por estos valores. Si hay algunos valores X que, por ejemplo, sean x veces el promedio de valores X, las líneas de regresión estimadas con o sin dichas observaciones serían muy diferentes. Con frecuencia estos valores atípicos son resultado de errores humanos de aritmética o de mezclar muestras de diferentes poblaciones.

Según (Castro, 2008), si se reemplaza parámetros desconocidos β_1, β_2 por dos números $b_1, b_2 \in \mathbb{R}$ se produce una discrepancia entre $b_1 + b_2 x_i$ y y_i escrita como

e_i tal que $e_i = y_i - (b_1 + b_2x_i)$ donde la idea central es minimizar suma de desvíos al cuadrado:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_1 - b_2x_i)^2$$

Tal que se puede definir números $\hat{\beta}_1, \hat{\beta}_2$ que minimizan $SRC(b_1, b_2) = \sum_{i=1}^n (y_i - b_1 - b_2x_i)^2$ donde $b_1, b_2 \in \mathbb{R}$ se llaman estimadores de mínimos cuadrados (EMC) de parámetros β_1, β_2 . Además, la notación:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2, S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Demostración:

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

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

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

Teorema. Bajo hipótesis H, estimadores de mínimos cuadrados (MC) de modelo $y_i = \beta_1 + \beta_2 X_i + u_i$ $i = 1, 2, 3, 4, \dots, n$ son:

$$\hat{\beta}_2 = \frac{S_{xy}}{S_{xx}} \text{ tq } \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

Demostración. Para minimizar la función $SRC(b_1, b_2) = \sum_{i=1}^n (y_i - b_1 - b_2x_i)^2$ se estiman derivadas parciales:

$$\frac{\partial SRC(b_1, b_2)}{\partial(b_1)} = -2 \sum_{i=1}^n (y_i - b_1 - b_2x_i)$$

$$\frac{\partial SRC(b_1, b_2)}{\partial(b_2)} = -2 \sum_{i=1}^n (y_i - b_1 - b_2x_i) x_i$$

Igualando a cero, se tiene el sistema de ecuaciones llamado Sistema Normal de Ecuaciones:

$$\begin{cases} \bar{y} - b_1 - b_2\bar{x} = 0 \\ \sum y_i x_i - n\bar{x}b_1 - b_2 \sum x_i^2 = 0 \end{cases}$$

Su solución está dada por el enunciado del teorema anterior.

1.3.4. Propiedades de estimadores MC

Definición. Cualquier estimador que sea combinación lineal de observaciones y_i se dice que es un estimador lineal, pues:

$$\hat{\beta}_2 = \frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})y_i = \sum_{i=1}^n a_i y_i \text{ tal que } a_i = \frac{x_i - \bar{x}}{S_{xx}}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} = \frac{1}{n} \sum_{i=1}^n y_i - \sum_{i=1}^n \bar{x} a_i y_i = \sum_{i=1}^n b_i y_i \text{ tal que } b_i = \frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}}$$

Teorema. Bajo hipótesis H, estimadores de MC son insesgados y:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{S_{xx}}, \text{Var}(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right), \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -\sigma^2 \left(\frac{\bar{x}}{S_{xx}} \right)$$

Demostración. Las igualdades $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} = \frac{1}{n} \sum_{i=1}^n y_i - \sum_{i=1}^n \bar{x} a_i y_i$ y $\hat{\beta}_2 = \frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ facilitan estas demostraciones, tal que por linealidad de esperanza:

$$E(\hat{\beta}_2) = \sum_{i=1}^n a_i E(y_i) = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} (\beta_1 + \beta_2 x_i) = \beta_2 \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} (x_i) = \beta_2$$

Tal que por correlación de variables y_i :

$$\text{Var}(\hat{\beta}_2) = \sum_{i=1}^n a_i^2 \text{Var}(y_i) = \sigma^2 \sum_{i=1}^n a_i^2 = \frac{\sigma^2}{S_{xx}}$$

Las otras igualdades se demuestran de forma semejante. La igualdad $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -\sigma^2 \left(\frac{\bar{x}}{S_{xx}} \right)$ manifiesta que estimadores $\hat{\beta}_1, \hat{\beta}_2$ son no correlacionados sólo si $\bar{x} = 0$, caso contrario si $\bar{x} > 0$, varía en sentidos opuestos; si la pendiente aumenta el punto de corte al eje Y baja:

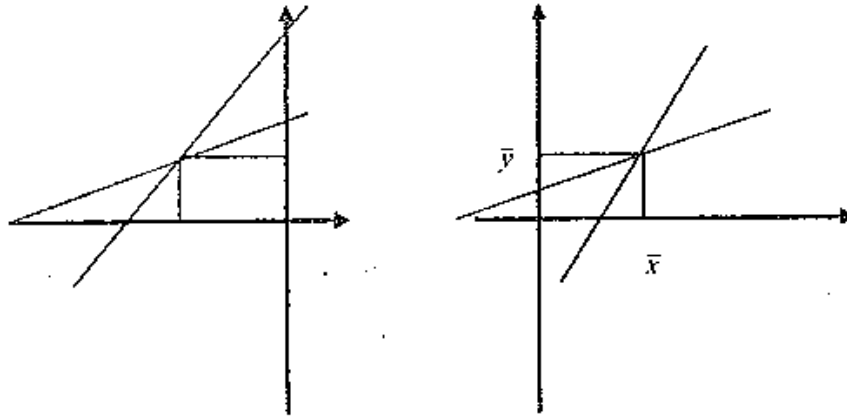


Figura. 1. 11. Patrones de desigualdades según aumento de la pendiente

Tal que se puede estimar la varianza de cualquier combinación lineal de estimadores:

$$\begin{aligned} \text{Var}(a\hat{\beta}_1 + b\hat{\beta}_2) &= a^2\text{Var}(\hat{\beta}_1) + 2ab\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) + b^2\text{Var}(\hat{\beta}_2) \\ &= \sigma^2 \left[a^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) - 2ab \left(\frac{\bar{x}}{S_{xx}} \right) + \frac{b^2}{S_{xx}} \right] \\ &= \sigma^2 \left[\frac{a^2}{n} + \frac{1}{S_{xx}} (a^2\bar{x}^2 - 2ab\bar{x} + b^2) \right] = \sigma^2 \left[\frac{a^2}{n} + \frac{(b - a\bar{x})^2}{S_{xx}} \right] \end{aligned}$$

1.3.5. Propiedades de estimadores de mínimos cuadrados ordinarios según supuestos de normalidad

Suponga que u_i sigue una distribución normal, pues supone que cada u_i está Normalmente Distribuida con Media ($E(u_i) = 0$), Varianza ($E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$), $\text{Cov}(u_i, u_j) = E\{[u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i, u_j) = 0$ $i \neq j$), expresados en forma resuma como $u_i \sim N(0, \sigma^2)$, por lo que con el supuesto de normalidad la ecuación anterior indica que u_i y u_j no están correlacionadas, sino

que están distribuidas independientemente tal que $u_i \sim \text{Normal e Independientemente Distribuido } (0, \sigma^2)$, los estimadores de MCO tienen las propiedades siguientes:

1. Son **Insesgados**.
2. Tienen **Varianza Mínima**. En combinación con 1 significa que son **estimadores insesgados con varianza mínima o eficientes**.
3. Presentan **Consistencia**. A medida que el tamaño muestral aumenta indefinidamente, los estimadores convergen hacia sus verdaderos valores poblacionales.
4. $\hat{\beta}_1$ está **Normalmente Distribuida** al ser función lineal de u_i con Media ($E(\hat{\beta}_1) = \beta_1$), $\text{Var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$, en forma más completa $\hat{\beta}_1 \sim N(\beta_1, \sigma_{\hat{\beta}_1}^2)$. Tal que, de acuerdo con las propiedades de distribución normal, variable Z es definida como $Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}}$ sigue una **distribución normal estándar**; es decir, una distribución normal con media cero y varianza unitaria (= 1) ó, en otras palabras, $Z \sim N(0, 1)$.
5. $\hat{\beta}_2$ está **Normalmente Distribuida** al ser función lineal de u_i con Media ($E(\hat{\beta}_2) = \beta_2$), $\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$, en forma más completa $\hat{\beta}_2 \sim N(\beta_2, \sigma_{\hat{\beta}_2}^2)$. Entonces, $Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$ sigue una distribución normal estándar.

Geométricamente, las distribuciones de probabilidad de $\hat{\beta}_1$ y $\hat{\beta}_2$ son:

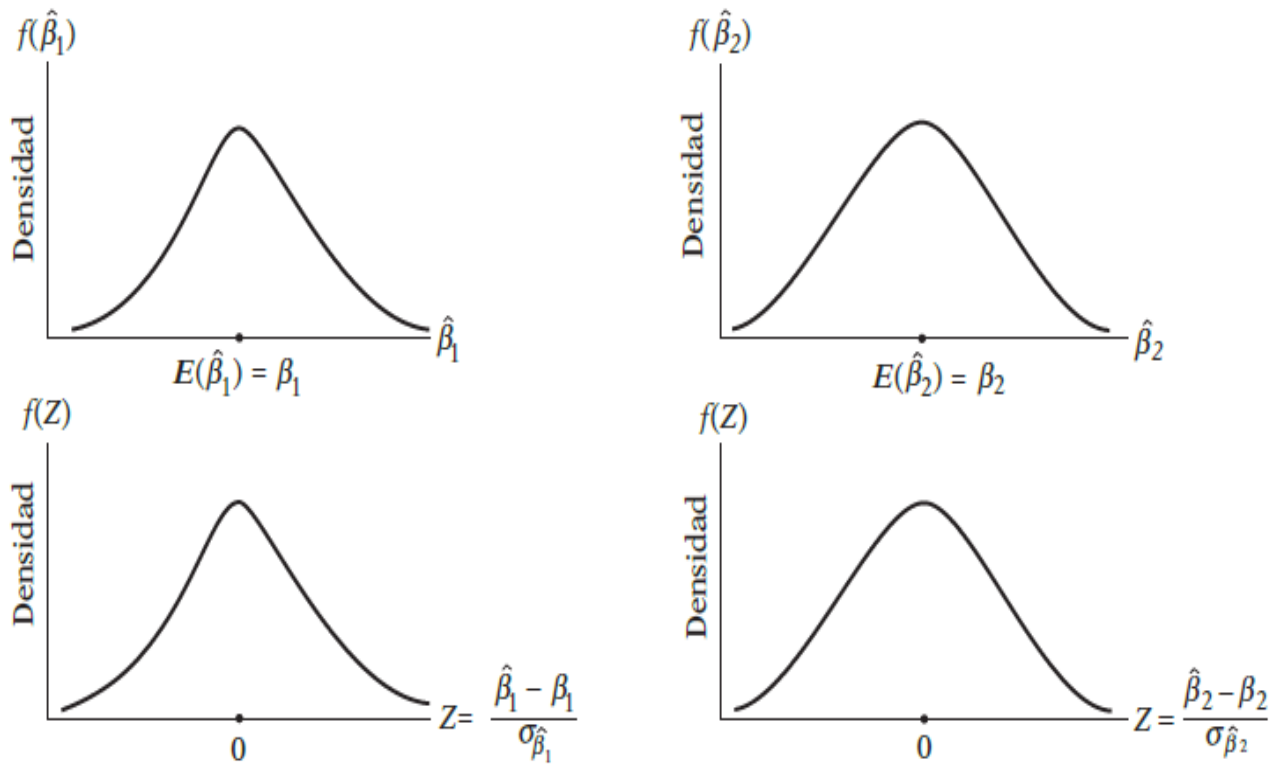


Figura. 1. 12. Distribuciones de probabilidad de $\hat{\beta}_1$ y $\hat{\beta}_2$

6. Como la **Distribución Ji Cuadrada (χ^2)** está distribuida $(n - 2)$ (Grads de libertad) $\left(\frac{\hat{\sigma}^2}{\sigma^2}\right)$. Esto ayuda a hacer inferencias respecto a la verdadera σ^2 a partir de $\hat{\sigma}^2$.

7. $(\hat{\beta}_1, \hat{\beta}_2)$ se distribuyen de manera independiente respecto a $\hat{\sigma}^2$.

8. $\hat{\beta}_1$ y $\hat{\beta}_2$ tienen varianza mínima entre todas las clases de **estimadores insesgados, lineales o no lineales**. Es eficaz debido a que, a diferencia del Teorema de Gauss-Markov, no se limita a la clase de estimadores lineales. Por tanto, los **Estimadores de Mínimos Cuadrados son los Mejores Estimadores Lineales Insesgados (MELI)**, pues tienen varianza mínima en toda clase de estimadores insesgados.

Finalmente, el supuesto de normalidad permite derivar las distribuciones de probabilidad o muestrales de $\hat{\beta}_1$ y $\hat{\beta}_2$ (ambas normales) y de $\hat{\sigma}^2$ relacionadas con Ji Cuadrada (χ^2).

CAPÍTULO II

MÁXIMA VEROSIMILITUD (MV)

CAPÍTULO II

MÁXIMA VEROSIMILITUD (MV)

2.1. Estimación de máxima verosimilitud del modelo de regresión con dos variables

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 varianzas $= \sigma^2$, 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 , esta función de densidad de probabilidad conjunta se escribe como el producto de las n funciones de densidad individuales como:

$$\begin{aligned} & f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2) \\ &= f(Y_1 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_3 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_4 | \beta_1 + \beta_2 X_i, \sigma^2) \\ & \quad \dots f(Y_n | \beta_1 + \beta_2 X_i, \sigma^2) \end{aligned}$$

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 varianzas dadas.

Sin embargo, al sustituir la ecuación anterior en ecuación $f(Y_1, Y_2, Y_3, Y_4, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2) = f(Y_1 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_3 | \beta_1 + \beta_2 X_i, \sigma^2) f(Y_4 | \beta_1 + \beta_2 X_i, \sigma^2)$

... $f(Y_n | \beta_1 + \beta_2 X_i, \sigma^2)$ se tiene:

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

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):

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

2.2. 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:

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

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 $\hat{\cdot}$ (tilde o angular) para estimadores de MV mientras que $\hat{\cdot}$ (acento circunflejo) para estimadores de MCO, denotan estimadores de MV para obtener:

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

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

$$\begin{aligned} \sum Y_i &= n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \\ \sum Y_i X_i &= \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \end{aligned}$$

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

$$\begin{aligned} \sum Y_i &= n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \approx \sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \\ \sum Y_i X_i &= \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \approx \sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \end{aligned}$$

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

$$\begin{aligned} \hat{\beta}_2 &= \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2 \bar{X} \end{aligned}$$

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

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

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, $Y_i = \beta_1 + \beta_2 X_i + u_i \Rightarrow \bar{Y} = \beta_1 + \beta_2 \bar{X}_i + \bar{u}$ 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:

$\sum \hat{u}_i^2 = (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + \sum (u_i - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u})$ tal que al tomar valores esperados en ambos lados se tiene

$$E(\sum \hat{u}_i^2) = \sum x_i^2 E(\hat{\beta}_2 - \beta_2)^2 + E[\sum (u_i - \bar{u})^2] - 2E[(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u})] = \sum x_i^2 \text{Var}(\hat{\beta}_2) + (n-1)\text{Var}(u_i) - 2E[\sum x_i u_i (x_i u_i)] = \sigma^2 + (n-1)\sigma^2 - 2E[\sum k_i x_i u_i^2] = (n-2)\sigma^2$$

donde si 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:

$$E(\hat{\beta}_2) = \beta_2 + \sum k_i E(u_i) = \beta_2.$$

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 $\tilde{\sigma}^2 = \frac{1}{n} \sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = \frac{1}{n} \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 = \frac{1}{n} \sum \hat{u}_i^2$ 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 $E(\sum \hat{u}_i^2) = (n-2)\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2$ 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 $E(\sum \hat{u}_i^2) = (n-2)\sigma^2 = \sigma^2 - \frac{2}{n}\sigma^2$, 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 $n \rightarrow \infty$. 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 .

2.3. Estimación de MCO y MV de Coeficientes de Regresión Parcial

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). 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 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 $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ tal que \hat{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 $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ 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)**). Gráficamente esto es:

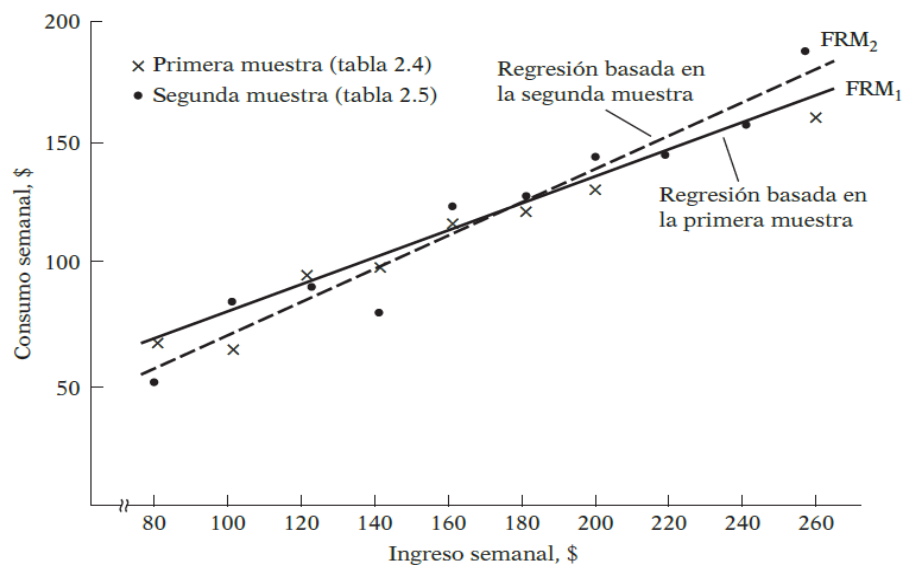
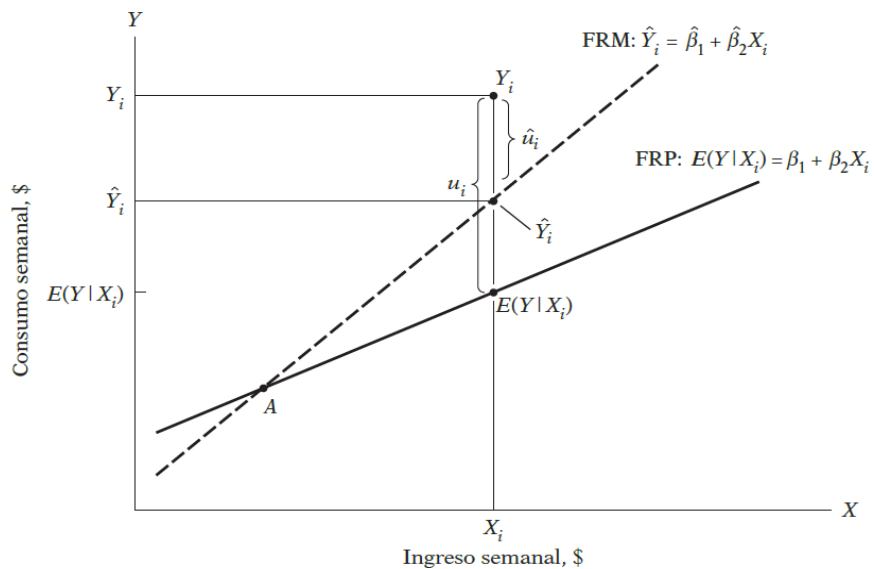


Figura. 1. 13. Diagrama de dispersión con dos líneas de regresión muestral con el fin de “ajustar” razonablemente bien las dispersiones FRM₁ y FRM₂



Fundamentos matemáticos de regresión lineal

Figura. 1. 14. Estimación de la Función de Regresión Poblacional (FRP) $Y_i = \beta_1 + \beta_2 X_i + u_i$ con base en Función de Regresión Muestral (FRM) $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

Para $X = X_i$ se tiene una observación muestral, $Y = Y_i$ que en términos de FRM, Y_i observada es $Y_i = \hat{Y}_i + \hat{u}_i$ y en términos de RFP $Y_i = E(Y|X_i) = u_i$.

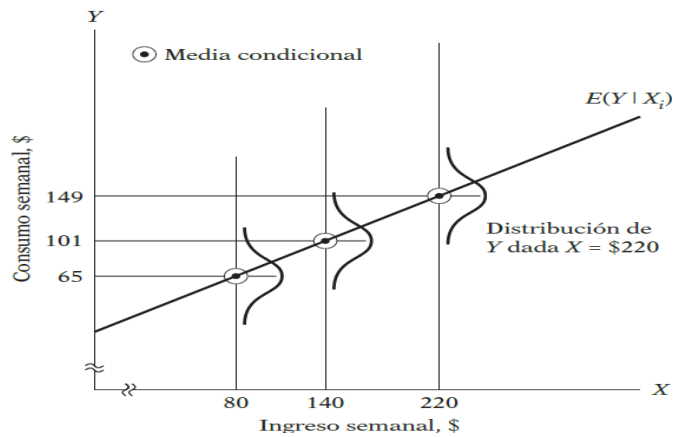
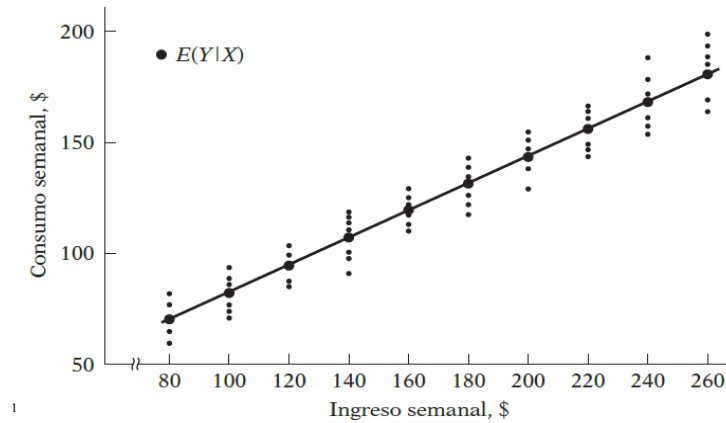


Figura. 1. 15. Representación simbólica de $E(Y|X_i) = f(X_i)$ tal que $f(X_i)$ denota alguna función de variable explicativa X . $E(Y|X_i)$ es una función lineal de X_i

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 . 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 varía con X . $E(Y|X_i) = \beta_1 + \beta_2 X_i$.

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). 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 $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$:

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

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:

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

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 $\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$ diferenciarla parcialmente respecto de las incógnitas, igualar a cero las expresiones resultantes y resolverlas al mismo tiempo. De $\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$ respecto de tres incógnitas e igualar a cero las ecuaciones resultantes se obtiene $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) (-1)$, $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) (-1 * X_{2i})$ y $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) (-1 * X_{3i})$, 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:

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

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

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

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

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

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 $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$, $\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$ 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 \mathbf{X}_2 y \mathbf{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.

2.4. Teorema de Gauss-Markov

Aunque se conoce como Teorema de Gauss-Markov, el método de Gauss de mínimos cuadrados en 1821 antecede al de Markov de varianza mínima en 1900. En contexto de regresión puede probarse que estimadores de MCO son MELI, basado en Teorema de Gauss-Markov: “Dados los supuestos del modelo clásico de regresión lineal, los estimadores de mínimos cuadrados, dentro de la clase de estimadores lineales insesgados, tienen varianza mínima o, en otras palabras, son MELI”.

Al resolver ecuaciones normales al mismo tiempo se obtiene $\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$, que con identidades algebraicas simples, esta fórmula se usa para estimar β_2 como:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum X_i Y_i}{\sum X_i^2 - n\bar{X}^2}$$

Donde $\sum x_i^2 = \sum (X_i - \bar{X})^2 = \sum X_i^2 - 2 \sum X_i \bar{X} + \sum \bar{X}^2 = \sum X_i^2 - 2\bar{X} \sum X_i + \sum \bar{X}^2$, pues \bar{X} es una constante. Además, $\sum X_i = n\bar{X}$ y $\sum \bar{X}^2 = n\bar{X}^2$, pues \bar{X} es una constante.

Finalmente, $\sum x_i^2 = \sum X_i^2 - n\bar{X}^2$. Los estimadores obtenidos se conocen como **estimadores de mínimos cuadrados**, pues se derivan del principio de mínimos cuadrados.

Las **propiedades numéricas** de estimadores obtenidos con método de MCO, que se mantienen como consecuencia del uso de mínimos cuadrados ordinarios sin considerar la forma como se generaron los datos, son:

- I. **Estimadores de MCO se expresan únicamente en términos de cantidades (X y Y) observables (muestras).** Por consiguiente, se calculan con facilidad.
- II. **Son estimadores puntuales.** Dada la muestra, cada estimador proporciona un solo valor (puntual) del parámetro poblacional pertinente, mientras que **estimadores por intervalos** proporcionan un intervalo de valores posibles para parámetros poblacionales no conocidos.
- III. **Obtenidos los estimadores de MCO de datos muestrales, se obtiene sin problemas la línea de regresión muestral:**

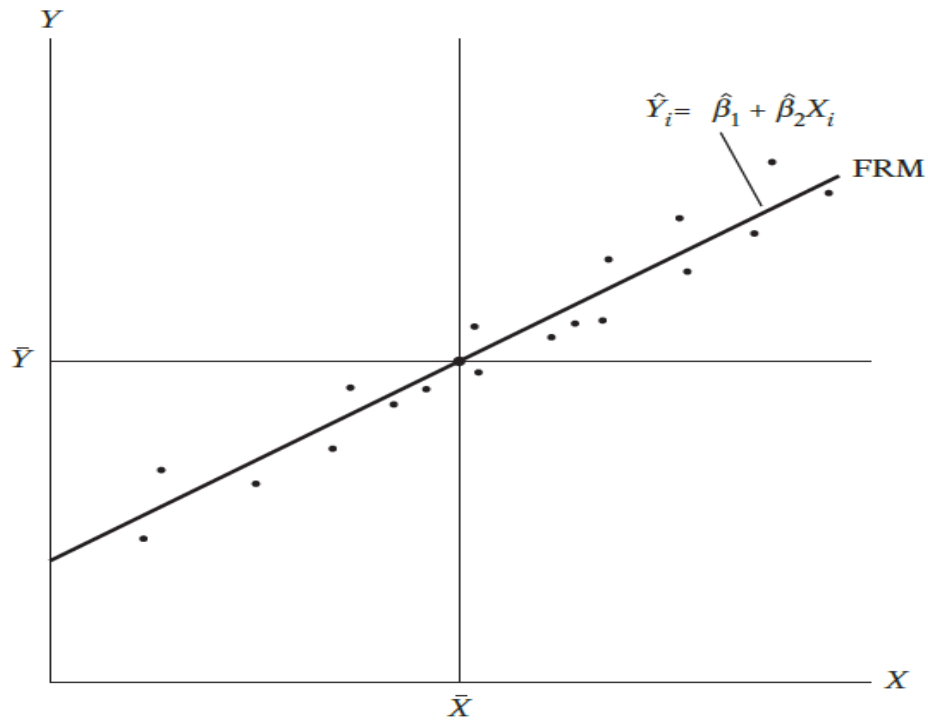


Figura. 2. 1. Obtención de la línea de regresión muestral mediante Función de Regresión Muestral (FRM)

Presenta las siguientes propiedades:

1. Pasa a través de medias muestrales de Y y X, evidente por ecuación

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y}{n \sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2 \bar{X}, \text{ pues esta ecuación puede escribirse } \bar{Y} =$$

$$\hat{\beta}_1 + \hat{\beta}_2 \bar{X}:$$

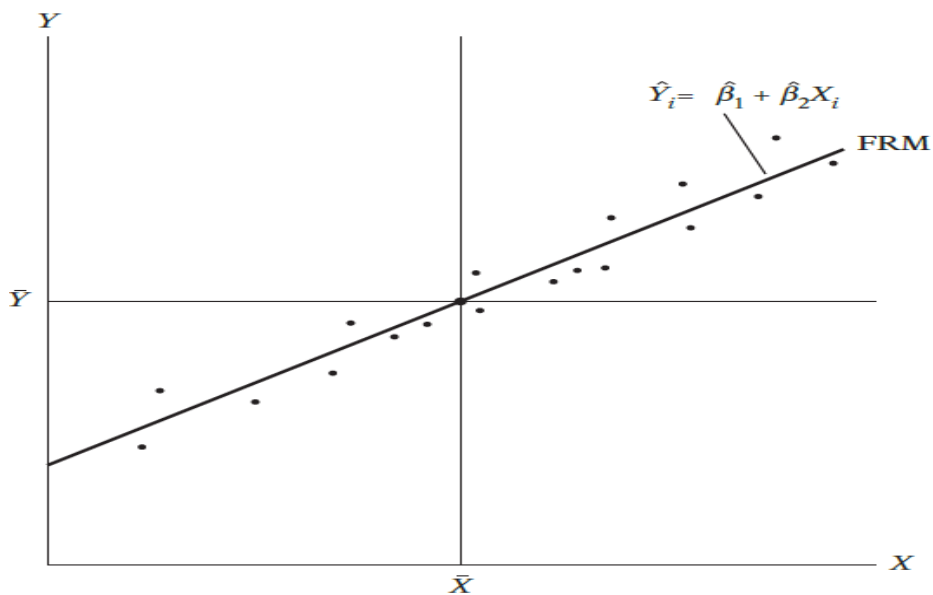


Figura. 2. 2. Obtención de la Función de Regresión Muestral (FRM) que pasa a través de medias muestrales de Y y X

2. El valor medio de Y estimada es igual a \hat{Y}_i es igual al valor medio de Y real para:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 (X_i - \bar{X})$$

Al sumar ambos lados de esta última igualdad sobre valores muestrales y dividir por el tamaño n de muestra se obtiene que $\hat{Y} = \bar{Y}$, tal que se aprovecha que $\sum(X_i - \bar{X}) = 0$.

3. El valor medio de residuos $\hat{u}_i = 0$, pues al diferenciar parcialmente $\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$ respecto de $\hat{\beta}_1$ y $\hat{\beta}_2$ se obtiene:

$$\frac{\partial(\sum \hat{u}_i^2)}{\partial \hat{\beta}_1} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2 \sum \hat{u}_i$$

Complementariamente:

$$\frac{\partial(\sum \hat{u}_i^2)}{\partial \hat{\beta}_2} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) X_i = -2 \sum \hat{u}_i X_i$$

Si se igualan a cero, se simplifican y manipulan algebraicamente se obtiene estimadores dados en ecuaciones $\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$ y $\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y}{n \sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2 \bar{X}$. Entonces, la ecuación $2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$, pero como $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$, la ecuación anterior se reduce a $2 \sum \hat{u}_i = 0 \Rightarrow \hat{u}_i = 0$, aunque requiere que el término del intercepto β_1 esté presente en el modelo.

Como resultado de la propiedad anterior, la regresión muestral es:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}$$

Puede definirse de forma en que Y y X se expresan como desviaciones de sus medias. FRM se limita a la población de valores Y que corresponden a valores fijos X, tal que con toda deliberación evita consideraciones muestrales; es decir, datos

poblacionales. Sin embargo, en la práctica lo que se obtiene al alcance no es más que una muestra de valores Y que corresponden a algunos valores fijos de X .

Por lo tanto, la labor es estimar la FRP (Función de Regresión Poblacional) con base en información muestral. Se pueden tomar dos o más muestras y representarse mediante **lineal de regresión muestral** tal que se obtendrían N FRM diferentes para N muestras diferentes y estas FRM no por fuerza son iguales.

Al igual que FRP, basada en línea poblacional, se desarrolla el concepto de **función de regresión muestral (FRM)** para representar la línea de regresión muestral que se escribe como \hat{Y}_i (Estimador de $E(Y|X_i)$) = $\hat{\beta}_1 + \hat{\beta}_2 X_i$. Un **estimador o estadístico** (muestral) es una regla, fórmula o método para estimar el parámetro poblacional a partir de información suministrada por la muestra disponible.

La FRP se expresa de dos maneras, $E(Y|X_i) = \beta_1 + \beta_2 X_i$ o $Y_i = E(Y|X_i) + u_i = \beta_1 + \beta_2 X_i + u_i$, la FRM se expresa en su forma estocástica como $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$. Por lo tanto, el objetivo del análisis de regresión es estimar FRP, $Y_i = \beta_1 + \beta_2 X_i + u_i$, con base en FRM, $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$ aunque, por fluctuaciones, la estimación de FRP basada e FRM es, en mejor de los casos, una aproximación.

Con base en esto, las **propiedades de linealidad e insesgamiento de estimadores de mínimos cuadrados** se muestran:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i Y_i$$

Donde $k_i = \frac{x_i}{(\sum x_i^2)}$ muestra que $\hat{\beta}_2$ es un estimador lineal, pues es una función lineal de Y ; de hecho, es un promedio ponderado de Y_i donde k_i representa las ponderaciones. Igualmente, se demuestra que $\hat{\beta}_1$ es un estimador lineal. Ahora sustituya la Función de Regresión Poblacional (FRP) $Y_i = \beta_1 + \beta_2 X_i + u_i$ en ecuación $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i Y_i$ para obtener:

$$\hat{\beta}_2 = \sum k_i (\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i = \beta_2 \sum k_i u_i$$

Donde se emplean las propiedades de k_i :

1. Como se supuso que X_i son no estocásticas, las k_i también son no estocásticas.
2. $\sum k_i = 0$.
3. $\sum k_i^2 = \frac{1}{\sum x_i^2}$.
4. $\sum k_i x_i = \sum k_i Y_i = 1$

Estas propiedades se verifican directamente con definición de k_i . Por ejemplo:

$$\sum k_i = \sum \left(\frac{x_i}{\sum x_i^2} \right) = \frac{1}{\sum x_i^2} \sum x_i = 0$$

Pues para una muestra dada se conoce $\sum x_i^2$ y es $= 0$ debido a que $\sum x_i$, suma de desviaciones de la media, es siempre cero. Tal que, al obtener los valores esperados de ecuación $\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i = \beta_2 \sum k_i u_i$ para ambos lados y advertir que las k_i , al ser no estocásticas, pueden tratarse como constantes, por lo que se obtiene:

$$E(\hat{\beta}_2) = \beta_2 + \sum k_i E(u_i) = \beta_2$$

Esto es porque $E(u_i) = 0$ por sustitución. En consecuencia, $\hat{\beta}_2$ es un estimador insesgado de β_2 e, igualmente, se demuestra que $\hat{\beta}_1$ también lo es de β_1 . Para apreciar lo anterior, sume $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$ en ambos lados para obtener:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + \sum \hat{u}_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + 0$$

Al dividir la anterior ecuación entre n , se obtiene:

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

Si se resta la ecuación $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$ de $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$, se obtiene:

$$Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i \text{ o } y_i = \hat{\beta}_2 x_i + \hat{u}_i$$

Donde y_i y x_i según lo escrito, representan desviaciones de valores respectivos de sus medias muestrales y esta ecuación se llama **forma de desviación**. $\hat{\beta}_1$ puede estimarse mediante $\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2 \bar{X}$, pues la línea de regresión muestral pasa mediante medias muestrales de Y y X.

Una ventaja de la forma de desviación es que está simplificada a menudo los cálculos de fórmulas. Es importante mencionar que la forma de desviación de Función de Regresión Muestral (FRM) es $\hat{y}_i = \hat{\beta}_2 x_i$ mientras que las unidades de medición originales, dicha expresión era $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$.

4. Residuos \hat{u}_i no están correlacionados con el valor pronosticado de Y_i , que se verifica de la manera siguiente:

$$\sum \hat{y}_i \hat{u}_i = \hat{\beta}_2 \sum x_i \hat{u}_i = \hat{\beta}_2 \sum x_i (y_i - \hat{\beta}_2 x_i) = \hat{\beta}_2^2 \sum x_i^2 - \hat{\beta}_2^2 \sum x_i^2 = 0$$

Donde se aprovecha que $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$.

5. Los residuos \hat{u}_i no están correlacionados con X_i ; es decir, $\sum \hat{u}_i X_i = 0$.

Esto se desprende de la ecuación $\frac{\partial (\sum \hat{u}_i^2)}{\partial \hat{\beta}_2} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) X_i = -2 \sum \hat{u}_i X_i$.

El Teorema de Gauss-Markov tiene importancia teórica y práctica a la vez. Enseguida se presenta la **distribución muestral** del estimador de MCO $\hat{\beta}_2$; es decir, la distribución de valores asumidos por $\hat{\beta}_2$ en experimentos repetidos de muestreo:

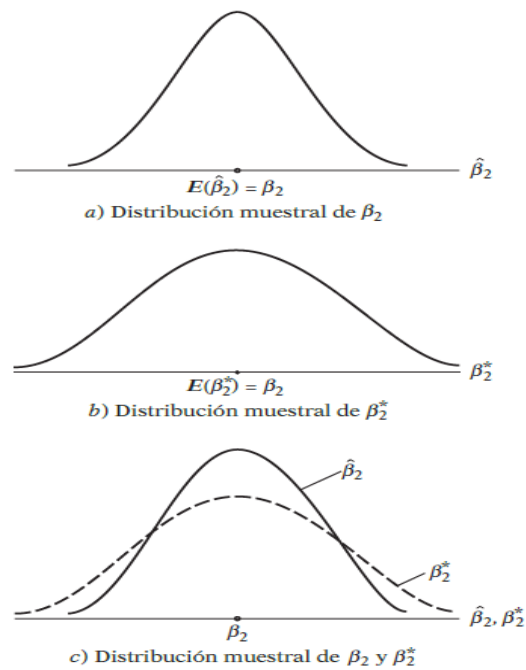


Figura. 2. 3. Distribución muestral de estimador de MCO $\hat{\beta}_2$ y estimador alternativo $\hat{\beta}_2^*$

La media de los valores $\hat{\beta}_2$, $E(\hat{\beta}_2)$, es igual al verdadero valor β_2 . Por lo que, se dice que $\hat{\beta}_2$ es un estimador insesgado de β_2 . Por conveniencia, suponga que $\hat{\beta}_2^*$, al igual que $\hat{\beta}_2$ es insesgado; es decir, que su valor promedio o esperado es igual a β_2 . Además, suponga que $\hat{\beta}_2$ y $\hat{\beta}_2^*$ son estimadores lineales; es decir, funciones lineales de Y .

No obstante, si bien $\hat{\beta}_2$ y $\hat{\beta}_2^*$ son insesgados, la distribución $\hat{\beta}_2^*$ es más difusa o dispersa alrededor del valor de la media que la distribución de $\hat{\beta}_2$. Es decir, la varianza de $\hat{\beta}_2^*$ es mayor que la varianza de $\hat{\beta}_2$. Asimismo, dados dos estimadores a la vez lineales e insesgados, sería preferible el estimador con la menor varianza, pues es probable que esté más cerca de β_2 , en comparación del alternativo.

Por lo tanto, se escogería el Mejor Estimador Lineal Insesgado (MELI).

El Teorema de Gauss-Markov no hace ninguna suposición respecto de distribución de probabilidad de variable aleatoria \hat{u}_i y, por consiguiente, tampoco respecto de Y_i . En medida que satisfagan los supuestos de Mínimos Cuadrados de Regresión Lineal (MCRL), el teorema será válido. No se requiere buscar otro

estimador insesgado lineal, pues no habrá otro estimador cuya varianza sea más pequeña que la del estimador de MCO.

Por supuesto, si no se cumple una o más de tales suposiciones, el teorema ya no es válido. Las propiedades antes vistas se conocen como **propiedades de muestras finitas**, pues se mantienen sin importar el tamaño de muestra en que se basen los estimadores.

Sin embargo, para entender este teorema se requiere considerar la **propiedad del mejor estimador lineal insesgado**: “ $\hat{u}'\hat{u} = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$ con reglas de diferenciación se obtiene $\frac{\partial(\hat{u}'\hat{u})}{\partial\hat{\beta}} = -2X'y + 2X'X\hat{\beta} \Rightarrow \frac{2X'y}{2} = X'X\hat{\beta} \Rightarrow X'X\hat{\beta} = X'y \therefore \hat{\beta} = (X'X)^{-1}X'y$ siempre que exista la inversa. Como $(X'X)^{-1}X'y$ es una matriz de números fijos, $\hat{\beta}$ es una función lineal de Y tal que, por definición, es un estimador lineal.

Es importante recordar que la Función de Regresión Poblacional (FRP) es $y = X\beta + u$, si se sustituye esta última ecuación en $\hat{\beta} = (X'X)^{-1}X'y$ se obtiene $\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = \beta[(X'X)^{-1}X'X] + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$ y, tomando el valor esperado de ésta última ecuación, se tiene $E(\hat{\beta}) = E(\beta) + (X'X)^{-1}X'E(u) = \beta$, pues $E(\beta) = \beta$ y $E(u) = 0$, según supuestos indica que $\hat{\beta}$ es un estimador insesgado de β .

Sea $\hat{\beta}^*$ cualquier otro estimador lineal de β , escrito como $\hat{\beta}^* = [(X'X)^{-1}X' + C]y$, donde C es una matriz de constantes.

Al sustituir y de $y = X\beta + u$ en $\hat{\beta}^* = [(X'X)^{-1}X' + C]y$ se obtiene $\hat{\beta}^* = [(X'X)^{-1}X' + C](X\beta + u) = \beta + CX\beta + (X'X)^{-1}X'u + Cu$ tal que si $\hat{\beta}^*$ es un estimador insesgado de β se tiene que $CX = 0$.

Entonces, con $\hat{\beta}^* = \beta + CX\beta + (X'X)^{-1}X'u + Cu$ y $CX = 0$ se escribe $\hat{\beta}^* - \beta = (X'X)^{-1}X'u + Cu$ y, por definición, la matriz de Var - Cov($\hat{\beta}^*$) = $E(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)$

$\hat{\beta}' = E \left[(X'X)^{-1}X'u + Cu \right] \left[(X'X)^{-1}X'u + Cu \right]'$ empleando propiedades de inversión y transposición de matrices, luego de simplificación algebraica se obtiene $\text{Var} - \text{Cov}(\hat{\beta}^*) = \sigma^2(X'X)^{-1} + \sigma^2CC' = \text{Var} - \text{Cov}(\hat{\beta}) + \sigma^2CC'$.

Esto indica que matriz de varianza – covarianza del estimador lineal e insesgado alterno $\hat{\beta}^*$ es igual a matriz varianza – covarianza del estimador MCO, $\hat{\beta}$ más σ^2 veces CC' , que es una matriz semidefinida positiva.

Por tanto, las varianzas de un elemento dado de $\hat{\beta}^*$ deben ser necesariamente iguales o mayores al elemento correspondiente de $\hat{\beta}$, que demuestra que $\hat{\beta}$ es Mejor Estimador Lineal Insesgado (MELI). Por supuesto, si C es una matriz nula (matriz cuyos elementos son todos cero y se denota por 0); es decir $C = 0$, entonces $\hat{\beta}^* = \hat{\beta}$, equivalente a que si encuentra un estimador MELI debe ser el estimador de mínimos cuadrados $\hat{\beta}$.

Con base en lo anterior, dados los supuestos del modelo clásico de regresión lineal, las estimaciones de mínimos cuadrados poseen algunas propiedades ideales u óptimas, contenidas en el **Teorema de Gauss-Markov** y siendo el estimador de MCO $\hat{\beta}_2$ es el mejor estimador lineal insesgado de β_2 o $\hat{\beta}^*$ de $\hat{\beta}$:

1. **Es lineal.** Función lineal de una variable aleatoria, como variable dependiente Y en modelo de regresión.
2. **Es insesgado.** Su valor promedio o esperado, $E(\hat{\beta}_2)$, es igual al valor verdadero, β_2 .

Estimador eficiente. Tiene varianza mínima dentro de la clase de todos los estimadores lineales insesgados. Un estimador insesgado con varianza mínima se conoce como estimador eficiente.

2.4.1. Precisión o errores estándar de estimaciones de Mínimos

Cuadrados Ordinarios

De las ecuaciones $\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$ y $\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \bar{Y} - \hat{\beta}_2 \bar{X}$ es evidente que estimaciones son función de datos muestrales. Sin embargo, como es probable que los datos cambien entre una muestra y otra, los valores estimados cambian ipso facto (expresión latina que significa por el mismo hecho, por el hecho mismo, inmediatamente o en el acto). Entonces, se requiere alguna medida de “confiabilidad” o **precisión** de estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$.

En estadística, la precisión de un valor estimado se mide por su error estándar, entendido como desviación estándar de la distribución muestral del estimador y la distribución muestral de un estimador es tan sólo una probabilidad o distribución de frecuencias del estimador; es decir, una distribución del conjunto de valores del estimador obtenidos de todas las muestras posibles de igual tamaño de una población dada.

Con las distribuciones muestrales se infieren los valores de los parámetros de la población, con base en los valores de los estimadores calculados a partir de una o más muestras. Dados los supuestos gaussianos, se muestra que los errores estándar de estimadores de MCO pueden obtenerse mediante:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{ee}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sum x_i^2}{n \sum x_i^2} \sigma^2$$

$$\text{ee}(\hat{\beta}_1) = \sqrt{\frac{\sum x_i^2}{n \sum x_i^2}} \sigma$$

Donde:

σ^2 : Varianza constante u homocedástica de u_i

ee: Error Estándar

Todas las cantidades que entran en anteriores ecuaciones, excepto σ^2 , pueden estimarse a partir de datos. Incluso, la misma σ^2 se estima mediante:

$$\hat{\sigma}^2_{\text{(Estimador de MCO de verdadera } \sigma^2 \text{ desconocida)}} = \frac{\sum \hat{u}_i^2 \text{ (Suma de valores residuales, errores al cuadrado o cuadrados residuales)}}{n - 2 \text{ (Número de grados de libertad)}}$$

2.4.2. Grados de libertad

Los **grados de libertad** se entiende como número total de observaciones en la muestra (n) menos el número de restricciones (lineales) independientes o de restricciones que se les impusieron o, como, la cantidad de observaciones independientes de un total de n observaciones y su regla general es $gl = n - \text{número de parámetros estimados}$. Error Estándar de estimación o error estándar de regresión (ee).

No es más que la desviación estándar de valores Y alrededor de la línea de regresión estimada, que suele servir como medida para resumir la “bondad del ajuste” de dicha línea. Entonces, dado X_i , σ^2 representa la varianza condicional de u_i y Y_i .

Por lo tanto, el error estándar de estimación también se denomina desviación estándar condicional de u_i y Y_i . Asimismo, es común, σ_Y^2 y σ_i representan la varianza incondicional y desviación estándar incondicional de Y , respectivamente.

Sin embargo, es importante mencionar las siguientes características de varianzas, así como de errores de estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$:

1. Varianza de $\hat{\beta}_2$ es directamente proporcional a $\sigma^2_{\text{(Varianza constante u homocedástica de } u_i)}$, pero inversamente proporcional a $\sum x_i^2$.

Indica que, dada σ^2 , entre más grande sea la variación en valores X menor será la varianza $\hat{\beta}_2$ y, por tanto, mayor será la precisión con que estimar β_2 y, dado $\sum x_i^2$, entre mayor sea la varianza σ^2 mayor será la de β_2 . Tal que, a medida que aumenta el tamaño n muestral, también lo hace el número de términos en la suma, $\sum x_i^2$. Es decir, a medida que aumenta n es mayor la precisión para estimar β_2 .

2. La varianza de $\hat{\beta}_1$ es directamente proporcional a σ^2 y a $\sum X_i^2$, pero inversamente proporcional a $\sum x_i^2$ y al tamaño n muestral.

3. $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores, no sólo variarán de una muestra a otra, sino que también, en una muestra dada, es probable que dependan entre sí. Tal que, esta dependencia se mide por covarianza entre ellos. Esto se demuestra con:

$$\begin{aligned} \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) &= E\{[\hat{\beta}_1 - E(\hat{\beta}_1)][\hat{\beta}_2 - E(\hat{\beta}_2)]\} = E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) = -\bar{X}E(\hat{\beta}_2 - \beta_2)^2 \\ &= -\bar{X}\text{Var}(\hat{\beta}_2) = -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) \end{aligned}$$

Donde $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$ y $E(\hat{\beta}_1) = \bar{Y} - \beta_2\bar{X}$ que es igual a $\hat{\beta}_1 - E(\hat{\beta}_1) = -\bar{X}(\hat{\beta}_2 - \beta_2)$.

La $\text{Var}(\hat{\beta}_2) = E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2 = E(\hat{\beta}_2 - \beta_2)^2$, pues $E(\hat{\beta}_2) = \beta_2$, implica que $E(\sum k_i u_i)^2$ con ecuación $\hat{\beta}_2 = \sum k_i (\beta_1 + \beta_2 X_i + u_i) = \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i = \beta_2 + \sum k_i u_i$ implica que sea $= E(k_1^2 u_1^2 + k_2^2 u_2^2 + k_3^2 u_3^2 + k_4^2 u_4^2 + \dots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + \dots + 2k_{n-1} k_n u_{n-1} u_n)$.

Por los supuestos $E(u_i^2) = \sigma^2$ para cada i y $E(u_i u_j) = 0, i \neq j \therefore \text{Var}(\hat{\beta}_2) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum x_i^2}$ con definición de k_i^2 .

La varianza de $\hat{\beta}_1$ se obtiene con el mismo razonamiento, pues una vez obtenidas las varianzas de $\hat{\beta}_1$ y $\hat{\beta}_2$ se obtiene también errores estándar correspondientes al tomar raíces cuadradas positivas.

Como $\text{Var}(\hat{\beta}_2)$ es siempre positiva, al igual que varianza de cualquier variable, la naturaleza de covarianza entre $\hat{\beta}_1$ y $\hat{\beta}_2$ depende del signo de \bar{X} , pues si \bar{X} es positiva, la covarianza será negativa. Así, el coeficiente de pendiente de β_2 está sobreestimada (pendiente muy pronunciada), el coeficiente del intercepto β_1 estará subestimado (intercepto será muy pequeño). La mayor utilidad del estudio de covarianzas entre coeficientes estimados de regresión será en Multicolinealidad.

2.4.3. Teorema de Gauss-Markov bajo las hipótesis H

Sea $\psi = a\beta_1 + b\beta_2$ donde $a, b \in \mathbb{R}$. Si $\hat{\beta}_1, \hat{\beta}_2$ son estimadores de MC , entonces $\hat{\psi} = a\hat{\beta}_1 + b\hat{\beta}_2$ es estimador lineal e insesgado de ψ , pero aún más es de menor varianza entre todos los estimadores lineales e insesgados. Entonces, $\hat{\psi}$ es mejor estimador lineal insesgado (*BLUE*) o estimador lineal insesgado eficiente.

2.4.4. Estimación σ^2

Según (Castro, 2008), medias $\mu_i = E(y_i|x_i) = \beta_1 + \beta_2 x_i$ son combinaciones lineales de parámetros β_1, β_2 tal que $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ es estimador lineal insesgado de varianza mínima de μ_i , por la fórmula $\text{Var}(a\hat{\beta}_1 + b\hat{\beta}_2) = a^2\text{Var}(\hat{\beta}_1) + 2ab\text{CovVar}(\hat{\beta}_1, \hat{\beta}_2) + b^2\text{Var}(\hat{\beta}_2) = \sigma^2 \left[a^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) - 2ab \left(\frac{\bar{x}}{S_{xx}} \right) + \frac{b^2}{S_{xx}} \right] = \sigma^2 \left[\frac{a^2}{n} + \frac{1}{S_{xx}} (a^2 \bar{x}^2 - 2ab\bar{x} + b^2) \right] = \sigma^2 \left[\frac{a^2}{n} + \frac{(b-a\bar{x})^2}{S_{xx}} \right]$ su varianza es:

$$\text{Var}(\hat{y}_i) = \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right]$$

Interpretación. La igualdad $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ se interpreta en sentido:

- Si X se incrementa en una unidad, entonces Y se incrementa o decreta en promedio $\hat{\beta}_2$ unidades.
- $\hat{\beta}_1$ es valor medio de Y cuando X es nula. Tiene sentido sólo si el rango de valores X incluye valor cero o si tiene sentido que X sea nula, a condición que la linealidad se mantenga en rango que incluye el valor cero.

Definición. La recta $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$ $x \in \mathbb{R}$ se llama recta de regresión. A \hat{y} se llama como predicción de Y dado $X = x$. Si bien la recta de regresión está definida en todo el eje real, su validez se limita al rango de observación. No debe olvidarse que una recta sólo es una aproximación de una curva más compleja que podría ser observada en un rango más amplio de X.

Varianzas y errores estándar de estimadores de MCO se obtienen después de los coeficientes de regresión parcial y errores estándar de estimadores, pues $\text{Var}(\hat{\beta}_2) = E[(\hat{\beta}_2 - E(\hat{\beta}_2))]^2 = E(\hat{\beta}_2 - \beta_2)^2$, pues $E(\hat{\beta}_2) = \beta_2$,

tal que $= E(\sum k_i u_i)^2$ 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 $= E(k_1^2 u_1^2 + k_2^2 u_2^2 + k_3^2 u_3^2 + k_4^2 u_4^2 + \dots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + 2k_3 k_4 u_3 u_4 + 2k_5 k_6 u_5 u_6 + 2k_7 k_8 u_7 u_8 + \dots + 2k_{n-1} k_n u_{n-1} u_n)$ aunque por supuestos de

$E(u_i^2) = \sigma^2 \forall i$ y $E(u_i u_j) = 0, i \neq j$ se deduce que $\text{Var}(\hat{\beta}_2) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum x_i^2}$, pero con definición de k_i^2 se concluye que $\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$, 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:

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

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

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

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

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

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

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

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:

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

Tal que:

$$\begin{aligned} \text{Var} - \text{Cov}(\hat{\beta}) &= E \left\{ [\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]' \right\} \\ &= \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_k) \end{bmatrix} \end{aligned}$$

Esto se demuestra por:

$$\text{Var} - \text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \Rightarrow \begin{matrix} \hat{\beta} \\ (k * 1) \end{matrix} = \begin{matrix} (X'X)^{-1} \\ (k * k) \end{matrix} * \begin{matrix} X' & y \\ (k * n) & (n * 1) \end{matrix}$$

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. $a' = [a_1 \ a_2 \ \dots \ a_n]$ 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, \dots, 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 $x'A_x = [x_1 \ x_2 \ \dots \ x_n] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

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, $\text{Var} - \text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1} \Rightarrow \hat{\beta} = \begin{matrix} (X'X)^{-1} & X' & y \\ (k * 1) & (k * k) & (k * n)(n * 1) \end{matrix}$,
 $\hat{u}'\hat{u} = (y - X\hat{\beta})' * (y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$, 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 $\hat{\beta} = \begin{matrix} (X'X)^{-1} & X' & y \\ (k * 1) & (k * k) & (k * n)(n * 1) \end{matrix}$, 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, $\text{Var} - \text{Cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left\{\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'\right\} = E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$ aprovechando que $(AB)' = A'B'$ y las X no son estocásticas, al tomar el valor esperado de ecuación anterior implica:

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

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

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 $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2X_{2i} - \hat{\beta}_3X_{3i} = y_i - \hat{\beta}_2x_{2i} - \hat{\beta}_3x_{3i}$ (letras minúsculas indican desviaciones respecto de valores de media) entonces

$$\sum \hat{u}_i^2 = \sum(\hat{u}_i\hat{u}_i) = \sum \hat{u}_i(y_i - \hat{\beta}_2x_{2i} - \hat{\beta}_3x_{3i}) = \sum \hat{u}_iy_i = \sum y_i\hat{u}_i = \sum y_i(y_i - \hat{\beta}_2x_{2i} - \hat{\beta}_3x_{3i}) = \sum y_i^2 - \hat{\beta}_2 \sum y_ix_{2i} - \hat{\beta}_3 \sum y_ix_{3i} \text{ tal que } \sum \hat{u}_ix_{2i} = \sum \hat{u}_ix_{3i} = 0:$$

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

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 $\sum \hat{u}_i^2 = \sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2X_{2i} - \hat{\beta}_3X_{3i} - \hat{\beta}_4X_{4i} \dots - \hat{\beta}_kX_{ki})^2$ parcialmente respecto a $\beta_1, \beta_2, \beta_3, \dots, \beta_k$, se obtiene:

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

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

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

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

.....

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

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

$$n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \hat{\beta}_4 \sum X_{4i} + \dots + \hat{\beta}_k \sum X_{ki} = \sum Y_i$$

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

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

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

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

$$\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \dots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \dots & \sum X_{3i}X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \dots & \sum X_{ki}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \dots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{32} & \dots & X_{3n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k1} & \dots & X_{k1} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{bmatrix}$$

(X'X) (X') (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:

$$I\hat{\beta} = (X'X)^{-1}X'y \Rightarrow \hat{\beta} = \begin{matrix} (k * 1) & & & & \\ & (k * k) & & & \\ & & (k * n) & & \\ & & & (n * 1) & \end{matrix} \begin{matrix} X' & y \\ & \end{matrix}$$

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.

Esta curva puede ser un modelo matemático-económico, 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):

➤ **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 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$:

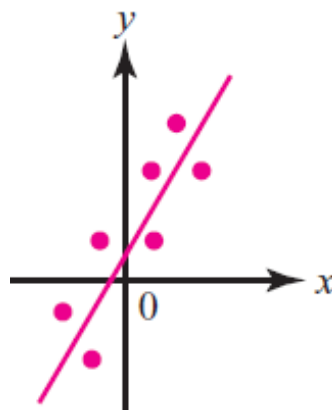


Figura. 2. 4. Representación de la recta o lineal de la forma $y = a + bx$ o $y = b + mx$

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”, $y = y_1$:

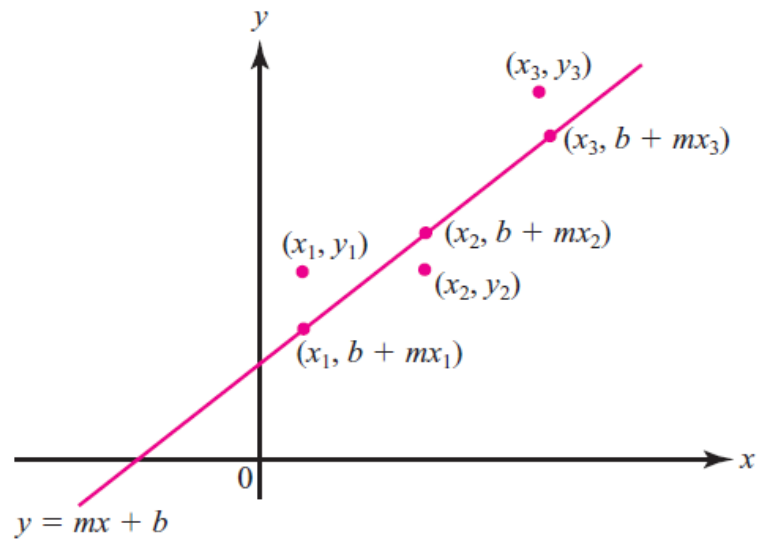


Figura. 2. 5. Representación de la recta o lineal, suponiendo que las variables x e y están relacionadas por la fórmula $y = b + mx$

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 encuentran números m y b tales que la suma $[y_1 - (b + mx_1)]^2 + [y_2 - (b + mx_2)]^2 + [y_3 - (b + mx_3)]^2 + [y_4 - (b + mx_4)]^2 + \dots + [y_n - (b + mx_n)]^2$ sea mínima. Para estos valores de m y b , la recta $y = a + bx$ o $y = b + mx$ se llama “**Aproximación por recta de Mínimos Cuadrados a datos $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$** ”.

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:

$$\begin{array}{ll} y_1 = a + bx_1 & y_1 = b + mx_1 \\ y_2 = a + bx_2 & y_2 = b + mx_2 \\ y_3 = a + bx_3 & y_3 = b + mx_3 \\ y_4 = a + bx_4 & y_4 = b + mx_4 \\ \dots & \dots \\ y_n = a + bx_n & y_n = b + mx_n \end{array} \quad \text{ó}$$

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

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

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

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

Encontrar el vector \vec{u} que minimiza no es tan difícil como parece. Como \mathbf{A} es una matriz de $n_{(\text{Número de filas})} * 2_{(\text{Número de columnas})}$ y \mathbf{u} una matriz de $2_{(\text{Número de filas})} * 1_{(\text{Número de columnas})}$, \mathbf{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.

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

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

Su **demostración** implica que sea $h = \text{proy}_H(\text{Imagen de A})v$ y $p = v - h$ y por **Definición Proyección**, “sea H un subespacio de \mathbb{R}^n con base ortonormal $\{u_1, u_2, u_3, u_4, \dots, u_k\}$. Si $v \in \mathbb{R}^n$, entonces la **proyección ortogonal** de v sobre H, denotada por $\text{proy}_H(\text{Imagen de A})v$ está dada por $\text{proy}_H(\text{Imagen de A})v = (v * u_1)u_1 + (v * u_2)u_2 + (v * u_3)u_3 + (v * u_4)u_4 + \dots + (v * u_k)u_k$, siendo que $\text{proy}_H(\text{Imagen de A})v \in H$ ”, 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: $h = (v * u_1)u_1 + (v * u_2)u_2 + (v * u_3)u_3 + (v * u_4)u_4 + \dots + (v * u_k)u_k$ y sea x un vector H tal que existen constantes $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_k$ tales que $x = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \dots + \alpha_k u_k$ entonces $p * x = (v - h) * x = [v - (v * u_1)u_1 - (v * u_2)u_2 + (v * u_3)u_3 + (v * u_4)u_4 + \dots + (v * u_k)u_k] * [\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \dots + \alpha_k u_k]$ aunque como $u_i * u_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$ es sencillo verificar que el producto escalar $p * x = (v - h) * x = [v - (v * u_1)u_1 - (v * u_2)u_2 + (v * u_3)u_3 + (v * u_4)u_4 + \dots + (v * u_k)u_k] * [\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \dots + \alpha_k u_k]$ está dado por $p * x = \sum_{i=1}^k \alpha_i (v * u_i) - \sum_{i=1}^k \alpha_i (v * u_i) = 0$, así $p * x = 0 \forall x \in H$ indicando que $p \in H^\perp$.

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

Entonces, $\{u_{k+1}, \dots, u_n\}$ es una base para H^\perp , tal que se puede decir que $v - \text{proy}_H(\text{Imagen de } A)v \in H^\perp$ se escribe $v - h = (v - \text{proy}_H(\text{Imagen de } A)v) + (\text{proy}_H(\text{Imagen de } A)v - h)$ siendo que el primer término de la derecha como el segundo está en H^\perp tal que $(v - \text{proy}_H(\text{Imagen de } A)v) \cdot (\text{proy}_H(\text{Imagen de } A)v - h) = 0$ 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 $v \in \mathbb{R}^n$ tal que $v = (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + (v \cdot u_3)u_3 + (v \cdot u_4)u_4 + \dots + (v \cdot u_k)u_k$ o $v = \text{proy}_{\mathbb{R}^n} v$, siendo su demostración que considera B como una base, puede escribir v de manera única como $v = c_1 u_1 + c_2 u_2 + c_3 u_3 + c_4 u_4 + \dots + c_n u_n$ por lo que $v \cdot u_i = c_1(u_1 \cdot u_i) + c_2(u_2 \cdot u_i) + c_3(u_3 \cdot u_i) + c_4(u_4 \cdot u_i) + \dots + c_i(u_i \cdot u_i) + \dots + c_n(u_n \cdot u_i) = c_i$, 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 $v = (v \cdot u_1)u_1 + (v \cdot u_2)u_2 + (v \cdot u_3)u_3 + (v \cdot u_4)u_4 + \dots + (v \cdot u_k)u_k + (v \cdot u_{k+1})u_{k+1} + \dots + (v \cdot u_n)u_n = \text{proy}_H v + \text{proy}_{H^\perp} v$ aunque para probar la unicidad, suponga que

$v = h_1 - p_1 = h_2 - p_2$ 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 $h_1 - h_2 \in H \cap H^\perp = \{0\}$, así que $h_1 - h_2 = 0$ y $p_1 - p_2 = 0$ completando la prueba”, por lo que v , tal que $|y - Au|$ es un mínimo cuadrado pues $Au = \text{proy}_H(\text{Imagen de } A)y$ 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 :

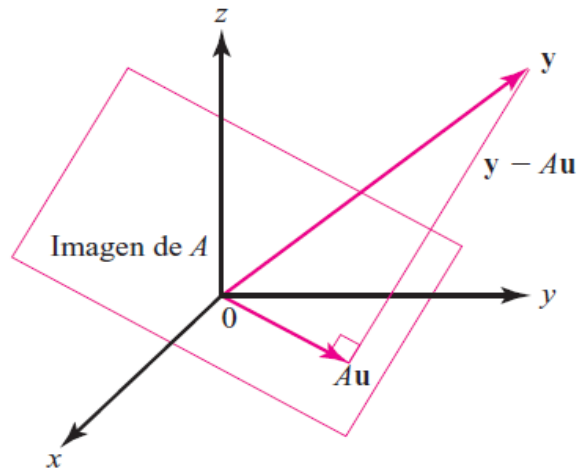


Figura. 2. 6. Gráfica en \mathbb{R}^3 de A , que es un plano o recta que pasa por el origen

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 $Au \perp (y - Au)$. Usando la definición de producto escalar en \mathbb{R}^n , se encuentra que $Au \perp (y - Au)$ se vuelve $Au * (y - Au) = 0$, $(Au)^T * (y - Au) = 0$ debido a que $a * b = a^T b$, $(A^T u^T) * (y - Au) = 0$ pues por **Teorema ii)** que supone en i) $(A^T)^T$, ii) $(AB)^T = A^T B^T$, iii) Si A y B son de $n * m$, entonces $(AB)^T = A^T + B^T$ y iv) Si A es invertible, entonces A^T es invertible y $(A^T)^{-1} = (A^{-1})^T$ o $a^T(A^T y - A^T * Au) = 0$ se cumple $\forall u \in \mathbb{R}^2 \Leftrightarrow A^T y - A^T * Au = 0$, pero al despejar \vec{u} de $A^T y - A^T * Au = 0$ 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:

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

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

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

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

➤ **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:

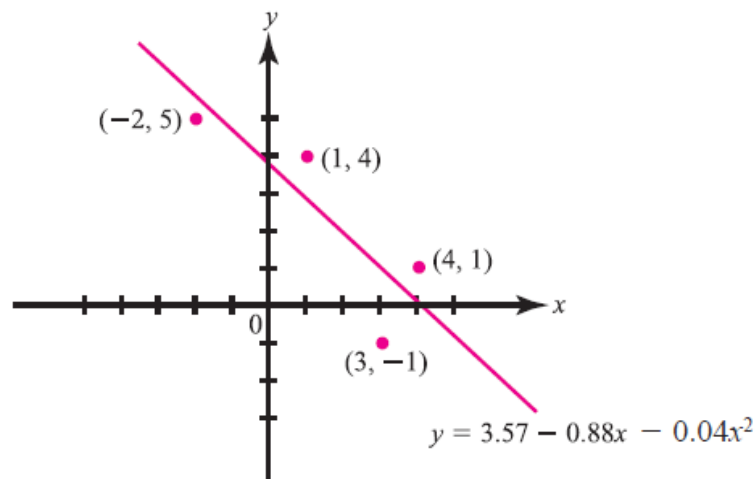


Figura. 2. 7. Ajuste de una curva cuadrática a n datos de forma $y = a + bx + cx^2$ o $y = b + mx + cx^2$

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

$$y_2 = a + bx_1 + cx_2^2 \quad \text{ó} \quad y_2 = b + mx + cx_2^2$$

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

$$\begin{array}{ll}
 y_4 = a + bx_1 + cx_4^2 & y_4 = b + mx + cx_4^2 \\
 \dots & \dots \\
 y_n = a + bx_n + cx_n^2 & y_n = b + mx + cx_n^2
 \end{array}$$

Aunque, este sistema se puede escribir como $\mathbf{y} = \mathbf{A}\mathbf{u}$ con:

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

Al igual que antes, si todos los datos no se encuentran sobre la misma parábola implica que $\mathbf{y} - \mathbf{A}\mathbf{u} \neq \mathbf{0}$ para cualquier vector \vec{u} y, de nuevo, el problema es “encontrar un vector \mathbf{u} en \mathbb{R}^2 tal que $|\mathbf{y} - \mathbf{A}\mathbf{u}|$ es mínima”.

Usando un razonamiento similar a éste, se puede demostrar que cuando menos tres de las x_i son diferentes; entonces, $\mathbf{A}^T\mathbf{A}$ es invertible y el vector que minimiza al vector \vec{u} está dado por $\vec{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. La siguiente imagen muestra la figura de un modelo matemático-econométrico cuadrático:

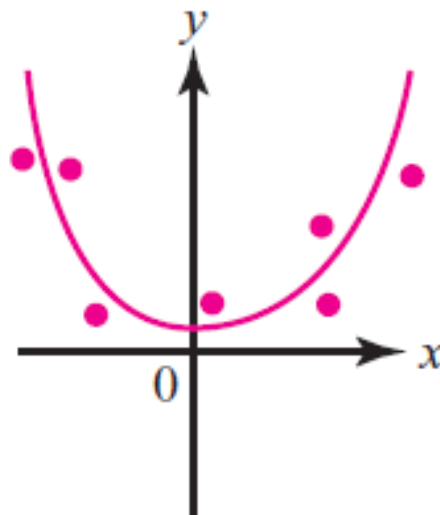


Figura. 2. 8. Modelo matemático-econométrico cuadrático

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 $A^T A$ es invertible de 2×2 :

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

Es importante mencionar que si $x_1 = x_2 = x_3 = x_4 = \dots = x_n$, 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:

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

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

$$A^T_{(\text{Matriz } A \text{ transpuesta})} A = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & x_4 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

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

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

Sea $u = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}$ y $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \dots \\ x_n \end{pmatrix}$. 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 $n \sum_{i=1}^n x_i^2 = \left(\sum_{i=1}^n x_i \right)^2$ puede establecerse como $|u||x|^2 = |u * x|^2$, sacando raíz cuadrada se puede obtener $|u * x| = |u||x|$.

2.5. Desigualdad de Cauchy-Schwarz en \mathbb{R}^n

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

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

$$\begin{aligned} 0 &\leq \left| \frac{u}{|u|} - \frac{v}{|v|} \right|^2 = \left(\frac{u}{|u|} - \frac{v}{|v|} \right) * \left(\frac{u}{|u|} - \frac{v}{|v|} \right) = \frac{u * u}{|u|^2} - \frac{2u * v}{|u||v|} + \frac{v * v}{|v|^2} \\ &= \frac{|u|^2}{|u|^2} - \frac{2u * v}{|u||v|} + \frac{|v|^2}{|v|^2} = 1 - \frac{2u * v}{|u||v|} + 1 = 2 - \frac{2u * v}{|u||v|} \Rightarrow 2 \\ &\leq \frac{2u * v}{|u||v|}, \frac{u * v}{|u||v|} \leq 1 \text{ y } u * v \leq |u||v| \end{aligned}$$

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. $|u * v| = |u||v| \Leftrightarrow u = 0$ 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:** $\frac{u * v}{|u||v|} = 1$.

$$\begin{aligned} \left| \frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right|^2 &= \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right) * \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right) = \frac{\mathbf{u} * \mathbf{u}}{|\mathbf{u}|^2} - \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + \frac{\mathbf{v} * \mathbf{v}}{|\mathbf{v}|^2} = \frac{|\mathbf{u}|^2}{|\mathbf{u}|^2} - \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + \frac{|\mathbf{v}|^2}{|\mathbf{v}|^2} \\ &= 1 - \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + 1 = 2 - \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = 2 - 2(1) = 0 \Rightarrow \frac{\mathbf{u}}{|\mathbf{u}|} = \frac{\mathbf{v}}{|\mathbf{v}|} \text{ ó } \mathbf{u} = \frac{\mathbf{u}}{|\mathbf{u}|} * \mathbf{v} \\ &= \lambda \mathbf{v} \end{aligned}$$

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

$$\begin{aligned} \left| \frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right|^2 &= \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right) * \left(\frac{\mathbf{u}}{|\mathbf{u}|} - \frac{\mathbf{v}}{|\mathbf{v}|} \right) = \frac{\mathbf{u} * \mathbf{u}}{|\mathbf{u}|^2} + \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + \frac{\mathbf{v} * \mathbf{v}}{|\mathbf{v}|^2} = \frac{|\mathbf{u}|^2}{|\mathbf{u}|^2} + \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + \frac{|\mathbf{v}|^2}{|\mathbf{v}|^2} \\ &= 1 + \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} + 1 = 2 + \frac{2\mathbf{u} * \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = 2 - 2(1) = 0 \Rightarrow \frac{\mathbf{u}}{|\mathbf{u}|} = -\frac{\mathbf{v}}{|\mathbf{v}|} \text{ ó } \mathbf{u} \\ &= -\frac{\mathbf{u}}{|\mathbf{u}|} * \mathbf{v} = \lambda \mathbf{v} \end{aligned}$$

➤ **Cúbico.**

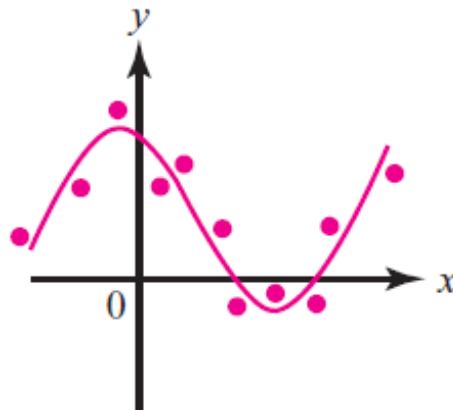


Figura. 2. 9. Representación cúbica de la recta o lineal de la forma $y = a + bx$ o $y = b + mx$

El objetivo es encontrar la curva del tipo específico que se ajuste “mejor” a los datos dados. Complementariamente, un sistema de ecuaciones se escribe $\mathbf{Ax} = \mathbf{b}$ tal que \mathbf{A} es una matriz de $m \times n$, $\mathbf{x} \in \mathbb{R}^n$ y $\mathbf{b} \in \mathbb{R}^m$. Asimismo, la función $\mathbf{Ax} = \mathbf{b}$ tiene las propiedades que caracteriza las transformaciones lineales, $\mathbf{A}(\alpha \mathbf{x}) = \alpha \mathbf{Ax}$ si α es una escalar y $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}$.

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 $\mathbf{v} \in V$ 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 $T(\alpha \mathbf{v}) = \alpha T\mathbf{v}$.

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:

- **Transformación lineal de \mathbb{R}^2 en \mathbb{R}^3 .**

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

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

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

$$T\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} \approx T\left[\alpha \begin{pmatrix} X \\ Y \end{pmatrix}\right] = T\begin{pmatrix} \alpha X \\ \alpha Y \end{pmatrix} = \begin{pmatrix} \alpha X + \alpha Y \\ \alpha X - \alpha Y \\ 3\alpha Y \end{pmatrix} = \alpha \begin{pmatrix} X + Y \\ X - Y \\ 3Y \end{pmatrix} = \alpha T\begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \therefore \mathbf{T \text{ es una}}$$

transformación lineal.

- **Transformación cero.**

Sean V y W espacios vectoriales y definida $T: V \rightarrow W$ 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**.

- **Transformación identidad.**

Sea V un espacio vectorial y definida $I: V \rightarrow V$ 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**.

- **Transformación de reflexión.**

Sea $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 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)):

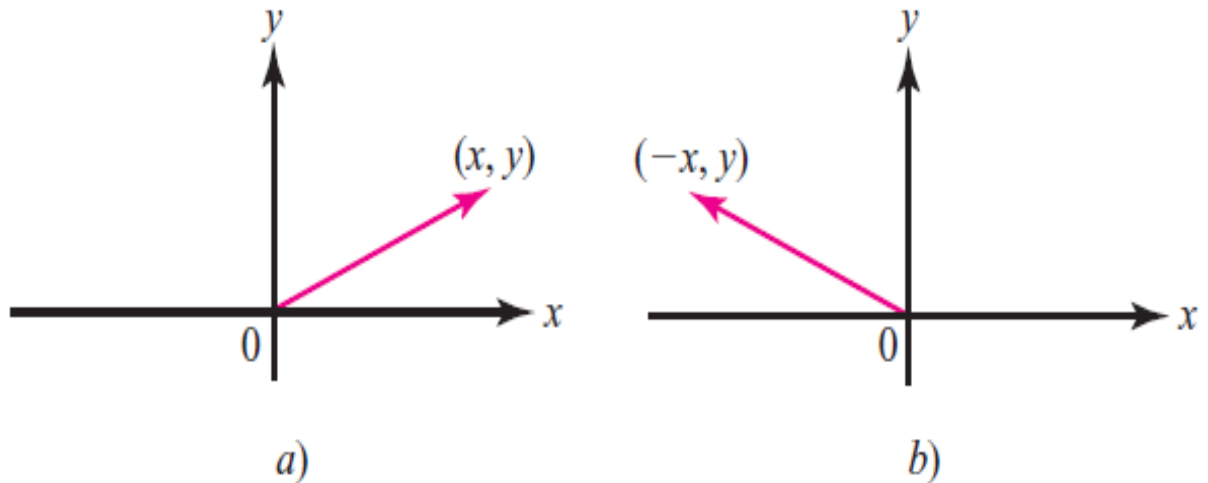


Figura. 2. 10. Transformación de reflexión de $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida por $T \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} -X \\ Y \end{pmatrix}$

- **Transformación lineal de $\mathbb{R}^n \rightarrow \mathbb{R}^m$ dada por multiplicación de matriz $m \times n$.**

Sea A una matriz de $m \times n$, definida $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 \mathbf{y} están en \mathbb{R}^n , se observa que T es una transformación lineal. Entonces, toda matriz A de $m \times n$ 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.

- **Transformación de rotación.**

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

Llame a este nuevo vector rotado $\vec{v}' = \begin{pmatrix} X' \\ Y' \end{pmatrix}$ (si r denota longitud de \vec{v} , que no cambia por rotación tal que $x = r \cos(\alpha) \Rightarrow x' = r \cos(\theta + \alpha)$ y $y = r \sin(\alpha) \Rightarrow y' = r \sin(\theta + \alpha)$, pues la definición estándar de $\cos(\theta)$ y $\sin(\theta)$ 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 $x = r \cos(\varphi)$ y $y = r \sin(\varphi)$, donde φ (Φ o Φ) es ángulo que forma $\vec{v}(X, Y)$ con lado positivo del eje X . Además, (X', Y') se obtiene rotando (X, Y) un ángulo θ :

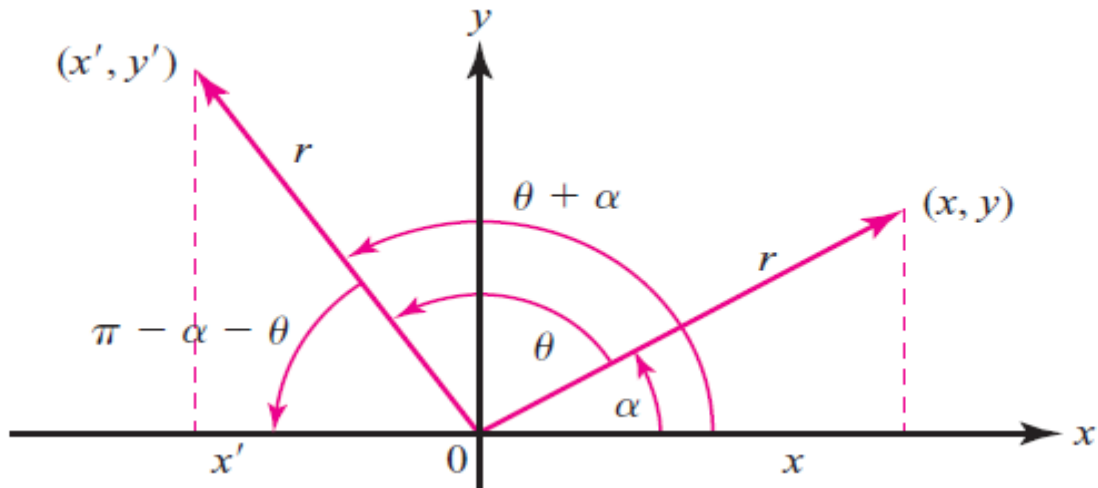


Figura. 2. 11. Transformación de rotación de $\vec{v} = \begin{pmatrix} X \\ Y \end{pmatrix}$ en $\Pi_{(XY)}$ se rota un θ , medido en grados ($^\circ$) y radianes (rad) en sentido contrario a manecillas del reloj

Pero $r \cos(\theta + \alpha) = r \cos(\theta)\cos(\alpha) - r \sin(\theta)\sin(\alpha) \Rightarrow x' = x \cos(\theta) - y \sin(\theta)$. Paralelamente, $r \sin(\theta + \alpha) = r \sin(\theta)\cos(\alpha) + r \cos(\theta)\sin(\alpha) \Rightarrow y' = x \sin(\theta) + y \cos(\theta)$. Sea $\mathbf{A}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \Rightarrow x' = x \cos(\theta) - y \sin(\theta)$ y $y' = x \sin(\theta) + y \cos(\theta)$ denota que $\mathbf{A}_\theta \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X' \\ Y' \end{pmatrix}$.

La transformación lineal $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida por $T\mathbf{v} = \mathbf{A}_\theta\mathbf{v}$, donde \mathbf{A}_θ está dado por $\mathbf{A}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ y recibe el nombre de **transformación de rotación**.

- **Transformación de proyección ortogonal** (perpendicular: \perp).

Sea H un subespacio de \mathbb{R}^n tal que la **transformación de proyección ortogonal** está dado por $P: V \rightarrow H$ se define por $P\mathbf{v} = \text{Proy}_H\mathbf{v}$. Sea la sucesión $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \dots, \mathbf{u}_k\}$ 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 $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \dots, \mathbf{u}_k\}$ tal que si $\mathbf{v} \in \mathbb{R}^n \Rightarrow$

proyección ortogonal de \mathbf{v} sobre H , denotada por $\text{Proy}_H \mathbf{v}$ está dada por $\text{Proy}_H \mathbf{v} = (\mathbf{v} * \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v} * \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v} * \mathbf{u}_3)\mathbf{u}_3 + (\mathbf{v} * \mathbf{u}_4)\mathbf{u}_4 + \dots + (\mathbf{v} * \mathbf{u}_k)\mathbf{u}_k$ tal que $\text{Proy}_H \mathbf{v} \in H$, se tiene que $P\mathbf{v} = (\mathbf{v}_1 * \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v}_1 * \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v}_1 * \mathbf{u}_3)\mathbf{u}_3 + (\mathbf{v}_1 * \mathbf{u}_4)\mathbf{u}_4 + \dots + (\mathbf{v}_1 * \mathbf{u}_k)\mathbf{u}_k$, 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.

- 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 $\vec{\nu}$ en espacio en $\Pi_{(xz)}$:

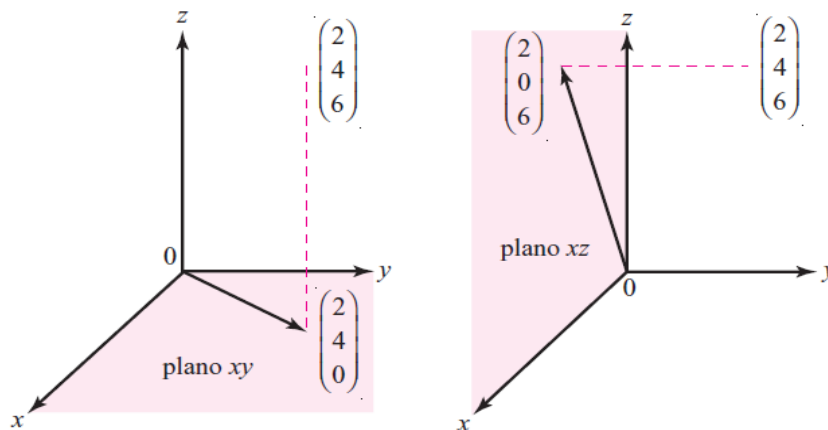


Figura. 2. 12. Representación de dos operadores de proyección definida por $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ por

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \Rightarrow T$$

- 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.

- Operador integral.

Sea $J: C[0, 1] \rightarrow \mathbb{R}$ 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 $\int_0^1 \alpha f(x) dx = \alpha \int_0^1 f(x) dx \Rightarrow J$ es un operador lineal.

Por ejemplo: $J(X^3) = \frac{1}{4} \Rightarrow J$ se denomina **operador integral**.

- **Operador diferencial.**

Suponga que $D: C^1[0, 1] \rightarrow C[0, 1]$ 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**.

- **Transformación no lineal.**

Suponga que $T: C[0, 1] \rightarrow \mathbb{R}$ 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.

Sin embargo, no toda transformación que parece lineal lo es. Por ejemplo: $T: \mathbb{R} \rightarrow \mathbb{R}$ por $Tx = 2x + 3 \Rightarrow$ la gráfica de $\{(x, Tx)\}: x \in \mathbb{R}$ es línea recta 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.

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 $Y_i^{(\lambda)} = \begin{cases} K_1(Y_i^\lambda - 1) & \text{si } \lambda \neq 0, \\ K_2 \text{ Ln}(Y_i) & \text{si } \lambda = 0 \end{cases}$ tal que K_2 es media geométrica de valores $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$; en consecuencia, $K_2 = \left(\prod_{i=1}^n (\text{Producto desde 1 hasta } n) Y_i\right)^{\frac{1}{n}} = (Y_1 * Y_2 * Y_3 * Y_4 * \dots * Y_n)^{\frac{1}{n}}$, 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 .

2.6. Regresión sin término constante

Si β_1 no está en modelo; es decir, si $y_i = \beta_2 x_i + u_i$, se minimiza:

$$SRC_{(\text{Suma de Residuos al Cuadrado})}(b_1) = \sum_{i=1}^n (y_i - b_2 x_i)^2$$

Derivando respecto a b_2 se tiene la única ecuación normal:

$$\sum_{i=1}^n (y_i - \hat{\beta}_2) x_i = 0$$

Tal que su solución es:

$$\hat{\beta}_2 = \left(\frac{\sum x_i y_i}{\sum x_i^2} \right)$$

La recta de regresión $\hat{y} = \hat{\beta}_2 x$ pasa por el origen de coordenadas. La ecuación normal $\sum_{i=1}^n (y_i - \hat{\beta}_2) x_i = 0$ permite afirmar:

$$\sum_{i=1}^n x_i \hat{u}_i = 0$$

Aunque, no se puede afirmar:

$$\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n (y_i - \hat{\beta}_2) x_i = 0$$

Esta se cumplirá sólo si $\hat{\beta}_2 = \left(\frac{\bar{y}}{\bar{x}} \right)$. Al igual que en regresión con término constante $\hat{\beta}_2$ es estimador insesgado de β_2 , pero su varianza es diferente:

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

Si β_1 no está en modelo i $\bar{y} \neq 0$ la igualdad $STC = SEC + SRC$ puede ser falsa pues $\sum_{i=1}^n \hat{u}_i$ no es necesariamente equivalente a cero. En vez de esta igualdad se tiene el resultado con validez general.

Teorema. En toda regresión lineal:

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + SRC$$

Demostración:

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n [(y_i - \hat{y}_i) + \hat{y}_i]^2 = \sum_{i=1}^n \hat{u}_i^2 + 2 \sum_{i=1}^n \hat{u}_i \hat{y}_i + \sum_{i=1}^n \hat{y}_i^2 = \sum_{i=1}^n \hat{u}_i^2 + \sum_{i=1}^n \hat{y}_i^2$$

La última desigualdad es consecuencia de lema $\sum_{i=1}^n \hat{u}_i \hat{y}_i = 0$.

Teorema. En una regresión lineal simple sin término constante y si $\bar{y} \neq 0$, el estimador insesgado de varianza es:

$$S_R^2 = \left(\frac{SRC}{n-1} \right)$$

Es importante mencionar que es dividido por $n-1$ y no por $n-2$.

2.1. Hipótesis de normalidad (inferencia)

Cuando se conoce la distribución de probabilidad de observaciones y_i se puede calcular estimadores de máxima verosimilitud (MV). Para el caso de regresión lineal y suponiendo que errores son independientes normalmente distribuidos se puede demostrar que estimadores de MC son iguales a estimadores de MV, salvo que σ^2 se estima con $\frac{SRC}{n}$.

2.6.1. Distribución de estimadores

Teorema. Bajo hipótesis N:

$$\hat{\beta}_1 \rightarrow N\left(\beta_1, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

$$\hat{\beta}_2 \rightarrow N\left(\beta_2, \frac{\sigma^2}{S_{xx}}\right)$$

Demostración:

Se hará para $\hat{\beta}_2$:

$$\hat{\beta}_2 = \sum_{i=1}^n a_i y_i \text{ con } a_i = \frac{x_i - \bar{x}}{S_{xx}}$$

Por hipótesis $\forall i y_i \rightarrow N(\beta_1 + \beta_2 x_i, \sigma^2)$ talque son independientes. Puesto que la combinación lineal de variables aleatorias de ley normal es normal. Se concluye:

$$\hat{\beta}_2 \rightarrow N\left(E(\hat{\beta}_2), \text{Var}(\hat{\beta}_2)\right)$$

Si el término constante no está en modelo:

$$\hat{\beta}_2 \rightarrow N\left(\hat{\beta}_2, \frac{\sigma^2}{\sum x_i^2}\right)$$

Corolario. Bajo las hipótesis N, si β_1 está en modelo si datos han sido centrados:

$$\frac{1}{\sigma^2} SRC \rightarrow \chi_{n-2}^2$$

Si β_1 no está en modelo y datos no están centrados, cambian únicamente los grados de libertad, pues son $n - 1$.

Teorema. Bajo las hipótesis N, para $\forall i$; $\hat{\beta}_1$ y $\hat{\beta}_2$ son independientes de \hat{u}_i .

Corolario. Bajo las hipótesis N, $\hat{\beta}_1$ y $\hat{\beta}_2$ son independientes de SRC.

Demostración:

La SRC es función únicamente de \hat{u}_i , que a su vez son independientes de $\hat{\beta}_1$ y $\hat{\beta}_2$. Por lo tanto, $\hat{\beta}_1$ y $\hat{\beta}_2$ son independientes de SRC.

Corolario. Bajo las hipótesis N y si el término constante está en modelo o si los datos han sido centrados:

$$\frac{\hat{\beta}_1 - \beta_1}{ee(\hat{\beta}_1)} \rightarrow t_{n-2}, \frac{\hat{\beta}_2 - \beta_2}{ee(\hat{\beta}_2)} \rightarrow t_{n-2}$$

Demostración:

Se demuestra la segunda, la primera se demuestra de manera equivalente:

$$\left(\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\left(\frac{\sigma^2}{S_{xx}}\right)}} \right) \rightarrow N(0,1) \text{ tal que } \frac{SRC}{\sigma^2} \rightarrow \chi_{n-2}^2$$

$$\left[\frac{\left(\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\left(\frac{\sigma^2}{S_{xx}}\right)}} \right)}{\left(\sqrt{\left(\frac{SRC}{(n-2)\sigma^2}\right)} \right)} \right] = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\left(\frac{S_R^2}{S_{xx}}\right)}} \rightarrow t_{n-2}$$

Si los datos han sido centrados, estrictamente se tiene que $\beta_1 = \hat{\beta}_1 = 0$. Siempre es posible la estimación de β_1 con igualdad $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ tal que todas las fórmulas anteriores son válidas. Se tendrá cuidado cuando la regresión no tiene término constante y $\bar{y} \neq 0$.

2.6.2. Normalidad asintótica

Si no se puede asegurar que vector U sigue una ley normal, se usar resultado asintótico tal que si n es suficientemente grande:

$$\hat{\beta} \approx N_k[\beta, \sigma^2(X^tX)^{-1}] \text{ tal que } \max_{1 \leq i \leq n} x_i^t(X^tX)^{-1}x_i \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

Se conoce que x_i son filas de matriz de diseño tal que indica que no existen puntos palanca.

La matriz X^tX es simétrica definida positiva, h_{ii} se interpreta como norma euclídeana de vector x_i respecto a $(X^tX)^{-1}$ tal que $h_{ii} = \|x_i\|^2(X^tX)^{-1}$ donde se demuestra que $h_{ii} = \left(\frac{1}{n}\right) [1 + \tilde{x}_i^t S_{xx}^{-1} \tilde{x}_i]$.

Donde $\tilde{x}_i = (x_{i2} - \bar{x}_2, x_{i3} - \bar{x}_3, x_{i4} - \bar{x}_4, \dots, x_{ik} - \bar{x}_k)$, S_{xx} es matriz de varianzas-covarianzas entre columnas de X tal que $\tilde{x}_i^t S_{xx}^{-1} \tilde{x}_i$ es Distancia Mahalanobis, medida de distancia introducida por Mahalanobis en 1936, su utilidad radica en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales, se diferencia de la distancia euclídea en que tiene en cuenta la correlación entre las variables aleatorias tal que distancia de Mahalanobis entre dos variables aleatorias con la misma distribución de probabilidad \vec{x}, \vec{y} con matriz de covarianza Σ se define como:

d_m (Distancia Mahalanobis) $(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})}$, de $(x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})$ al centro $(\bar{x}_2, \bar{x}_3, \bar{x}_4, \dots, \bar{x}_k)$ tal que un valor alto de h_{ii} indica que $(x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})$ está alejado del centro. Además, en teorema de Gauss-Markov se demuestra $\hat{U} = (I - H)U$ tal que $\hat{u}_i = u_i - \sum_{j=1}^n h_{ij}u_j$ tal que $\hat{u}_i \approx u_i$ ssi $\sum_{j=1}^n h_{ij}u_j \approx 0$ pues $E(\hat{u}_i) = 0$ y $\text{Var}(\sum_{j=1}^n h_{ij}u_j) = (\sigma^2)(\sum_{j=1}^n h_{ij}^2) = \sigma^2 h_{ii} \Rightarrow \hat{u}_i \approx u_i$ ssi $h_{ii} \approx 0$, el residuo \hat{u}_i es buen indicador del error u_i ssi $h_{ii} \approx 0$ tal que se puede demostrar que $\frac{1}{n} \leq h_{ii} < 1$ tal que el residuo es informativo sólo cuando $h_{ii} \approx \frac{1}{n}$ y si n es grande $h_{ii} \approx 0$. Si $h_{ii} \approx 1$, por teorema bajo hipótesis H , si matriz X es de rango completo y no aleatoria:

$E[\hat{U}] = 0$ y $\text{Var}[\hat{U}] = \sigma^2(I - H)$, $\text{Var}(\hat{U}) = (\sigma^2)(I - H)$ donde $H = XX^t(X^tX)^{-1} \Rightarrow \forall_i \text{Var}(\hat{u}_i) = \sigma^2(1 - h_{ii})$, $\text{Var}(\hat{u}_i) \approx 0 \Rightarrow \hat{u}_i \approx 0$ o, equivalente a, $y_i \approx \hat{y}_i$. La superficie de regresión pasaría muy cerca de punto (x_i, y_i) sin tomar en cuenta valor y_i ; es decir, es punto palanca.

Estos afectan a normalidad de errores, pero pueden tener influencia alta o baja en estimaciones de parámetros tal que para constatar su influencia se elimina el punto y se estiman nuevamente parámetros, si hay cambios importantes indica que es un punto atípico influyente y requiere tratamiento “especial”.

La influencia de un punto cualquiera se mide mediante estadístico de Cook: sea \hat{y}_i predicción de y en $x_i = (1, x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})^t$, con todas las observaciones mientras $\hat{y}_{i(i)}$ predicción de y en x_i , cuando se elimina (x_i, y_i) en estimación de recta de regresión. El estadístico de Cook es:

$$D_i = \left[\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{k(\widehat{\text{Var}}(\hat{y}_i))} \right] = \left[\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{k(s_R^2)(h_{ii})} \right] = \left[\left(\frac{r_i^2}{k} \right) \left(\frac{h_{ii}}{1 - h_{ii}} \right) \right]$$

Se considera que un punto es moderada e influyente, respectivamente, si:

$$\frac{|\hat{y}_i - \hat{y}_{i(i)}|}{\sqrt{\widehat{\text{Var}}(\hat{y}_i)}} > 1; \frac{|\hat{y}_i - \hat{y}_{i(i)}|}{\sqrt{\widehat{\text{Var}}(\hat{y}_i)}} > \sqrt{2}$$

Es decir, si $D_i > \left(\frac{1}{k}\right)$ o $D_i > \left(\frac{2}{k}\right)$ en orden. Además, si un punto palanca no es influyente puede afectar la normalidad de vector β , pues nunca estará seguro por completo de la normalidad de errores y, si existen puntos con valor h_{ii} alto se puede usar resultado asintótico $\hat{\beta} \approx N_k[\beta, \sigma^2(X^tX)^{-1}]$ tal que $\max_{1 \leq i \leq n} x_i^t(X^tX)^{-1}x_i \rightarrow 0$ cuando $n \rightarrow \infty$. Además, residuos estudentizados se pueden estimar mediante:

$$\hat{t}_i = \frac{\hat{u}_i}{S_{R(i)}^2 \text{SRC}(\text{Suma de Residuos al Cuadrado después de eliminar } i\text{-ésimo punto}) \sqrt{1-h_{ii}}} \Rightarrow$$

punto atípico si $|\hat{t}_i| > 3$.

Es sencillo comprobar:

$$h_{ii} = x_i x_i^t (X^t X)^{-1}$$

Forma la diagonal de matriz sombrero H tal que, si se tienen suficientes observaciones y no existen puntos palanca, intervalos de confianza de Student, así como pruebas t y F , tendrán niveles asintóticos, aproximados, pues entre más grande sea n su aproximación será mayor.

2.6.3. Intervalos de confianza

Se está en capacidad de estimar intervalos de confianza para los parámetros. Si θ es un parámetro y si $(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)$ son variables aleatorias independientes y su distribución depende de θ tal que:

$$[A(Y_1, Y_2, Y_3, Y_4, \dots, Y_n), B(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)]$$

Es un intervalo de confianza para θ , de nivel η y sólo si:

- $A(Y_1, Y_2, Y_3, Y_4, \dots, Y_n) \leq B(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)$ y
- $\Pr^\theta(A(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)) \leq \theta \leq B(Y_1, Y_2, Y_3, Y_4, \dots, Y_n) \geq \eta$

Esto significa que el verdadero valor del parámetro está dentro del intervalo de confianza (aleatorio) con probabilidad superior o igual a η . Si se tiene la igualdad se dice que el intervalo de confianza es de nivel exacto η . El valor típico para $\eta = 0.95$. Si:

$$\Pr^\theta(A(Y_1, Y_2, Y_3, Y_4, \dots, Y_n)) \leq \theta \leq B(Y_1, Y_2, Y_3, Y_4, \dots, Y_n) = \eta$$

El intervalo de confianza es de nivel exacto η ().

Corolario. Bajo hipótesis N_y si el término constante está en regresión o si datos han sido centrados:

$$\hat{\beta}_1 \pm ee(\hat{\beta}_1)t_{n-2}\left(\frac{\alpha}{2}\right)$$

$$\hat{\beta}_2 \pm ee(\hat{\beta}_2)t_{n-2}\left(\frac{\alpha}{2}\right)$$

Tal que $t_{n-2}\left(\frac{\alpha}{2}\right)$ es fractil de orden $1 - \left(\frac{\alpha}{2}\right)$ de Ley de Student con $n - 2$ grados de libertad, son intervalos de confianza de nivel exacto $\eta = 1 - \alpha$ para β_1 y β_2 , respectivamente.

Demostración:

$$\begin{aligned} & \Pr\left(\hat{\beta}_2 - ee(\hat{\beta}_2)t_{n-2}\left(\frac{\alpha}{2}\right) \leq \beta_2 \leq \hat{\beta}_2 + ee(\hat{\beta}_2)t_{n-2}\left(\frac{\alpha}{2}\right)\right) \\ &= \Pr\left(-t_{n-2}\left(\frac{\alpha}{2}\right) \leq \frac{\hat{\beta}_2 - \beta_2}{ee(\hat{\beta}_2)} \leq t_{n-2}\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha \end{aligned}$$

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 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ 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.

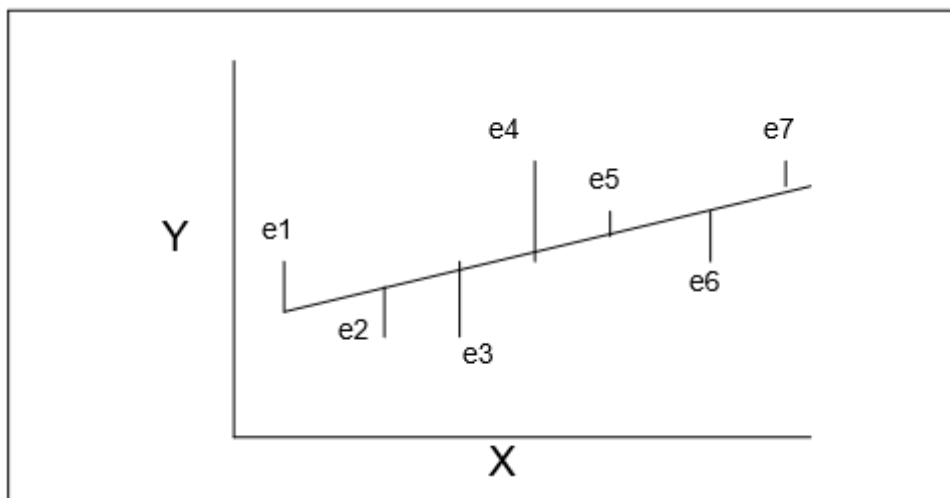


Figura. 2. 13. Representación de cada residuo (error), entendido como la distancia que existe entre valores observados de Y_i y valores predichos por el modelo

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), $b_1 \rightarrow \beta_1$, $e_i \rightarrow \varepsilon_i$ 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 $b_0 + b_1 X_i$ en esta expresión, el resultado es $\sum (Y_i - b_0 + b_1 X_i)^2 = \sum e_i^2$

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, $\frac{\sum e_i^2}{b_0} = 2 \sum (Y_i - b_0 + b_1 X_i) (-1) = 0$ y $\frac{\sum e_i^2}{b_1} = 2 \sum (Y_i - b_0 + b_1 X_i) (-X_i) = 0$ tal que arreglando términos se obtienen las ecuaciones normales: $\sum Y_i = n b_0 + b_1 \sum X_i$ y $\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2$. 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:

$$\begin{aligned} n \sum X_i Y_i &= n b_0 \sum X_i + n b_1 \sum X_i^2 \\ \sum X_i \sum Y_i &= n b_0 \sum X_i + b_1 \left(\sum X_i \right)^2 \\ \hline n \sum X_i Y_i - \sum X_i \sum Y_i &= n b_1 \sum X_i^2 - b_1 \left(\sum X_i \right)^2 \end{aligned}$$

Despejando b_1 de esta última ecuación queda $b_1 = \frac{[n \sum X_i Y_i - \sum X_i \sum Y_i]}{[n \sum X_i^2 - (\sum X_i)^2]}$ o, en forma más conocida, se divide por n tanto arriba como abajo $b_1 = \frac{[\sum X_i Y_i - (\frac{\sum X_i \sum Y_i}{n})]}{[\sum X_i^2 - (\frac{(\sum X_i)^2}{n})]}$ 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 : $Y = b_0 + b_1 X$. Tal que Y y X son las medias de Y y de X , respectivamente. Por lo tanto, $b_0 = Y - b_1 X$ obtiene la ecuación de ajuste que permite obtener cualquier valor de Y dado un X y resolviendo, de acuerdo a los coeficientes estimados, se obtiene $Y_i = b_0 + b_1 X_i$, definida como **ecuación de predicción**.

En consecuencia, sea el modelo de regresión lineal simple $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ 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 tienen capacidad explicativa de variación de Y.

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):

Fundamentos matemáticos de regresión lineal

Tabla. 2. 1. Análisis de varianza ANOVA para Regresión Lineal Simple (RLS)

Análisis de Varianza o ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher					
F. de V (Factor de Variación) o VF (Variation factor)	Gl (Grados de Libertad) o Df (Degrees of freedom)	SC (Suma de Cuadrados) o Sum Sq (Sum of square)	SCM (Cuadrado Medio) o Sum Mean Sq (Mean squares)	F (calculada) o F value (Calculated F)	F (tablas)
Total	n grados de libertad (total de observaciones)	$\sum Y_i^2$			
Media	1 grado de libertad (siempre)	$\frac{(\sum Y_i)^2}{n}$			
Regresión	1 grado de libertad (sólo 1 variable)	$\frac{[\sum X_i Y_i - (\sum X_i \sum Y_i)/n]^2}{[\sum X_i^2 - ((\sum X_i)^2)/n]}$	$\frac{S.C.\text{-Regresión}}{gl.\text{-Regresión}}$	$\frac{SCM(\text{Regresión})}{SCM(\text{Error})}$	
Intragrupo o Error	(n - 2)	$\sum e_i^2$	$\frac{S.C.\text{-Error}}{gl.\text{-Error}}$		

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:

Tabla. 2. 2. Ecuaciones de estimadores de interés para un X dado

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	$\sigma^2 b_0 = \sigma_e^2 \left[\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \right]$
$\hat{y}_i = b_0 + b_1 X_i$	$Y_i = \beta_0 + \beta X_i$	$\sigma_{V/X}^2 = \sigma_e^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{n \sum (X_i - \bar{X})^2} \right]$

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:

Tabla. 2. 3. Ecuaciones de estimadores con distribución normal y “t”

Estimador	Elemento esperado
Para b_1	$\left[\frac{b_0 - \beta_0}{\sigma_{b_0}^2} \right] = t_{(n-2, \alpha)}$
Para b_0	$\left[\frac{b_1 - \beta_1}{\sigma_{b_1}^2} \right] = t_{(n-2, \alpha)}$
Para $Y_{i/X}$	$\left[\frac{Y_{i/X} - \mu_{i/X}}{\sigma_{V/X}^2} \right] = t_{(n-2, \alpha)}$

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 $H_a: \beta_1 \neq 0 \Rightarrow t_c = \frac{(b_1 - 0)}{\sigma_{b_1}^2}$.

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 $Y_i = b_0 + b_1X_i$ 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 \bar{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 logarítmica básica es la función $y = \log_b x$, donde $b > 0$ y $b \neq 1$):

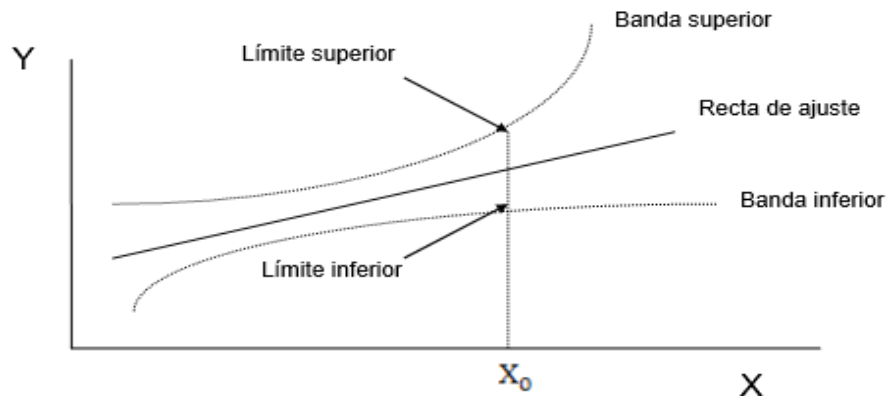


Figura. 2. 14. Representación de intervalos de confianza para una regresión lineal o recta de ajuste

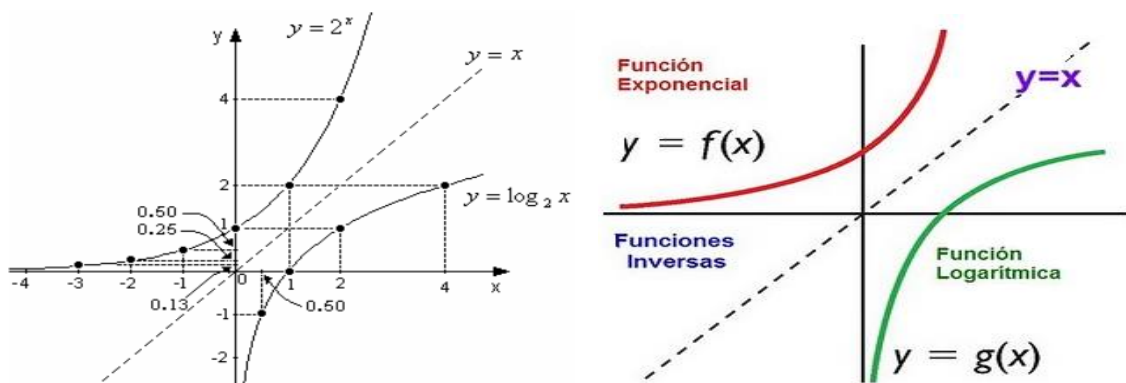


Figura. 2. 15. Función que es la operación inversa a exponenciación de base del logaritmo; es decir, e^x y su dominio es el conjunto de \mathbb{R} .

2.7. Ejemplos:

a) Con base en las siguientes observaciones :

Tabla. 2. 4. Observaciones para ejemplo

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 $Y(\quad)$. Los errores estándar de estimadores b_0 y b_1 son:

$$\sigma_{b_0}^2 = \sigma_e^2 \left[\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \right] = 7 * \left[\frac{42}{4 * 6} \right] = \sqrt{12.25} = \mathbf{3.5}_{(\text{Desviación Estándar o } \sigma_{b_0})}$$

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

Para probar que $b_0 \neq 0$, se divide el estimador b_0 entre su error estándar y se obtiene la "t_c": $t_c = \frac{b_0}{\sigma_{b_0}} = -\frac{1}{3.5} = -\mathbf{0.2857}$.

Después, este valor se compara con el obtenido en tabla de "t_{t(n-2; α=0.05)}" = 4.303 ⇒ $|t_c| < |t_t|$ ∴ 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 $t_c = \frac{b_1}{\sigma_{b_1}} = \frac{2}{1.08} = \mathbf{1.8} \Rightarrow |t_c| < |t_t|$ ∴ 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:

$$\sigma_{V/X}^2 = \sigma_e^2 \left[\frac{1}{n} * \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] = 7 * \left[\frac{(6-3)^2}{6} \right] = 12.25 \Rightarrow \sigma_{V/X} = \sqrt{12.25} =$$

3.50 (Desviación Estándar o σ_{b_0}). El valor de “ $t_{t(Gl(Error)); \alpha=0.05}$ ” = 4.303 y “ $t_{t(Gl(Error)); \alpha=0.01}$ ” = 9.925 \Rightarrow : el intervalo de confianza con 95 o 99% de confiabilidad estadística será $\hat{Y}_i \pm (t_{t(Gl(Error)); \alpha} * \sigma_{V/X})$ tal que el verdadero valor de \hat{Y}_i cuando X_0 presenta valores X_i :

Tabla. 2. 5. Estimación de intervalos de confianza para ejemplo

X_0	\hat{Y}_i	$\sigma_{V/X}^2$	$\sigma_{V/X}$	$t_{(Tablas)}$		Intervalo de confianza	
				95 %	99 %	\underline{L} (Límite inferior)	\bar{L} (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	9	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.

2.8. Pruebas de hipótesis

Las pruebas de hipótesis en la regresión lineal múltiple se emplean para determinar la significación de la regresión; es decir, si globalmente las variables aportan información al modelo y realizar pruebas sobre valores de coeficientes individuales para examinar si una variable particular es significativa en modelo y merece ser incluida en la ecuación. El intervalo de confianza delimita la región en que probablemente se encuentra el verdadero valor del parámetro tal que existe

a lo sumo una probabilidad $\alpha = 1 - \eta$ que esté fuera del intervalo. Esto se usa para contrastar o probar las hipótesis:

$$\begin{cases} \mathbf{H}_0: \beta_1 = 0 \\ \mathbf{H}_0 \text{ o } \mathbf{H}_1: \beta_1 \neq 0 \end{cases} \quad \begin{cases} \mathbf{H}_0: \beta_2 = 0 \\ \mathbf{H}_0 \text{ o } \mathbf{H}_1: \beta_2 \neq 0 \end{cases}$$

H_0 se lee “H subíndice 0”, pues H significa hipótesis y el subíndice 0 señala “no hay diferencia”. En cambio, H_1 o H_a , leído como “H subíndice 1 o a” señala que “al menos 1 es diferente respecto al resto de los tratamientos”.

La demostración matemática que lo sustenta se encuentra en la siguiente nota al pie y da razón que exista H_0 e H_a en una prueba de dos colas.

Además, con base en una muestra de una variable aleatoria X de ley P_θ en decidir entre dos hipótesis:

$H_0: \theta \in \theta_0$ (hipótesis nula o privilegiada) vs $H_1: \theta \in \theta_1$ (hipótesis alternativa o alterna) tal que $\theta_0 \cup \theta_1 = \theta$ y $\theta_0 \cap \theta_1 = \emptyset$. Si d_0 y d_1 representan, respectivamente, las decisiones de no rechazar H_0 o H_1 y $D = \{d_0, d_1\}$.

Entonces: sea $X: \Omega \rightarrow E \subseteq \mathbb{R}$ una variable aleatoria de ley P_θ . Se llama “Prueba de Hipótesis Pura” o test puro a toda aplicación: $\emptyset: E^n \rightarrow D$ tal que \emptyset equivale a particionar E^n en dos conjuntos:

$W = \emptyset^{-1}(d_0) = \{x \in E^n: \text{si se observa } x, \text{ no se acepta } H_0\}$ y $W^c = \emptyset^{-1}(d_1) = \{x \in E^n: \text{si se observa } x, \text{ no se rechaza } H_0\}$. El nombre de H_0 proviene que H_0 se asumirá como verdadera, salvo que datos muestrales indiquen su falsedad. Nula debe entenderse como neutra.

H_0 nunca se considera probada o demostrada, salvo estudiando todos los datos de la población, puede diferir en un valor pequeño imperceptible para el muestreo, que puede ser imposible; aunque, puede no ser aceptada por los datos. Con base en esto, no se debe afirmar “se acepta H_0 ”, siendo lo correcto “no se rechaza H_0 ” y, por abuso de lenguaje, es común hallar la expresión “se acepta la H_0 ” en lugar de “no se rechaza H_0 ”.

Generalmente, H_0 se elige según con el principio de simplicidad científica, que establece que únicamente se abandona un modelo simple a favor de otro más complejo cuando la evidencia a favor de este último sea fuerte. Por el carácter dicotómico, si no se acepta H_0 , automáticamente no se rechaza H_1 .

Finalmente, se llama **riesgo de primera especie a la probabilidad de rechazar H_0 cuando es verdadera**: $\alpha_0(\emptyset) = P_0(W_{(\text{Región crítica de prueba})})$, mientras que el **riesgo de segunda especie es la probabilidad de aceptar H_0 cuando es falsa**: $\beta_\theta(\emptyset) = P_1(W_{(\text{Región de aceptación})}^c) = 1 - P_1(W)$. Se llama "Potencia de una prueba" a la probabilidad de no aceptar H_0 cuando es falsa $-P_1(W) = 1 - \beta_\theta(\emptyset) = \gamma$, también, se denomina "Nivel de significación de una prueba" al valor $\alpha = \sup_{\theta \in \theta_0} \alpha_\theta(\emptyset)$.

Cuando $\theta \subseteq \mathbb{R}$ se estudian problemas del tipo:

$$P_0: \begin{cases} H_0: \theta = \theta_0, \\ H_1: \theta = \theta_1 \text{ con } \theta_0 \neq \theta_1 \end{cases}, \quad P_1: \begin{cases} H_0: \theta \leq \theta_0, \\ H_1: \theta > \theta_0 \text{ (Prueba de cola derecha)} \end{cases},$$

$$P_2: \begin{cases} H_0: \theta \geq \theta_0, \\ H_1: \theta < \theta_0 \text{ (Prueba de cola izquierda)} \end{cases},$$

$$P_3: \begin{cases} H_0: \theta_0 \leq \theta \leq \theta_1, \\ H_1: \theta < \theta_0 \text{ (Prueba de cola izquierda)} \text{ o } \theta > \theta_1 \text{ (Prueba de cola derecha)} \text{ (} \theta_0 < \theta_1 \text{)} \end{cases},$$

$$P_4: \begin{cases} H_0: \theta < \theta_0 \text{ o } \theta > \theta_1, \\ H_1: \theta_0 \leq \theta \leq \theta_1, \end{cases} \text{ o } P_5: \begin{cases} H_0: \theta = \theta_0, \\ H_1: \theta \neq \theta_1 \end{cases}.$$

El **teorema Neyman-Pearson** indica, en contexto de prueba propuesta: $\forall \alpha \in [0, 1], \epsilon$ un test puro \emptyset , de nivel α , de potencia máxima, definido por la región crítica: $W = \{x \in E^n: \frac{L(\theta_0, x)}{L(\theta_1, x)} \leq k\} \rightarrow k$ se determina de la condición $\alpha = P_0(W)$. L representa la función de verosimilitud y x una observación de la muestra. Otras pruebas importantes, de nivel α , son:

$$A): \begin{cases} H_0: \mu > \mu_0, \\ H_1: \mu \leq \mu_0 \end{cases} \text{ y } B): \begin{cases} H_0: \mu < \mu_0, \\ H_1: \mu \geq \mu_0 \end{cases}$$

Se define el estadístico razón de t de Student:

$$t_j = \frac{\hat{\beta}_j}{ee(\hat{\beta}_j)}$$

Corolario. Bajo hipótesis N , si el término constante está en regresión o si datos han sido centrados. Se rechaza hipótesis nula (H_0) a nivel α si sólo si:

$$|t_j| \geq t_{n-2} \left(\frac{\alpha}{2} \right)$$

Tal que, el nivel α de una prueba es una cota superior para la probabilidad de cometer un error tipo I, rechazar H_0 cuando es verdadera. Cuando se hace una prueba es recomendable estimar la probabilidad crítica (ρ – Value), que es valor mayor de α en que no se rechaza H_0 .

Es decir, ρ – value es la probabilidad de observar un valor muestral tan extremo o más extremo que el valor observado dado que la H_0 es verdadera o es una manera de expresar la probabilidad que H_0 no sea verdadera o, según (Galindo, 2006), “mínimo valor del nivel de significación para que datos observados indican que H_0 será rechazada”.

Complementariamente, $\Pr(> F)$ es el nivel de probabilidad que H_a caiga en la zona de rechazo. Las investigaciones sociales usualmente trabajan con $\alpha = 0.10$ o 90 % de confiabilidad estadística, Diseños Experimentales con $\alpha \leq 0.05$ o 95 % de confiabilidad estadística, Control de Calidad $\alpha \leq 0.01$ o 99 % de confiabilidad estadística y Control de Calidad con 6σ $\alpha \leq 0.001$ o ≥ 99.9 % de confiabilidad estadística.

Demostración:

No se rechaza H_0 si i sólo si:

$$|t_j| < t_{n-2} \left(\frac{\hat{\alpha}}{2} \right)$$

Como la función de distribución es monótona creciente tal que:

$$\Pr(T_{n-2} < |t_j|) = 1 - \frac{\hat{\alpha}}{2} \text{ tal que } \hat{\alpha} = 2\Pr(T_{n-2} \geq |t_j|)$$

Además, estimar la potencia de prueba es importante, pues la probabilidad de rechazar H_0 cuando H_a o H_1 es la hipótesis verdadera; es decir, es correcto rechazar H_0 . Cuando $B_j = b_j \neq 0$ se llama η a función potencia:

$$\eta(b_j) = \Pr (|t_j| < t_{n-2}(\frac{\hat{\alpha}}{2}) | B_j = b_j)$$

Bajo hipótesis $B_j = b_j$, la razón t_j sigue una ley de Student descentrada con $n - 2$ grados de libertad y parámetro descentramiento $\delta^2 = \frac{b_j^2}{\text{Var}(\hat{B}_j)}$, que puede ser aproximada por una ley de Student centrada. Los paquetes estadísticos presentan potencias de pruebas para $B_j = \hat{B}_j$ y reemplazando $\text{Var}(\hat{B}_j)$ por $\hat{\text{Var}}(\hat{B}_j)$; es decir, se reemplaza σ^2 por S_R^2 . Si rechaza que pendiente de recta es nula se interpreta como regresión es significativa.

En este caso, no se rechaza que Y depende linealmente de X ; sino $\beta_2 = 0$ tal que Y no depende linealmente de X (gráfica izquierda) que puede ser motivo de 1. Y no depende funcionalmente de X o relación no es lineal (gráfica derecha es función cuadrática):

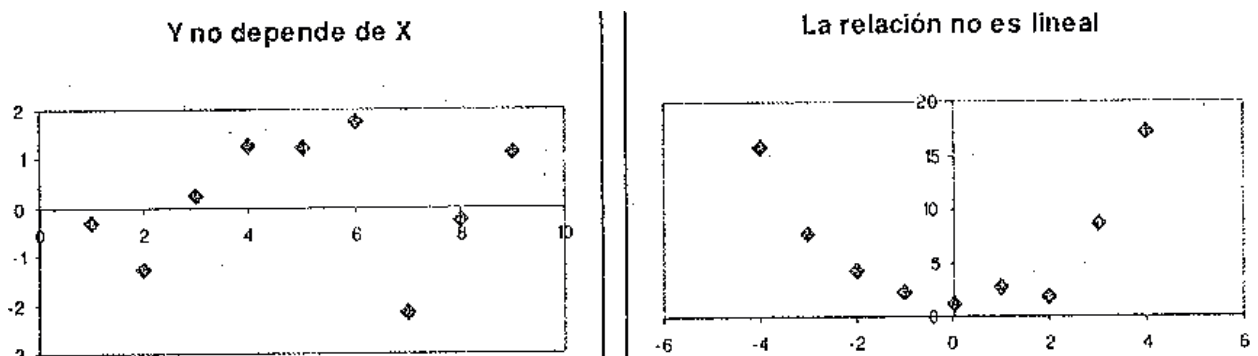


Figura. 2. 16. Dependencia lineal de Y respecto a X

Para probar la significancia de regresión se acostumbra usar prueba de Fisher, que en caso de regresión lineal simple es equivalente a prueba t de Student para la pendiente. El cálculo del estadístico F de Fisher o F de Snedecor se acostumbra escribir la tabla de análisis de varianza (ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher).

Una tabla ANOVA descompone la varianza en diferentes rubros, que en regresión lineal se descompone en varianza explicada por regresión y varianza debido al error.

En otras palabras, está asociado a un nivel crítico de probabilidad de obtener valores, como el obtenido o mayores. Si éste es mayor al nivel de error asumido, habitualmente 95 % de confiabilidad estadística, error o nivel de significancia (α) igual a 0.05, se interpreta “con base en resultados obtenidos de datos muestrales (poblacionales), no se acepta hipótesis nula, que indica igualdad de medias de tratamientos ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_k = \mu(0)$ o $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \dots \tau_k = 0$) y no se rechaza la alternativa, que indica diferencia de al menos una media de tratamiento respecto al resto ($H_a: \mu_i \neq \mu_j$ para algún $i \neq j$ o $H_a: \tau_i \neq 0$ para algún i)”.

Con diferencia parcial respecto a (Gujarati & Porter, 2010) que afirman “con base en una prueba de significancia, como prueba t, se dice “aceptar” la hipótesis nula, todo lo que se afirma es que, con base en la evidencia dada por la muestra, no existe razón para rechazarla, no se sostiene que la hipótesis nula sea verdadera con absoluta certeza” y, de acuerdo con (González, 1985), “de la misma manera que un tribunal se pronuncia un veredicto de “no culpable” en vez de “inocente”, así la conclusión de una prueba estadística es “no rechazar” en vez de “aceptar”.

$|F_{\text{Calculada}}| > |F_{\text{Tablas}}(\alpha, k-1, N-k)|$ F_{Tablas} de Fisher Tablas $(\alpha_{\text{(Tamaño de error o nivel de significancia)}, k-1_{\text{(Grados de libertad del numerador)}, N-k_{\text{(Grados de libertad del denominador)}})$ se tiene el criterio más formal y/o riguroso de lectura de una ANOVA que no se acepta hipótesis nula, que indica igualdad de medias de tratamientos ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_k = \mu(0)$ o $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \dots \tau_k = 0$) y no se rechaza la alternativa, que indica diferencia de al menos una media de tratamiento respecto al resto ($H_a: \mu_i \neq \mu_j$ para algún $i \neq j$ o $H_a: \tau_i \neq 0$ para algún i).

Si

$|F_{\text{Calculada}}| < |F_{\text{Tablas}}(\alpha, k-1, N-k)|$ F_{Tablas} de Fisher Tablas

(α (Tamaño de error o nivel de significancia), $k - 1$ (Grados de libertad del numerador), $N - k$ (Grados de libertad del denominador)) se tiene el criterio más formal y/o riguroso de lectura de una ANOVA que no se rechaza hipótesis nula, que indica igualdad de medias de tratamientos ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots = \mu_k = \mu (0)$ o $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \dots \tau_k = 0$) y no se acepta la alternativa, que indica diferencia de al menos una media de tratamiento respecto al resto ($H_a: \mu_i \neq \mu_j$ para algún $i \neq j$ o $H_a: \tau_i \neq 0$ para algún i).

Tabla. 2. 6. Análisis de varianza de Fisher

Análisis de Varianza o ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher					
F. de V (Factor de Variación) VF (Variation factor)	Gl (Grados de Libertad) Df (Degrees of freedom)	SC (Suma de Cuadrados) Sum Sq (Sum of square)	CM (Cuadrado Medio) Mean Sq (Mean squares)	F (calculada) F value (Calculated F)	Pr(> F) ρ - Value
Total	(n - 1)	STC (Suma Total de Cuadrados)			
Intergrupo Regresión	(1) (gl numerador)	SEC (Suma de Estimaciones al Cuadrado)	SEC / gl (numerador)	$\frac{C. M. Regresión}{C. M. Error}$	$\frac{gl(\text{numerador})}{gl(\text{denominador})}$
Intragrupo Error	(n - 2) (gl denominador)	SRC (Suma de Residuos al Cuadrado)	SRC / gl (denominador)		

Las sumas de cuadrados en que la regresión tiene término constante o datos han sido centradas. Las medias o promedios de sumas de cuadrados se obtienen dividiendo las sumas de cuadrados por grados de libertad. Es decir, son el número de contrastes ortogonales menos el número de restricciones impuestas, que pueden hacerse en un grupo de datos.

Los grados de libertad pueden descomponerse al igual que la suma de cuadrados en

Gl total = Gl entre grupos o Tratamientos + Gl dentro de grupos o Error; aunque, sus divisiones pueden aumentar según el diseño experimental que se trabajó.

Además, refiere a un número de valores a escoger libremente son el número de datos que son libres de variar cuando se calcula tal prueba o indican el número de dimensiones en que vectores pueden variar libremente, introducida la expresión por Fisher, de un conjunto de observaciones están dados por el número de valores que pueden ser asignados arbitrariamente, antes que el resto de variables queden completamente determinadas.

Según (Castro, 2008), los grados de libertad corresponden a dimensiones de sub espacios vectoriales en que se encuentran los datos tal que un conjunto generador se representa a partir que vectores $v_1, v_2, v_3, v_4, \dots, v_n$ 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, $\forall v \in V_{(\text{Espacio Vectorial})} \exists$ escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n \Rightarrow v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$.

Además, un espacio generado por un conjunto de vectores se genera a partir que sea $v_1, v_2, v_3, v_4, \dots, v_k$ k vectores de un espacio vectorial V . El espacio generado por $\{v_1, v_2, v_3, v_4, \dots, v_k\}$ es el conjunto de combinaciones lineales $v_1, v_2, v_3, v_4, \dots, v_k$; es decir,

$$\text{gen}\{v_1, v_2, v_3, v_4, \dots, v_k\} = \{v | v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_k v_k\}.$$

Por lo tanto, se dice que es **linealmente dependiente** (L. D. o D. L. en inglés) si sean $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$, n vectores en espacio vectorial V tal que los vectores son linealmente dependientes si existen n escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n$ NO TODOS SON IGUALES A CERO; es decir, con algún $\alpha_i \neq 0 \Rightarrow \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n = \vec{0}$.

Caso contrario, estos serán **linealmente independientes** (L. I. o I. L. en inglés) tal que, la única combinación lineal que da cero en la TRIVIAL es $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \dots = \alpha_n = 0$. Por ejemplo: $STC_{(Suma\ Total\ de\ Cuadrados)}$ es norma del vector cuyas componentes son $\{y_i - \bar{y} | i = 1, 2, 3, 4, \dots, n\}$ que se encuentra en un sub espacio vectorial de \mathbb{R}^n de dimensión $n - 1$. $SEC_{(Suma\ de\ Estimaciones\ al\ Cuadrado)}$ es suma de predicciones que se hallan sobre una recta tal que está asociada a un sub espacio vectorial de dimensión 1.

Para la suma $\sum_{i=1}^n y_i^2$ los grados de libertad son n , pues es la norma del vector $(y_1, y_2, y_3, y_4, \dots, y_n) \in \mathbb{R}$ tal que para $SRC_{(Suma\ de\ Residuos\ al\ Cuadrado)}$ en una regresión sin término constante los grados de libertad son $n - 1$. Sin embargo, si los datos han sido centrados $\sum y_i = 0$, el vector $(y_1, y_2, y_3, y_4, \dots, y_n)$ se encuentra sobre un sub espacio vectorial de \mathbb{R}^n de dimensión $n - 1$.

Teorema.

Bajo hipótesis N , si la regresión tiene término constante para $\beta_2 = 0$ tal que la razón F sigue la ley Fisher-Snedecor con $(1, n - 2)$ grados de libertad ($F \rightarrow F_{(1, n-2)}$). Este teorema permite no aceptar $H_0: \beta_2 = 0$ a favor de H_a o $H_1: \beta_2 \neq 0$ al nivel de α si $F_{(Calculado)}$ es superior al fractil $(F_{(1, n-2), \alpha})$ de ley de Fisher con $(1, n - 2)$ grados de libertad, de orden $1 - \alpha$; es decir, la cola derecha de la función de densidad es igual a α .

La probabilidad crítica para esta prueba es:

$$\hat{\alpha} = \Pr(F_{(1, n-2)} \geq F_{(Tablas)})$$

De acuerdo con (González, 1985), “de la misma manera que un tribunal se pronuncia un veredicto de “no culpable” en vez de “inocente”, así la conclusión de una prueba estadística es “no rechazar” en vez de “aceptar”. Esto se basa, en parte, en las siguientes razones:

A) (Wackerly, Mendehall, & Scheaffer, 2010) señalan “la probabilidad α recibe el nombre de nivel de significancia o, en forma más sencilla, nivel de prueba. Aun cuando se recomiendan con frecuencia pequeños valores de α , el valor real α para usar en un análisis es un tanto arbitrario. Sin embargo, **ρ – value es el nivel de significancia alcanzado, relacionado con una prueba y es un estadístico que representa el valor más pequeño de α para el cual la información muestral indica que H_0 puede ser rechazada.** En cierto sentido, ρ – value permite al lector de la investigación evaluar la magnitud de la discrepancia entre datos observados e H_0 ”.

B) (Navidi, 2006) afirma “si se rechaza H_0 se concluye que era falsa, en cambio si H_0 no se rechaza no se concluye que H_0 es verdadera, pues sólo se puede concluir que H_0 es factible, pero nunca se puede llegar a la conclusión H_0 es verdadera debido a que, por un lado, el estadístico de prueba es consistente con hipótesis H_a y está un poco en desacuerdo con H_0 tal que la única cuestión es si el nivel de desacuerdo medido con ρ – value es suficientemente grande para presentar H_0 como no factible y, por otro lado, **una regla general conveniente, considerando ρ – value como menor que un umbral específico, indica rechazar o no aceptar H_0 cada vez que $\rho \leq 0.05$, interpretando que es “estadísticamente significativo”;** aunque, **este criterio no tiene ninguna base científica y, además, ésta es una mala práctica por varias razones:**

1) No proporciona ninguna manera de decir si ρ – value era apenas menor que 0.05 o si era mucho menor,

2) Notificar que un resultado era estadísticamente significativo a un nivel de 5% implica que hay gran diferencia entre ρ – value justo debajo de 0.05 y uno justo arriba de 0.05, cuando efectivamente hay una diferencia pequeña,

3) Un trabajo así no permite al lector decidir por ellos mismos si ρ – value es suficientemente pequeño para rechazar o no aceptar H_0 . Por ejemplo: Si un lector cree que H_0 no debe rechazarse a menos que $\rho < 0.01$ entonces informar que solamente que $\rho < 0.05$ no permite al lector determinar si rechaza o no acepta o acepta o no rechaza H_0 ,

4) Valores pequeños de ρ – value señalan que H_0 es improbable que sea verdadera, tal que es tentador pensar que ρ – value representa la probabilidad que H_0 sea verdadera. No obstante, la verdad o falsedad de H_0 no se puede cambiar mediante la repetición del experimento; por lo tanto, no es correcto hablar de “probabilidad” que H_0 sea verdadera,

5) La clase de probabilidad que analiza si no se rechaza o no se acepta H_0 , útil solamente cuando se aplica a resultados que pueden resultar en formas diferentes cuando se repiten experimentos debido a que para definir el ρ – value como probabilidad de observar un valor extremo de un estadístico como \bar{X} , pues su valor podría ser diferente si el experimento se repite, se llama **probabilidad frecuentista** (la frecuencia de un suceso en una muestra se define como el cociente entre número de veces que ha ocurrido el suceso en la muestra y el tamaño de la misma.

Empíricamente, se observa que al ir aumentando el tamaño de una muestra, la frecuencia de sucesos tiende a estabilizarse de un número fijo tal que se ha denominado como **ley de estabilidad de frecuencias** o **ley única del azar** y ese número ideal, límite que alcanzaría la frecuencia de un suceso si se obtuviera una muestra infinita del experimento es el primer concepto de probabilidad de un suceso),

6) La **probabilidad subjetiva** calcula la probabilidad que un enunciado, como H_0 , sea verdadero y es importante en la teoría de Estadística Bayesiana (es un subconjunto del campo de la estadística en la que la evidencia sobre el verdadero estado del mundo se expresa en términos de grados de creencia o, más específicamente, las probabilidades bayesianas. Es sólo una de una serie de

interpretaciones de la probabilidad y hay otras técnicas estadísticas que no se basan en "grados de creencia".

La inferencia bayesiana es un enfoque de la inferencia estadística, que es distinta de la inferencia frecuentista. Se basa específicamente en el uso de probabilidades bayesianas al resumir las pruebas.

La formulación de modelos estadísticos para su uso en la estadística bayesiana tiene la característica adicional, no está presente en otros tipos de técnicas estadísticas, que requiere la formulación de un conjunto de distribuciones previas para los parámetros desconocidos, sus consideraciones habituales en diseño de experimentos se extienden en el caso de diseño Bayesiano de experimentos para incluir la influencia de las creencias anteriores y requiere ciertas técnicas computacionales modernas para la inferencia bayesiana, como diversos tipos de técnicas Monte Carlo de cadenas de Markov) y

7) Si un resultado tiene un ρ – value pequeño se dice que es “estadísticamente significativo”, representa “importante” y, en consecuencia, resulta tentador pensar que resultados estadísticamente significativos siempre son importantes. No obstante, a veces este tipo de resultados no tienen importancia científica o práctica.

Entonces, **un resultado puede ser estadísticamente significativo sin ser lo suficientemente grande para tener importancia práctica**, pues una diferencia es estadísticamente significativa cuando es grande comparada con su σ o s_e , incluso, cuando σ o s_e es muy pequeña puede ser estadísticamente significativa.

Por lo tanto, **ρ – value no mide la significancia práctica, sino el grado de confianza que se puede tener que el valor verdadero es muy diferente del valor especificado por H_0** tal que cuando ρ – value es pequeño se tiene la confianza que el valor verdadero es, en verdad, muy diferente, pero no implica que la diferencia sea lo bastante grande para que tenga importancia práctica”.

C) (Lind, Marchal, & Mason, 2006) sostienen “en una prueba de hipótesis sólo se puede tomar una de dos decisiones: aceptar o rechazar H_0 . En lugar de

“aceptar” H_0 algunos investigadores prefieren enunciar la decisión como “No se rechaza H_0 ”, “No se puede rechazar H_0 ” o, también, “Los resultados muestrales no permiten rechazar H_0 ”. En pruebas de hipótesis p – value, proporciona la probabilidad de obtener, suponiendo que H_0 sea verdadera, un valor muestral estadístico de prueba tan extremo, por lo menos o más extremo, como el obtenido tal que p – value compara la probabilidad con el nivel de significancia”. Por lo tanto, si no se acepta H_0 , el siguiente paso es efectuar la prueba de significancia entre medias de tratamientos con el fin de conocer cuál o cuáles son estadísticamente mejores.

D) (Triola, 2004) afirma “estadístico de prueba es un valor calculado a partir de datos muestrales, que se usa para tomar la decisión sobre rechazo de H_0 . Por lo tanto, sirve para determinar si existe evidencia significativa en contra de H_0 . p es la probabilidad de obtener un valor del estadístico de prueba que sea al menos tan extremo como el que representa a datos muestrales, suponiendo que H_0 es verdadera.

Algunos libros dicen “aceptar H_0 ” en vez de “no rechazar H_0 ”, pero no se está probando H_0 , sea que use el término “aceptar” o “no rechazar”, tal que únicamente se afirma que la evidencia muestral no es lo suficientemente fuerte como para justificar el rechazo de H_0 . El término “aceptar” es un poco confuso, pues parece indicar incorrectamente que H_0 ha sido aprobada tal que la frase “no rechazar” **indica correctamente que la evidencia disponible no es lo suficientemente fuerte para justificar el rechazo de H_0** ”.

Usualmente, en ejercicios se pide demostrar que $t^2 = F$ tal que $t = \frac{\hat{\beta}_2}{ee(\hat{\beta}_2)}$ quemuestra que las pruebas Student y Fisher son equivalentes.

Observación.

Si $\beta_2 = b \neq 0$, F sigue una ley de Fisher descentrada con igual grados de libertad y parámetro de descentramiento $\delta^2 = \frac{b_2^2}{\text{Var}(\hat{\beta}_2)}$. La potencia de prueba es:

$$\eta(b) = Pr(F \leq F_{(1, n-2), \alpha} | \beta_2 = b)$$

2.9. Coeficiente de determinación

Es otra medida de relación entre las variables llamada coeficiente de determinación R^2 . Se debe a que da mayor fuerza de interpretación a la relación entre las variables. Además, es un buen indicador de “calidad” de regresión, pero no es determinante ni suficiente para decidir sobre la adecuación del modelo. Su uso es muy difundido, pero en general inapropiado pues no tienen presentes las limitaciones y alcances de este indicador. Además, no estudian residuos.

2.9.1. Coeficiente R^2

Una regresión será “buena” si la variabilidad explicada por regresión es relativamente alta respecto a variabilidad total de Y ; es decir, si $SEC \approx STC$. La proporción de variabilidad explicada por regresión se mide con coeficiente de determinación.

Definición. Se llama coeficiente de determinación R^2 al cociente:

$$R^2 = \frac{SEC}{STC}$$

Teorema. Si la regresión tiene término constante o si $\bar{y} = 0$, $0 \leq R^2 \leq 1$.

Demostración.

Se sabe que si la regresión tiene término constante o si $\bar{y} = 0$, entonces:

$$STC = SEC + SRC$$

Tal que, la suma de cuadrados es positiva y el cociente de dos números positivos es positivo. Si la regresión no incluye el término constante i $\bar{y} \neq 0$ no existe razón para que R^2 esté entre 0 y 1.

Interpretación. Se multiplica por 100 y se interpreta en términos de porcentaje tal que es el porciento de variabilidad explicada por la regresión con respecto a variabilidad total. La definición e interpretación de R^2 tiene sentido únicamente cuando la regresión lineal incluye el término constante o los datos han sido centrados y, por consecuencia, no hay término constante en la regresión.

Es importante que R^2 sea cercano a 1, pero no es determinante para calidad de regresión. Su calidad se decide en función de todos sus indicadores, como coeficiente de determinación, razón F, signos esperados, rangos esperados para estimadores de parámetros β_j , pruebas de hipótesis sobre parámetros, gráficos de residuos y pruebas de hipótesis sobre residuos. Se puede demostrar que $F = (n - 2) \left(\frac{R^2}{1 - R^2} \right)$ tal que R^2 será significativo si F lo es. El coeficiente puede ser “bajo” y la regresión significativa.

El coeficiente R^2 se puede usar en comparaciones de dos modelos de regresión tal que se prefiere el modelo con mayor valor de R^2 . Para comparar dos modelos con el coeficiente R^2 es importante que la variable dependiente sea la misma.

Por ejemplo $y_i = \beta_0 + \beta_1 x_i + u_i$ i $\ln(y_i) = \beta_0 + \beta_1 x_i + u_i$ no son comparables con R^2 . En primer caso R^2 es propoción de variabilidad de Y explicada por X mientras que en segundo caso es la variabilidad de $Z = \ln(Y)$ explicada por la misma variable. Este coeficiente puede ser fuertemente afectado por datos atípicos.

Coeficiente de Determinación r^2 (dos variables) ó R^2 (regresión múltiple) como medida de “bondad de ajuste”. La bondad del ajuste de la línea de regresión de un conjunto de datos refiere a cuán “bien” se ajusta la línea de regresión a los datos. Con base en la siguiente figura, es claro que si todas las observaciones caen en la línea de regresión se obtiene un ajuste “perfecto”, pero rara ocasión se presenta:

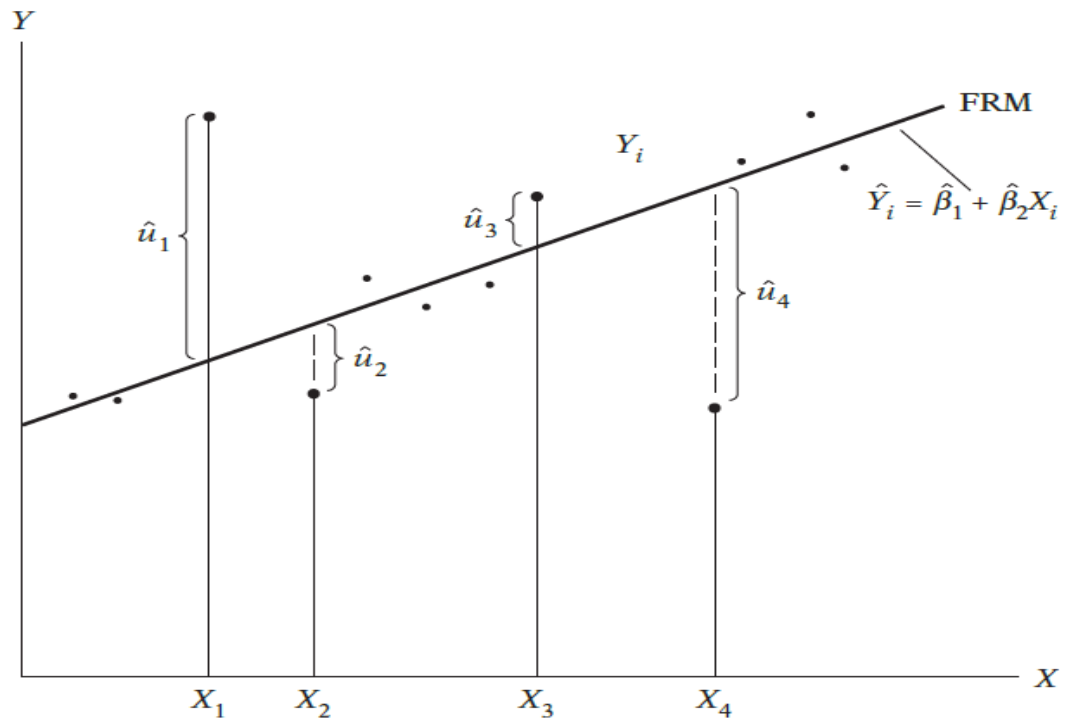


Figura. 2. 17. Bondad del ajuste de la línea de regresión de un conjunto de datos refiere a cuán “bien” se ajusta la línea de regresión a los datos

En general, hay algunas \hat{u}_i positivas y algunas \hat{u}_i negativas. Se tiene la esperanza que estos residuos alrededor de la línea de regresión sean lo más pequeños posibles. Entonces, el **coeficiente de determinación r^2** (dos variables) ó **R^2** (regresión múltiple) es una medida comprendida que indica cuán bien se ajusta la línea de regresión muestral a los datos. Una explicación del significado de heurística () de r^2 ó R^2 en términos gráficos, conocida como **Diagrama de Venn, Euler** o **Ballentine** (r^2 : a) = 0; f) $r^2 = 1$) es:

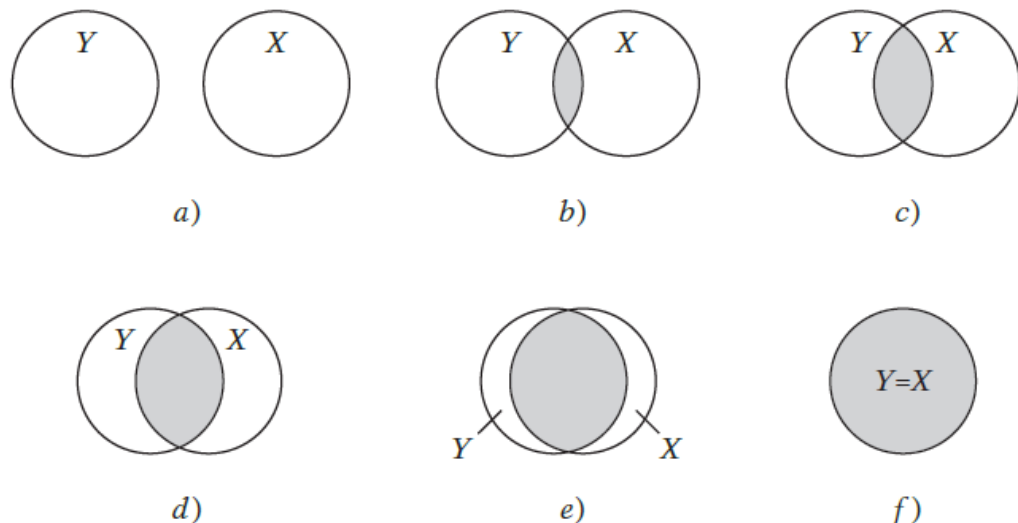


Figura. 2. 18. Diagrama de Venn, Euler o Ballentine ($r^2: a) = 0; f) r^2 = 1)$

El círculo Y representa la variación en variable explicada, dependiente o endógena Y, el círculo X es la variación en la variable explicativa, independiente o exógena X (Varianza_(Suma de cuadrados dividida por grados de libertad apropiados) = Variación_{(Suma de cuadrados de desviaciones de una variable respecto a su media)/Gl), mientras que la intersección de círculos (área sombreada) señala la media en que la variación en Y se explica por variación en X, como regresión de MCO.}

A mayor medida de intersección, mayor será la variación en Y que se explica por X, pues a medida que va de izquierda a derecha, el área de intersección aumenta o, en otras palabras, hay una proporción cada vez mayor de la variación en Y explicada por X.

Entonces, r^2 ó R^2 es una medida numérica de esta intersección. Por lo tanto, cuando no hay intersección r^2 ó $R^2 = 0$, cuando es completa r^2 ó $R^2 = 1$, entendida como 100% de variación de Y se explica por X y, por lógica, r^2 ó R^2 varía ± 1 . El cálculo de r^2 ó R^2 se hace de la siguiente forma, pues si:

$$Y_i = \hat{Y}_i + \hat{u}_i \text{ o, en términos de desviación, } y_i = \hat{y}_i + \hat{u}_i$$

Donde se emplean ecuaciones $y_i = \hat{\beta}_2 x_i + \hat{u}_i$ y $\hat{y}_i = \hat{\beta}_2 x_i$ tal que al elevar al cuadrado $y_i = \hat{y}_i + \hat{u}_i$ en ambos lados y sumar sobre la muestra se obtiene:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2$$

Pues $\sum \hat{y}_i \hat{u}_i = 0$ y $\hat{y}_i = \hat{\beta}_2 x_i$. Las diversas sumas de cuadrados en ecuación anterior se describen de la forma siguiente:

$$\sum y_i^2 = \sum (Y_i - \bar{Y})^2$$

Hace referencia a **variación total de valores reales de Y** respecto de su media muestral, denominada **Suma de Cuadrados Total (SCT)**.

$$\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum x_i^2$$

Refiere a la **variación de valores de Y** estimador alrededor de su media ($\hat{Y}_i = \bar{Y}$), que apropiadamente puede llamarse **Suma de Cuadrados por Regresión**, debida a variables explicativas, independientes o exógenas, o, simplemente, **Suma de Cuadrados explicada (SCE)**. Adicionalmente, $\sum \hat{u}_i^2$ es **variación residual o variación no explicada** de valores de Y alrededor de la línea de regresión o, también, llamada **Suma de Cuadrados de Residuos (SCR)**. Así, ecuación:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \Rightarrow$$

$$\therefore \text{SCT} = \text{SCE} + \text{SCR}$$

Esta ecuación muestra que la **variación total** en valores Y observados alrededor del valor de su media puede dividirse en dos partes, una atribuible a la línea de regresión y otra a fuerzas aleatorias, pues no todas las observaciones Y caen sobre línea ajustada. Geométricamente es:

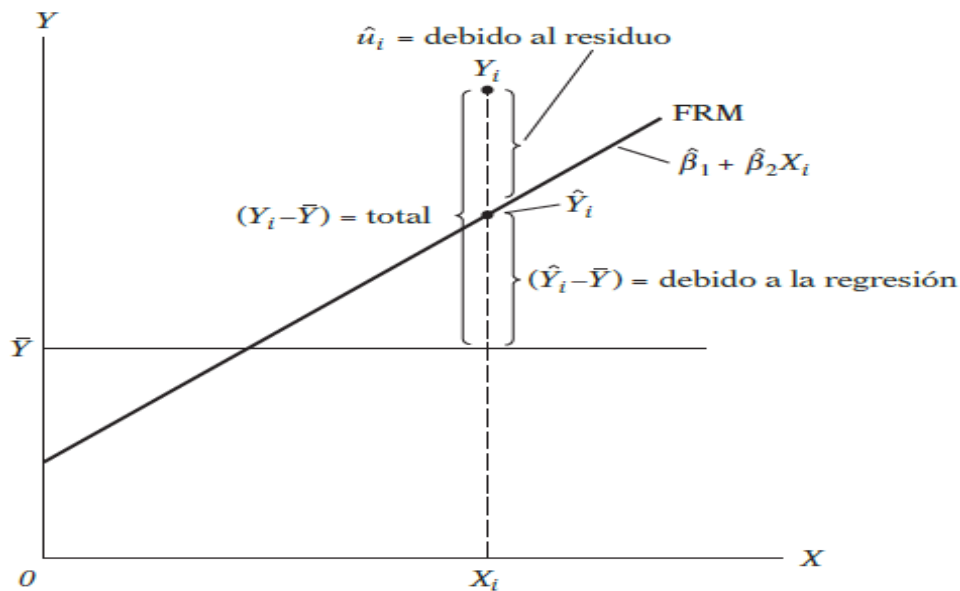


Figura. 2. 19. Variación total en valores Y observados alrededor del valor de su media puede dividirse en dos partes, una atribuible a la línea de regresión y otra a fuerzas aleatorias

Al dividir la ecuación $\text{SCT} = \text{SCE} + \text{SCR}$ entre SCT en ambos lados, se obtiene:

$$\frac{\text{SCT}}{\text{SCT}} = \frac{\text{SCE}}{\text{SCT}} + \frac{\text{SCR}}{\text{SCT}} \Rightarrow 1 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

Ahora bien, se define r^2 :

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{SCE_{(Error)}}{SCT_{(Total)}}$$

ó

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{SCR_{(Regresión)}}{SCT_{(Total)}}$$

La cantidad definida de esta manera se conoce como **“Coeficiente de Determinación Muestral”**, entendida como la medida más común de bondad de ajuste de una lineal de regresión o, verbalmente, r^2 mide la proporción o por ciento de variación total en Y explicada por el modelo de regresión. Asimismo, r^2 observa dos propiedades:

1. Es una cantidad no negativa, pues al elevar al cuadrado cualquier número, sea positivo o negativo, el resultado es positivo (ley de signos).

2. Sus límites son $0 \leq r^2 \leq 1$. $r^2 = 1$ indica un ajuste perfecto; es decir, $\hat{Y}_i = Y_i$ por cada i . Por otro lado, $r^2 = 0$ significa que no hay una relación alguna entre la variable regresada, dependiente, explicada o endógena y variable regresora, independiente, explicativa o exógena (coeficiente de pendiente o de variable exógena $\hat{\beta}_1 = 0$). En este caso, según $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i = \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})$, $\hat{Y}_i = \hat{\beta}_1 = \bar{Y}$; es decir, la mejor predicción de cualquier valor de Y es simplemente el valor de su media. En consecuencia, la línea de regresión será horizontal al eje X.

No obstante, r^2 puede calcularse directamente a partir de la definición o, más rápido, con la fórmula siguiente:

$$\begin{aligned} r^2 &= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{SCR_{(Regresión)}}{SCT_{(Total)}} = \frac{SCE_{(Error)}}{SCT_{(Total)}} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} \\ &= \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right) \end{aligned}$$

Si se dividen el numerado y denominador de la ecuación anterior por el tamaño n muestral (n – 1 si la muestra es pequeña), sabiendo que $S_x^2 - S_y^2$ son varianzas muestrales X – Y respectivamente, se obtiene:

$$r^2 = \hat{\beta}_2^2 \left(\frac{\frac{\sum x_i^2}{n}}{\frac{\sum y_i^2}{n}} \right) = \hat{\beta}_2^2 \left(\frac{S_x^2}{S_y^2} \right)$$

Como $\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum x_i^2}$, la ecuación $\hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right)$ también se expresa:

$$r^2 = \frac{(\sum X_i Y_i)^2}{\sum x_i^2 \sum y_i^2} = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)}$$

Con la definición de r^2 , $SCE_{(Error)}$ y $SCR_{(Regresión)}$ explicadas antes, se expresan así:

$$SCE_{(Error)} = r^2 * SCT_{(Total)} = (r^2) \left(\sum y_i^2 \right)$$

Por lo tanto:

$$\begin{aligned} SCR_{(Regresión)} &= SCT_{(Total)} - SCE_{(Error)} = SCT_{(Total)} \left(1 - \frac{SCR_{(Regresión)}}{SCT_{(Total)}} \right) \\ &= \left(\sum y_i^2 \right) (1 - r^2) \end{aligned}$$

Por lo tanto, se escribe:

$$SCT_{(Total)} = SCE_{(Error)} + SCR_{(Regresión)}$$

ó

$$\sum y_i^2 = r^2 \sum y_i^2 + (1 - r^2) \left(\sum y_i^2 \right)$$

2.9.2. Coeficiente de correlación

Si se tienen dos variables aleatorias, una medida de la relación que existe entre ellas es el coeficiente de correlación ρ . Para determinar si existe una relación lineal entre las variables productora y de respuesta se utiliza el coeficiente de correlación lineal de Pearson, denotado por r, que se define por:

$$r = \frac{SC_{xy}}{\sqrt{SC_{xx}SC_{yy}}} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)}}$$

Incluso, dada una pareja de variables aleatorias (X, Y) se define el coeficiente de correlación:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{[\text{Var}(X)\text{Var}(Y)]}} = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Tal que $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$.

Teorema. Para $\forall (X, Y)$ par de variables aleatorias, $|\rho| \leq 1$.

Teorema. Si las variables X, Y son independientes tal que $\rho = 0$. Sin embargo, el recíproco de este teorema es, en general, falso. Si la distribución conjunta es normal, entonces independencia es equivalente a no correlación. Se puede demostrar que si las parejas $(X_i, Y_i)_{i=1,2,3,4,\dots,n}$ son independientes, igualmente distribuidas con ley normal tal que el coeficiente de correlación (empírico) definido:

$$r = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_i (y_i - \bar{y})^2 \sum_i (x_i - \bar{x})^2}}$$

Es estimador de máxima verosimilitud de ρ . Además, si:

$$z = \tanh^{-1}(r) = \frac{1}{2} \text{Ln} \left(\frac{1+r}{1-r} \right)$$

Se tiene:

$$\sqrt{n-3}(z - \tanh^{-1}(\rho)) \rightarrow_{\text{Ley}} N(0,1)$$

Esto indica que para n suficientemente grande:

$$z \approx N \left(\tanh^{-1}(\rho), \frac{1}{n-3} \right)$$

Se puede usar este resultado asintótico para estimar intervalos de confianza para ρ o para contrastar hipótesis. Sin embargo, el coeficiente de correlación r ($\sqrt{r^2}$) es la cantidad estrechamente relacionada con el concepto de Coeficiente de

Determinación r^2 (dos variables) ó R^2 (regresión múltiple) como medida de “bondad de ajuste”.

El **Coefficiente de Correlación Muestral** explica el grado de asociación, mide la fuerza o grado de asociación lineal, entre dos variables. Su cálculo es a partir de:

$$r = \pm\sqrt{r^2} = \frac{\sum X_i Y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} = \frac{n \sum X_i Y_i - [(\sum X_i)(\sum Y_i)]}{\sqrt{[n \sum x_i^2 - (\sum X_i)^2][n \sum y_i^2 - (\sum Y_i)^2]}}$$

El **Coefficiente de Correlación Poblacional** ρ (letra griega minúscula Rho o ro) es:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{[\text{Var}(X)\text{Var}(Y)]}} = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Entonces, ρ es una medida de Asociación Lineal entre dos variables y su valor se sitúa entre ± 1 , donde -1 indica una perfecta asociación negativa y $+1$ es una perfecta asociación positiva. Con base en lo anterior, se deduce:

$$\text{Cov}(X, Y) = \rho(\sigma_X \sigma_Y)$$

Algunas propiedades de r son:

I. Tener signo positivo o negativo, según el signo del término en numerador de $r = \pm\sqrt{r^2} = \frac{\sum X_i Y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} = \frac{n \sum X_i Y_i - [(\sum X_i)(\sum Y_i)]}{\sqrt{[n \sum x_i^2 - (\sum X_i)^2][n \sum y_i^2 - (\sum Y_i)^2]}}$ que mide

la Covariación Muestral de dos variables.

II. Se ubica entre límites -1 y $+1$; es decir, $-1 \leq r \leq 1$.

III. Es simétrico por naturaleza; es decir, el Coeficiente de Correlación entre X y Y , denotado por r_{XY} .

IV. Es independiente del origen y escala; es decir, si se define $X_i^* = aX_i + C$ y $Y_i^* = bY_i + d$ tal que $a > 0$, $b > 0$, c y d son constantes. Entonces, r entre X^* y Y^* es igual a r entre variables originales X y Y .

V. Si X y Y son estadísticamente independientes, pues dos variables aleatorias X y Y son estadísticamente independientes si y sólo si (\Leftrightarrow) $f(x, y) = f(x) f(y)$; es decir, si Función de Densidad de Probabilidad conjunta

se expresa como el producto de las FDP marginales, el Coeficiente de Correlación ente ellas es cero, aunque no significa que dos variables sean independientes. En otras palabras, una correlación igual a cero no necesariamente implica independencia (siguiente figura h).

Sea X una variable discreta que toma valores diferentes $x_1, x_2, x_3, x_4, \dots, x_n$ tal que la función $f(x) = P(X = x_i)$ para $i = 1, 2, 3, 4, \dots, n$ y $f(x) = P(X = x_i) = 0$ para $x \neq x_i$ se denomina **Función de Densidad de Probabilidad Discreta (FDP)** de X , donde $P(X = x_i)$ indica probabilidad que la variable discreta X tome valor de x_i .

En cambio, **Función de Densidad de Probabilidad de Variable Aleatoria Continua (FDP)** afirma que X es una variable aleatoria continua, se dice que $f(x)$ es FDP de X si satisfacen las condiciones que $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x) \delta x = 1$ y $\int_a^b f(x) \delta x = P(a \leq x \leq b)$ donde $f(x) \delta x$ es elemento probabilístico (probabilidad asociada a un pequeño interval de una variable continua) y $P(a \leq x \leq b)$ es probabilidad que X se encuentre en intervalo a a b tal que para una variable aleatoria continua, en contraste con una variable aleatoria discreta, la probabilidad que X tome valor específico es cero pues $\int_a^a f(x) \delta x = 0$, la probabilidad para tal variable solo se mide sobre un rango o intervalo dado, como (a, b) :

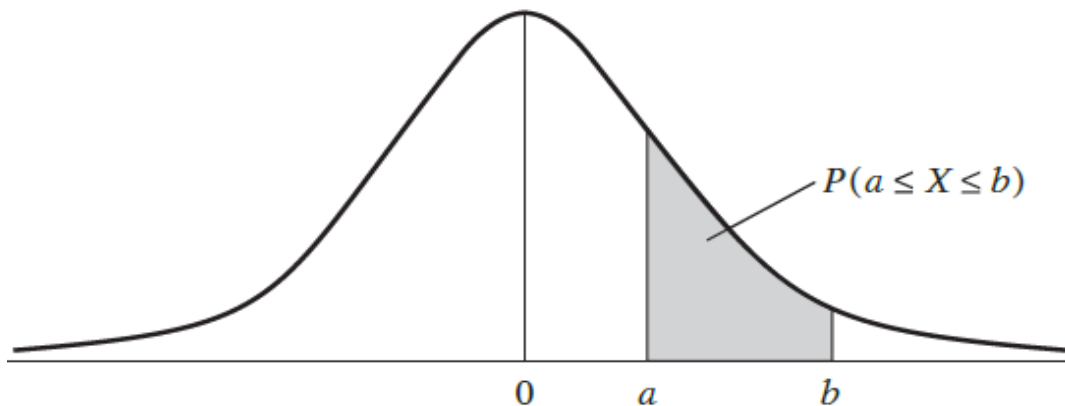


Figura. 2. 20. $P(a \leq x \leq b)$ es probabilidad que X se encuentre en intervalo a a b tal que para una variable aleatoria continua

La **Función de Densidad de Probabilidad Conjunta Discreta** indica que sean X y Y dos variables discretas tal que la función $f(x, y) = P(X = x \text{ y } Y = y) = 0$ cuando

$X \neq x$ y $Y \neq y$ y da probabilidad (conjunta) que X tome valor de x y Y tome valor y . La **Función de Densidad de Probabilidad Individual o Marginal** indica la relación con $f(x,y)$, $f(x)$ y $f(y)$ se denominan funciones de densidad de probabilidad individuales o marginales, obtenidas mediante $f(x) = \sum_y f(x,y)$ (FDP marginal de X) y $f(y) = \sum_x f(x,y)$ (FDP marginal de Y). Finalmente, **Función de Probabilidad Condicional** del comportamiento de una variable condicional respecto a valores de otra (s) variable (s): $f(x|y) = P(X = x|Y = y)$ tal que FDP condicional. Similarmente, $f(y|x) = P(Y = y|X = x)$, FDP condicional de Y .

Las FDP condicionales se obtienen por $f(x|y) = \frac{f(x,y)}{f(y)}$ (FDP condicional de X) y $f(y|x) = \frac{f(x,y)}{f(x)}$ (FDP condicional de Y) tal que la FDP de una variable se expresa como la razón de FDP conjunta respecto de FDP marginal de otra variable (condicionante).

VI. Es una medida de Asociación Lineal o Dependencia Lineal solamente, su uso es la descripción de relaciones no lineales no tiene significado. Así, en la siguiente figura h, $Y = X^2$ es una relación exacta y $r = 0$, pues es una expresión cuadrática (parábola positiva) y, en consecuencia, una relación no lineal.

VII. Es una medida de asociación lineal entre dos variables y no implica obligatoriamente una relación causa-efecto: “Una relación estadística por más fuerte y sugerente que sea nunca podrá establecer una conexión causal, por lo que ideas de causalidad provendrán de estadísticas externas y, finalmente, de una u otra teoría”. Por lo tanto, “una relación estadística por sí misma no puede, lógicamente, implicar causalidad”.

En el contexto de regresión, r^2 es una medida con más significado que r , pues la primera indica la proporción de variación en variable dependiente, explicada, predicha, regresada, respuesta, endógena, resultado o controlada explicada por la (s) variable (s) independiente (s), explicativa (s), predictora (s), regresora (s), estímulo (s), exógena (s), covariante (s) o de control (s). En consecuencia, constituye

una medida global del grado en que la variación en una variable determina la variación de otra r no tiene este valor y, además, la interpretación de r (R) en un modelo de regresión múltiple es de valor dudoso.

Enseguida se muestran los patrones de correlación de Theil citado por (Gujarati & Porter, 2010):

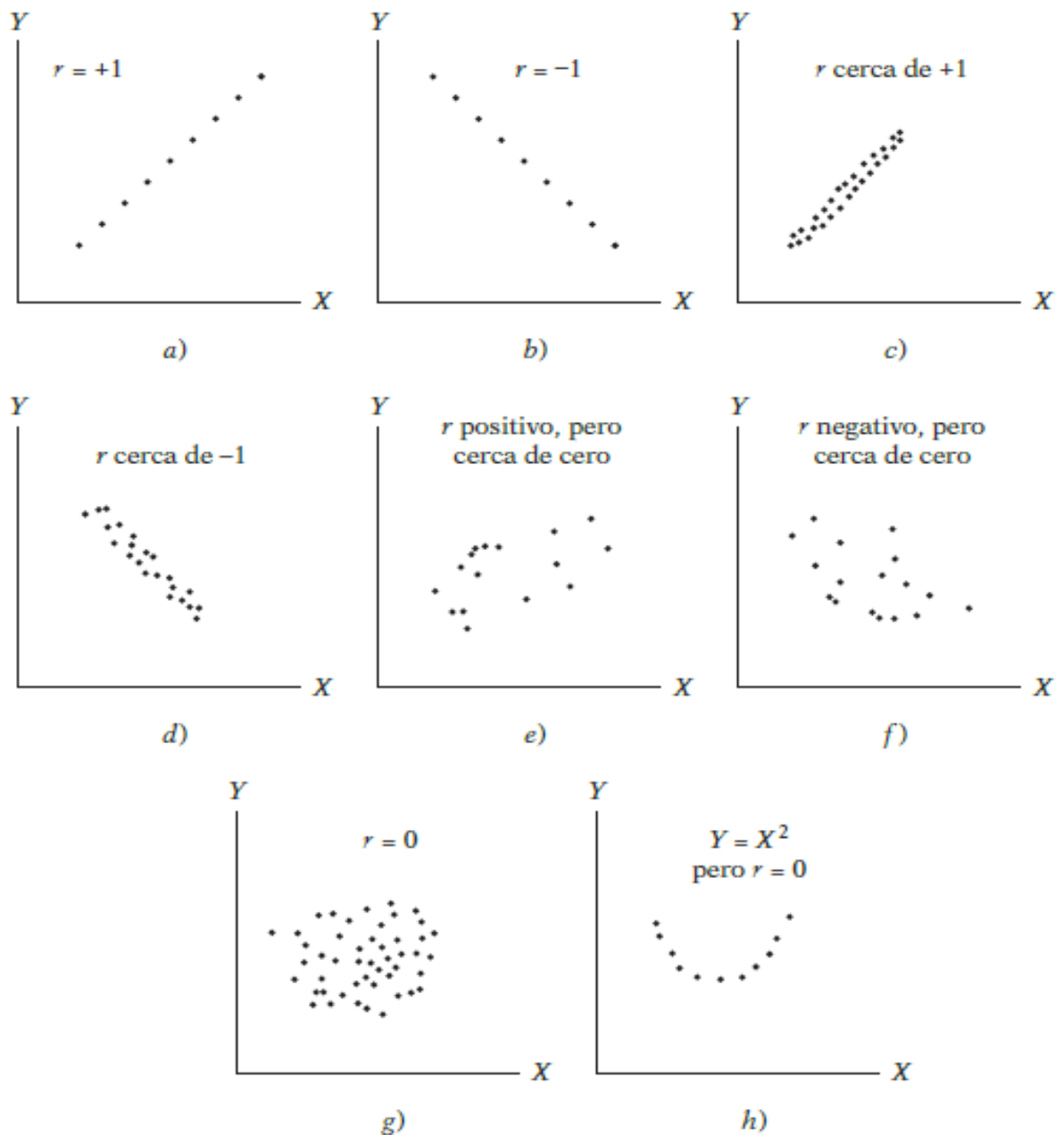


Figura. 2. 21. Patrones de correlación de r

Otra forma de considerarlo es:

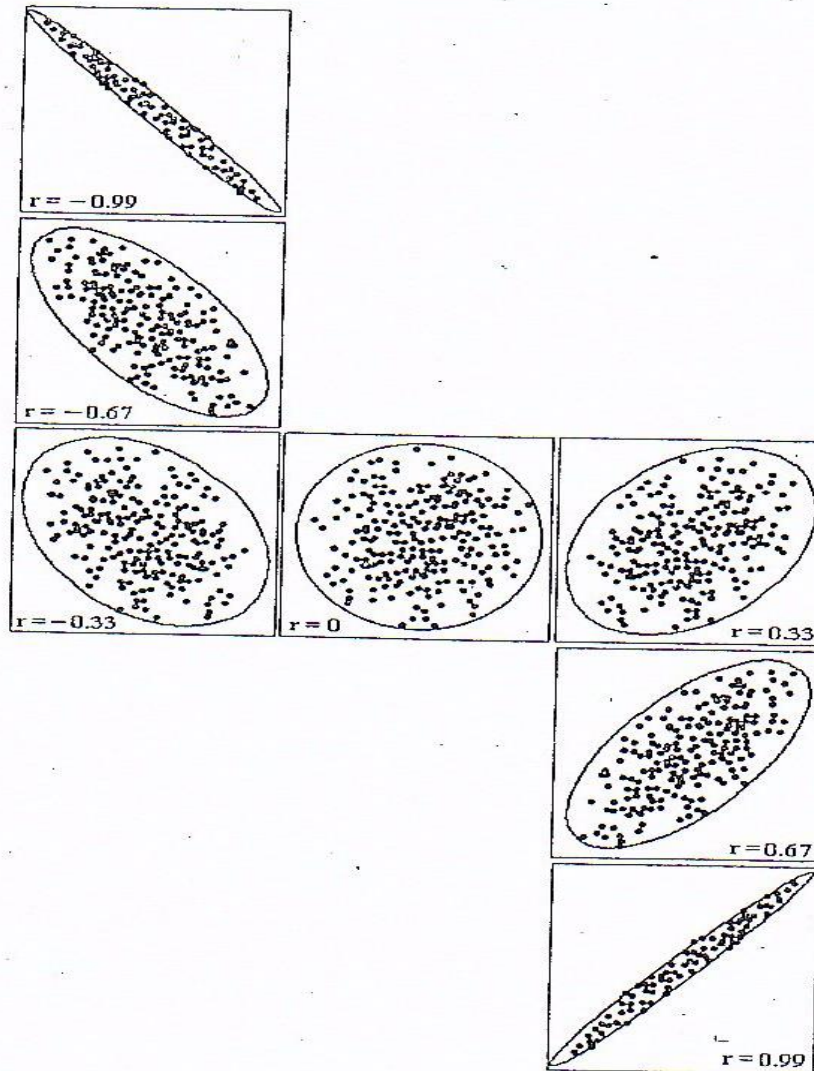


Figura. 2. 22. Patrones de correlación de r

Teorema. $R^2 = r^2$

Demostración:

$$r^2 = \left(\frac{1}{S_{yy}} \right) \left(\frac{S_{xy}^2}{S_{xx}} \right) = \left(\frac{SEC_{(\text{Suma de Estimaciones al Cuadrado})}}{STC_{(\text{Suma Total de Cuadrados})}} \right) = R^2$$

Tanto R^2 como r son una medida de relación lineal entre el conjunto de puntos (x_i, y_i) , linealidad que puede expresarse a todos los $x \in [x_{(\min)}, x_{(\max)}]$. Se demuestra fácilmente que $R^2 = 1$ si y sólo si existen constantes a, b tal que:

$$y_i = a + bx_i \quad i = 1, 2, 3, 4 \dots, n$$

Si la relación no es lineal o si no existe ningún tipo de relación funcional entre X, Y su coeficiente de correlación será nulo o casi nulo. Por ejemplo, los puntos $\{(-2,4), (-1,1), (0,0), (1,1), (2,4)\}$ están sobre una parábola y su coeficiente de correlación es nulo.

Prueba de hipótesis sobre ρ :

1. Hipótesis nula $H_0: \rho = 0$.
2. Hipótesis alternativa H_1 o $H_a: \rho \neq 0$.
3. Estadístico de prueba $t_{\text{Calculada}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$
4. Región de Rechazo. Se rechaza si H_0 si $t_{\text{obs}} < -t_{\alpha/2}(n-2)$ o $t_{\text{obs}} > t_{\alpha/2}(n-2)$

También, se pueden realizar pruebas unilaterales sobre ρ , pero ellas solo tienen un valor estadístico y su valor práctico es restringido.

2.10. Varianza de regresión

Al proponer el modelo probabilístico:

$$y = \beta_0 + \beta_1 + \varepsilon$$

Se estableció que el error aleatorio está distribuido según una ley normal de media cero y varianza σ^2 . Entonces, cada valor observado de y está influido por tal error. Además, a los errores asociados con distintas observaciones se los considera mutuamente independientes. La varianza σ^2 del error es desconocida y su estimador ingresado es:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} = \frac{SC_E}{n-2}$$

Así, s^2 es la desviación estándar de la regresión muestra y , también, es error estándar de estimación. Aunque s^2 se puede considerar una medida de la calidad de ajustes, su principal utilidad se encuentra en la determinación de la bondad de ajustes, sea mediante intervalo de confianza o con una prueba de hipótesis.

2.10.1. Inferencias de pendiente recta β_1

En primer lugar, se desea estudiar si existe o no existe relación entre las variables xy . Se contestar a la pregunta ¿aportar x información para predecir y ? Esta pregunta se refiere a β_1 , pues afirmar que y no se relaciona linealmente con x equivalente a decir que $\beta_1 = 0$. Entonces, se desea probar la hipótesis nula « x no contribuye con información para predecir y »; o la hipótesis alternativa, «las variables están relacionadas de forma lineal con una pendiente de cero»; es decir,

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

Para realización de la prueba habrá que encontrar la distribución de muestreo de β_1 . Distribución de muestreo de β_1 . Si los componentes del error son variables aleatorias independientes normalmente distribuida: media cero y varianza σ^2 , la distribución de muestreo del estimador β_1 es normal con media desviación estándar:

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{SC_{xx}}}$$

Esto quiere decir que β_1 es un estimador insesgado para β_1 , pues $E(\beta_1) = \beta_1$ y que la desviación estándar de β_1 puede estimarse mediante:

$$s_{b_1} = \frac{s}{\sqrt{SC_{xx}}}$$

Donde s es el error estándar de la estimación. Entonces la variable aleatoria $t = (\beta_1 - \hat{\beta}_1)/s_{b_1}$ sigue una ley t a $(n - 2)$ grado de libertad.

a. Prueba de hipótesis acerca de β_1 :

1. Hipótesis nula. $H_0 = 0$.
2. Hipótesis alternativa $H_1: \beta_1 < 0$ o bien $H_1: (\beta_1 > 0)$
3. Estadísticos de prueba $t_{(\text{calculada})} = \frac{b_1}{s/\sqrt{SC_{xx}}}$

4. Región de rechazo. Se rechaza H_0 si $t_{(\text{calculada})} >$

$t_{(\text{tablas}; \frac{\alpha}{2}; n-2)}$ cuando $H_1: (\beta_1 > 0)$.

b) prueba bilateral

1. Hipótesis nula $H_0: \beta_1 = 0$

2. Hipótesis alternativa $H_0: \beta_1 \neq 0$

3. Estadísticos de prueba $t_{(\text{calculada})} = \frac{b_1}{s/\sqrt{SC_{xx}}}$

4. Región de rechazo. No se acepta si H_0 si $t_{(\text{calculada})} \leq t_{(\text{tablas}; \frac{\alpha}{2}; n-2)}$ o si $t_{(\text{calculada})} > t_{(\text{tablas}; \frac{\alpha}{2}; n-2)}$. Intervalo de confianza de $100(1 - \alpha) \%$ para β_1 . Un intervalo de confianza para la pendiente β_1 , a nivel $100(1 - \alpha) \%$ es:

$$b = \frac{t\alpha/2(n-2)s}{\sqrt{SC_{xx}}}; + \frac{t\alpha/2(n-2)s}{\sqrt{SC_{xx}}}$$

2.10.2. Inferencias de pendiente recta β_0

Suponga que se desea averiguar si β_0 es igual o, no a un valor específico. Se logra realizando una prueba de hipótesis o con un intervalo de confianza, de manera similar a la descrita para β_1 . Se puede demostrar que la desviación estándar de b_0 es

$$\delta b_0 = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} \frac{\sum_{i=1}^n 2}{nSC_{xx}}}$$

El estadístico $t = \frac{(\beta_0 - \beta_{00})\sqrt{nSC_{xx}}}{\sqrt{\sum_{i=1}^n x_i^2}}$ sigue una distribución t con $(n - 2)$ grados de

libertad. Prueba de hipótesis acerca de β_0 :

1. Hipótesis Nula. $H_0: \beta_0 = \beta_{00}$.

2. Hipótesis Alternativa. $H_0: \beta_0 \neq \beta_{00}$.

3. Estadístico de Prueba. $t_{(\text{calculada})} = \frac{b_0 - \beta_{00}}{s/\sqrt{SC_{xx}}} \sqrt{\frac{nSC_{xx}}{\sum_{i=1}^n x_i^2}}$

4. Región de rechazo. No se acepta si H_0 si $t_{(\text{calculada})} \leq t_{(\text{tablas}; \frac{\alpha}{2}; n-2)}$ o si $t_{(\text{calculada})} > t_{(\text{tablas}; \frac{\alpha}{2}; n-2)}$. Intervalo de confianza de $100(1 - \alpha) \%$ para β_1 .

Un intervalo de confianza para la pendiente β_1 , a nivel $100(1 - \alpha) \%$ es:

$$b = \frac{t_{\alpha/2}(n-2)s}{\sqrt{SC_{xx}}}; + \frac{t_{\alpha/2}(n-2)s}{\sqrt{SC_{xx}}}$$

Intervalo de confianza de $100(1 - \alpha) \%$ para β_0

$$\left(b_0 - t_{\alpha/2}(n-2) \frac{\sqrt{\sum_{i=1}^n x_1^2}}{\sqrt{nSC_{xx}}} \quad b_0 + t_{\alpha/2}(n-2) \frac{\sqrt{\sum_{i=1}^n x_1^2}}{\sqrt{nSC_{xx}}} \right)$$

2.10.3. Adecuación de modelo

Una vez que se halla una estimación de la recta de regresión, es necesario determinar si la ecuación obtenida es un buen modelo para los datos y cuantificar el error que comete cuando se emplea tal ecuación.

Esto se logra mediante el empleo de los coeficientes de correlación y determinación y a través de la realización de pruebas estadísticas sobre los parámetros. La regresión lineal simple se basa en hipótesis:

H₁: Para $\forall i$ $E(u_i) = 0$.

H₂: Para $\forall i$ $Var(u_i) = \sigma^2$ llamada hipótesis de homocedasticidad de errores.

H₃: Para $\forall i$, todo $i \neq i'$ $Cov(u_i, u_{i'}) = 0$. \nexists correlación de errores. Cuando los datos son cronológicos se habla de correlación serial o autocorrelación en vez de correlación. La aditividad del error, la no aleatoriedad de variable independiente y parámetros, permiten que estas hipótesis se pueden escribir en función de momentos y_i .

H₁' Para $\forall i$ $E(Y_i|x_i) = \beta_1 + \beta_2 x_i$. Para enfatizar que las esperanzas dependen de los valores x_i se escribe $E(Y_i|x_i)$ en vez de $E(Y_i)$.

H_2 Para $\forall i$ $\text{Var}(Y_i|x_i) = \sigma^2$. La varianza es constante y no depende de x_i .

H_3 Para $\forall i, \text{todo } i' \neq i$ $\text{Cov}(Y_i, Y_{i'}|x_i, x_{i'}) = 0$. Las covarianzas son nulas y no dependen de $x_i, x_{i'}$.

Sin embargo, nunca han sido verificadas, pues los errores no son observables. Un primer indicio de su incumplimiento se tiene con el gráfico de dispersión $X - Y$ tal que si los puntos tienden a formar una curva (no recta) indica que la relación no es constante. Además, mayor cantidad de información respecto a hipótesis se con base en residuos puesto que dan información sobre errores.

2.10.4. Prueba de linealidad

Si existen observaciones repetidas, se puede hacer una prueba de ajuste a la linealidad. Suponga que para cada valor x_i se han realizado n_i mediciones u observaciones de variable respuesta tal que se tiene el modelo:

$$y_{ir} = \beta_1 + \beta_2 x_i + u_{ir} \text{ tal que } i = 1, 2, 3, 4, \dots, m; \quad r = 1, 2, 3, 4, \dots, n_i. \text{ Supone } \sum_{i=1}^m n_i = n, m > 2$$

Las medias:

$$\mu_i = E(y_{ir}) \quad \forall i, \forall r$$

Pueden ser estimados con promedios:

$$\bar{y}_i = \left(\frac{1}{n_i}\right) \left(\sum_{r=1}^{n_i} y_{ir}\right) \quad i = 1, 2, 3, 4, \dots, m$$

Con medias estimadas con base en regresión lineal por método MC (Mínimos Cuadrados):

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + u_{ir} \text{ tal que } i = 1, 2, 3, 4, \dots, m$$

Si la relación es lineal, efectivamente, las dos estimaciones deben ser aproximadamente iguales tal que la suma de cuadrados es:

$$\sum_{r=1}^m n_i (\bar{y}_i - \hat{y}_i)^2$$

Será pequeña respecto a variabilidad entre clases tal que la suma de variabilidades de cada uno de los grupos $\{y_{i1}, y_{i2}, y_{i3}, y_{i4}, \dots, y_{in_i}\}$.

Teorema. No se acepta hipótesis $H_0: \mu_i = \beta_1 + \beta_2 x_i + u_{ir}$ tal que $i = 1, 2, 3, 4, \dots, m$ al nivel α si i sólo si:

$$\left(\frac{n - m}{m - 2} \right) \left(\frac{\sum_{r=1}^m n_i (\bar{y}_i - \hat{y}_i)^2 \text{ (Error por mala adecuación)}}{\sum_{i=1}^m \sum_{r=1}^{n_i} (y_{ir} - \bar{y}_i)^2 \text{ (Error puro)}} \right) \geq F_{(m-2, n-m, \alpha)}$$

Definición. Estos resultados se escriben en una ANOVA tal que puede ser insertada en ANOVA principal:

Tabla. 2. 7. Análisis de varianza de Fisher

Análisis de Varianza o ANOVA, ADEVA, ANDEVA, ANVA, AVAR o ANOVA de Fisher					
F. de V o VF (Factor de Variación Variation factor)	Gl (Grados de Libertad) o Df (Degrees of freedom)	SC (Suma de Cuadrados) o Sum Sq (Sum of square)	CM (Cuadrado Medio) o Mean Sq (Mean squares)	F (calculada) o F value (calculated F)	Pr (> F) o ρ - Value
Error Total	$(n - 2)$	$\sum_{i=1}^m \sum_{r=1}^{n_i} (y_{ir} - \hat{y}_i)^2$			
Error por mala adecuación	$(m - 2)$	$\sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2$	SEMA/(m - 2)	$\frac{C. M_{SEMA}}{C. M_{SEP}}$	$\frac{(m - 2)}{(n - m)}$
Error puro	$(n - m)$	$\sum_{i=1}^m \sum_{r=1}^{n_i} (y_{ir} - \bar{y}_i)^2$	SEP/(n - m)		

2.10.5. Gráficos de residuos

No se puede conocer errores u_i , pero si calcular residuos $\hat{u}_i = y_i - \hat{y}_i$, que no son estimadores de errores, pero si proporcionan información respecto a ellos. Las hipótesis asumidas son linealidad, igualdad de varianzas (homocedasticidad), no correlación y normalidad y deben reflejarse en gráficos de residuos. Los residuos pueden señalar:

- Función de regresión no es lineal.
- Varianza de errores no es constante.
- Errores son correlacionados.
- Existen observaciones singulares.
- Errores no son normalmente distribuidos.
- Uno o más regresores son omitidos del modelo.

Residuos en función \hat{Y} y de X . Se gráfica puntos $\{(\hat{y}_i, \hat{u}_i) | i = 1, 2, 3, 4, \dots, n\}$ y, también, $\{(x_i, \hat{u}_i) | i = 1, 2, 3, 4, \dots, n\}$. Si los puntos están dentro de una franja horizontal simétrica a eje X (A -no existe evidencia de violación a hipótesis mientras que B -la relación no es lineal- C-varianza no es constante que crece con valores del eje abscisas- D-puede ser por error en análisis o cálculos- hipótesis son falsas):

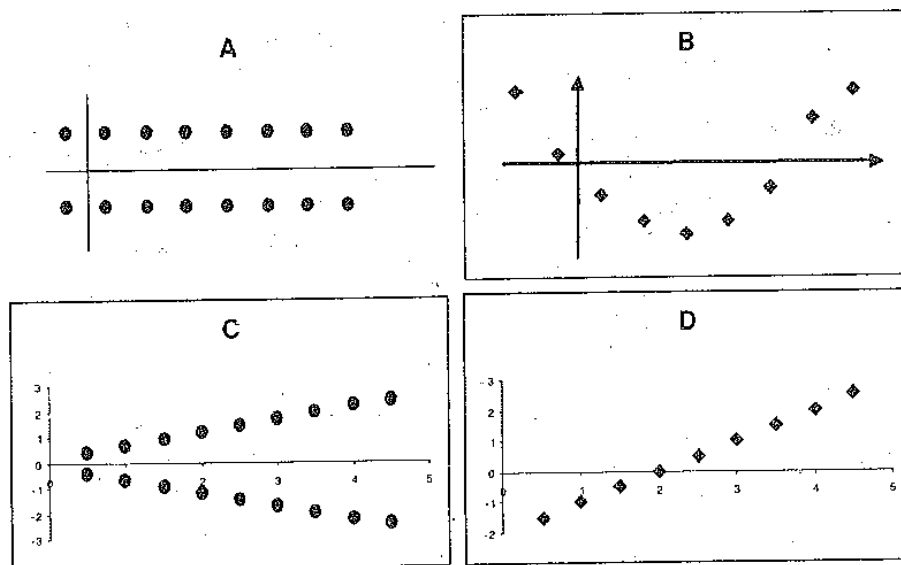


Figura. 2. 23. Gráfica de puntos $\{(\hat{y}_i, \hat{u}_i) | i = 1, 2, 3, 4, \dots, n\}$ y, también, $\{(x_i, \hat{u}_i) | i = 1, 2, 3, 4, \dots, n\}$

La falta de linealidad puede ser por:

- Relación es de forma $g(y) = f(x)$ tal que f y g son funciones. Para alcanzar la linealidad se puede ensayar una transformación de una de las variables o ambas. Además, se puede recurrir a regresión polinomial o no paramétrica.
- Relación parece no lineal por presencia de puntos atípicos o ausencia de variables importantes para explicar Y .

Para solucionar el problema de varianza no es constante (caso C):

- Transformar variable y para estabilizar varianza. Es posible que para conservar la linealidad haya que transformar variable x . En muchas ocasiones la heterocedasticidad está acompañada de falta de linealidad, los dos problemas se solucionan simultáneamente mediante transformación de variables.
- Uso de mínimos cuadrados ponderados.

En gráfico D puede deberse por error en análisis, cálculo o, también, a una mala especificación del modelo. Si variable y es constante respecto a algunos valores de x , los puntos (\hat{y}_i, \hat{u}_i) formarán una recta de pendiente -1 sin que sea, necesariamente, un error en análisis o en cálculos.

Residuos en función de una variable omitida en modelo. Se puede graficar residuos en función de una variable omitida en modelo, pero que puede influir sobre Y . Si datos son cronológicos es conveniente graficar residuos en función del tiempo tal que sería en función del índice. Estos gráficos indicarían:

- Correlación entre errores ϵ tal que cuando hay pocos o muchos cambios de signo, pero serán respaldadas por pruebas de hipótesis adecuadas.

- Varianza del error aumenta o varía con el tiempo, gráfico C, se recurre a mínimos cuadrados ponderados.

Respuesta depende del tiempo, gráficos B o D, se introducirá el tiempo como regresor.

2.10.6. Detección de normalidad

Para estudiar normalidad es conveniente estandarizar residuos. Algunos textos y/o paquetes estadísticos lo hacen simple:

$$d_i = \frac{\hat{u}_i}{\sqrt{S_R^2}}$$

Esta ecuación supone que todos \hat{u}_i tiene la misma varianza (σ^2), que es erróneo:

$$y_i = \hat{y}_i + \hat{u}_i$$

Al estar variables \hat{u}_i, \hat{y}_i no correlacionadas:

$$\text{Var}(y_i) = \text{Var}(\hat{y}_i) + \text{Var}(\hat{u}_i)$$

Tal que:

$$\text{Var}(\hat{u}_i) = \text{Var}(y_i) - \text{Var}(\hat{y}_i)$$

Por igualdad $\text{Var}(\hat{y}_i) = (\sigma^2) \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$:

$$\text{Var}(\hat{u}_i) = (\sigma^2 - \sigma^2) \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right) = \sigma^2(1 - h_{ii}) \text{ tal que } h_{ii} = \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

Tal que para puntos $x_i \approx \bar{x}$:

$$\text{Var}(\hat{u}_i) = (\sigma^2) \left(\frac{n-1}{n} \right) \approx \sigma^2 \text{ con } n \text{ suficientemente grande}$$

Caso contrario, si x_i se aleja demasiado de \bar{x} (punto palanca):

$$\text{Var}(\hat{u}_i) \approx \left(\frac{\sigma^2}{n} \right)$$

Definición. Residuos estandarizados son definidos por:

$$r_i = \frac{\hat{u}_i}{\sqrt{(S_R^2)(1 - h_{ii})}}$$

Tal que $\forall i \ d_i x_i < r_i$. Los residuos d_i subestiman magnitud residual real. Si todos h_{ii} son aproximadamente iguales, no existen h_{ii} muy grandes. Los residuos r_i siguen aproximadamente una ley normal con parámetros (0,1). Para probar normalidad se recurre al gráfico de probabilidad normal (normal probabilit plot).

Gráfico de probabilidad normal. El proceso de construcción del gráfico de probabilidad normal es:

- Se ordenen residuos r_i en orden ascendente.
- Se calcula función de distribución empírica acumulada mediante:

$$\hat{F}_i = \left(\frac{i - 0.5}{n} \right)$$

- Se estiman frac tiles de ley normal centrada y reducida con ordenes \hat{F}_i . Sea $\Phi^{-1}(\hat{F}_i)$ estos órdenes.

- Se grafican puntos $(r_i, \Phi^{-1}(\hat{F}_i))$.

Uso de gráfico de probabilidad normal. Con base en lo anterior, se desprende que si los puntos tienden a formar una línea recta no se rechaza la normalidad. Algunos programas estadísticos añaden una banda de confianza tal que si hay puntos que caen fuera de esta no se acepta la normalidad. Las desviaciones de línea recta sugieren diferentes anormalidades.

Los puntos, ubicados al inicio o al fin del gráfico y se alejan mucho de la recta, generalmente, corresponden a puntos atípicos. Una curvatura S sugiere una ley simétrica, pero con diferentes curtosis, una curva cóncava o convexa sugiere una distribución asimétrica. Un ejemplo con datos simulados es (Normal con parámetros (2,4) –el gráfico de probabilidad normal debería mostrar una recta,

ecuación aproximada $y = \frac{(x-1)}{2}$, uniforme en intervalo (0,1) -pues la ley es simétrica se espera una curva S- y exponencial de parámetro 2 -ley asimétrica se espera una curva convexa o cóncava-):

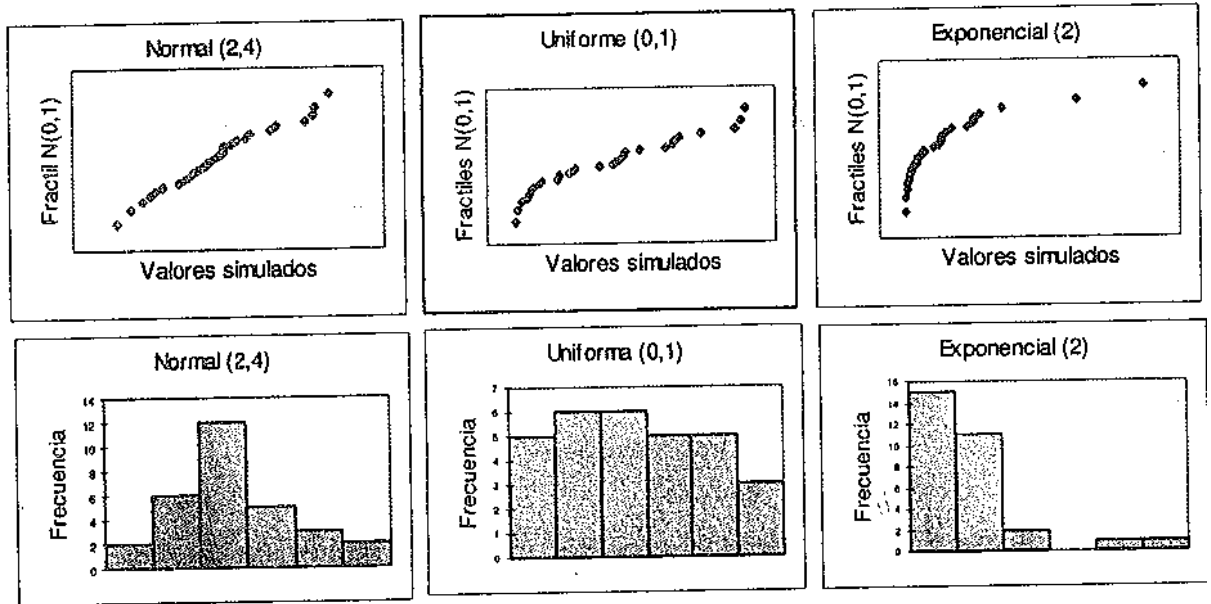


Figura. 2. 24. Ejemplo con datos simulados

Si residuos han sido estandarizados se espera que aproximadamente el 95% de errores estén entre -2 y 2 i 99% entre -3 y 3. Estas proporciones se deben observar en gráficos de dispersión. También, se puede comparar curtosis y Coeficiente de correlación de asimetría. La curtosis de ley normal es 3 y su coeficiente de asimetría es 0. Existen pruebas, como Jarque-Bera, basadas en estos coeficientes.

En mayoría de casos la falta de normalidad se debe a una fuerte asimetría o una curtosis elevada. Estadístico de Jarque-Bera (JB), Carlos M. Jarque Uribe -- y Anil K. Bera --, prueba normalidad de los términos de error y, también, presenta la probabilidad de obtener los estadísticos indicados. Cuanto más alta sea probabilidad de obtener el estadístico JB observado, mayor será la evidencia en favor de la hipótesis nula de que los términos de error están distribuidos normalmente.

Estadístico de Jarque-Bera (JB) se distribuye asintóticamente como una distribución chi cuadrado con dos grados de libertad y puede usarse para probar

la hipótesis nula que datos pertenecen a una distribución normal. La hipótesis nula es una hipótesis conjunta que asimetría y exceso de curtosis son nulos (asimetría = 0 y curtosis = 3). Considera los elementos para probar la normalidad de errores de modelo regresión lineal:

$$y = X\beta + u$$

Tal que $E[u] = 0$, $E[uu'] = \sigma^2$, si $u \sim N(0,1) \Rightarrow \mu_3 = [u_t^3] = 0$; $\mu_4 = [u_t^4] = 3\sigma^4$. La prueba Jarque-Bera (JB) toma el principio “qué tanto se desvían coeficientes de asimetría y curtosis de residuos con coeficientes de asimetría y curtosis de distribución normal”. Se define como:

JB_(Estadístico de Jarque-Bera)

$$= \left[\left(\frac{n(\text{número de observaciones o grados de libertad en general}) - k(\text{número de regresores}) + 1}{6} \right)^2 \left(S_{(\text{Coeficiente de asimetría de muestra})}^2 + \frac{1}{4} (C_{(\text{Coeficiente de curtosis de muestra})} - 3)^2 \right) \right]$$

Es importante mencionar que en todas las pruebas de normalidad se tiene H_0 : ley es normal, H_1 : ley es diferente a normal tal que lo deseable es no rechazar H_0 . Desgraciadamente, todas las pruebas de normalidad son poco potentes por:

- H_a es muy general, no se contrasta contra una ley H_1 en particular.
- En muestras grandes, por teorema del límite central), los residuos tienden a ser normales aún si errores no lo son.

Pruebas se basan en resultados asintóticos tal que se requiere un n gran con respecto a k .

2.11. Otros indicadores de normalidad

Histograma. Es conveniente trazar histograma de residuos y comprar con la curva de función de densidad de una ley normal. Particularmente, mirar la simetría pues la ley normal lo es.

2.11.1. ¿Qué hacer ante falta de normalidad?

Los puntos singulares, no son producto de un error en datos, pueden deberse a una mala especificación del modelo por falta de linealidad o variables no incluidas en modelo. Siempre que existen puntos singulares se estudian para identificar si se deben a variables omitidas en modelo.

Si el histograma de residuos muestra dos o más modas se puede deber a que datos provienen de dos o más poblaciones diferentes, si es el caso se introduce la variable cualitativa población. Sin embargo, se puede construir normalidad transformando variables, como histograma de variable dependiente Y muestra fuerte asimetría hacia valores superiores de la media tal que $\ln(Y)$ puede ser buena transformación.

2.11.2. Puntos singulares y atípicos

Un punto es singular si se aleja “notablemente” del resto de puntos. Puede ser visualizado en el gráfico de dispersión o de preferencia en gráficos de residuos tal que los gráficos de residuos estandarizados son una buena herramienta para su detección. Un punto es atípico si no ha sido generado por el modelo. Un punto puede ser singular por:

- Tiene un valor x_i que se aleja mucho de media \bar{x} tal que se aleja de la gran masa de puntos.
- Su residuo es muy alto, implica más de 3 desviaciones estándar, en valor absoluto.

Para tener una idea más cercana, cuán grande es un residuo, en valor absoluto, se compara con desviación estándar tal que es conveniente trabajar con residuos con desviación estándar de uno. Residuos estandarizados d_i y r_i cumplen este requisito, pero los primeros sub estiman el valor residual tal que se usan los r_i .

2.11.3. Puntos palanca

La pendiente $\hat{\beta}_2$ es un promedio ponderado de pendientes de rectas que pasan por puntos (x_i, y_i) , (\bar{x}, \bar{y}) :

$$\hat{\beta}_2 = \sum_{i=1}^n \left(\frac{(x_i - \bar{x})(y_i - \bar{y})}{S_{xx}} \right) = \sum_{i=1}^n \left(\frac{(x_i - \bar{x})^2}{S_{xx}} \right) \left(\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right) = \sum_{i=1}^n (\rho_i) \left(\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right)$$

El cociente $\left(\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right)$ es pendiente de recta que pasa por puntos (x_i, y_i) , (\bar{x}, \bar{y}) mientras que los pesos $\rho_i = \left(\frac{(x_i - \bar{x})^2}{S_{xx}} \right)$ son números positivos de suma total igual a uno. Si existe un punto x_i suficientemente alejado del resto, su peso ρ_i será próximo a uno y puede ser decisivo para el valor de pendiente, la recta de regresión pasará por (\bar{x}, \bar{y}) tal que estos se llaman puntos palanca.

Si todos los pesos fuesen iguales serían $\rho_i = \left(\frac{1}{n} \right)$ tal que se consideran puntos con alto efecto palanca aquellos que:

$$\rho_i = \left(\frac{(x_i - \bar{x})^2}{S_{xx}} \right) \geq \frac{4}{n} \text{ o } h_{ii} \geq \frac{5}{n}$$

Tal que:

$$\text{Var}(\hat{u}_i) = \sigma^2(1 - h_{ii}) = \sigma^2 \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

Tal que para los puntos con alto valor palanca la varianza del residuo es muy pequeña y, también, el residuo será pequeño en valor absoluto, pues su esperanza es 0. Además, se puede graficar puntos (h_{ii}, r_i) con el objeto de visualizar puntos palanca y puntos con residuo alto en valor absoluto.

2.11.4. Puntos influyentes

Un punto influyente si al eliminarlo implica que valores de estimaciones, desviaciones estándar, razones t y/o R^2 cambian de forma considerable. La influencia de un punto se puede medir con estadístico de Cook. Sea \hat{y}_i predicción

de y en x_i , con todas las observaciones mientras que $\hat{y}_{i(i)}$ será predicción de y en x_i talque cuando se elimina el punto (x_i, y_i) en estimación de recta de regresión.

El estadístico D de Cook es:

$$D_i = \left(\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{2\widehat{\text{Var}}(\hat{y}_i)} \right) = \left(\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{2s_R^2 h_{ii}} \right) = \left(\frac{r_i^2}{2} \right) \left(\frac{h_{ii}}{1 - h_{ii}} \right)$$

En teorema de Gauss-Markov se demuestra $\hat{U} = (I - H)U$ tal que $\hat{u}_i = u_i - \sum_{j=1}^n h_{ij}u_j$ tal que $\hat{u}_i \approx u_i$ ssi $\sum_{j=1}^n h_{ij}u_j \approx 0$ pues $E(\hat{u}_i) = 0$ y $\text{Var}(\sum_{j=1}^n h_{ij}u_j) = (\sigma^2)(\sum_{j=1}^n h_{ij}^2) = \sigma^2 h_{ii} \Rightarrow \hat{u}_i \approx u_i$ ssi $h_{ii} \approx 0$, el residuo \hat{u}_i es buen indicador del error u_i ssi $h_{ii} \approx 0$ tal que se puede demostrar que $\frac{1}{n} \leq h_{ii} < 1$ tal que el residuo es informativo sólo cuando $h_{ii} \approx \frac{1}{n}$ y si n es grande $h_{ii} \approx 0$. Si $h_{ii} \approx 1$, por teorema bajo hipótesis H , si matriz X es de rango completo y no aleatoria: $E[\hat{U}] = 0$ y $\text{Var}[\hat{U}] = \sigma^2(I - H)$, $\text{Var}(\hat{U}) = (\sigma^2)(I - H)$ donde $H = XX^t(X^tX)^{-1} \Rightarrow \forall_i = \text{Var}(\hat{u}_i) = \sigma^2(1 - h_{ii})$, $\text{Var}(\hat{u}_i) \approx 0 \Rightarrow \hat{u}_i \approx 0$ o, equivalente a, $y_i \approx \hat{y}_i$. La superficie de regresión pasaría muy cerca de punto (x_i, y_i) sin tomar en cuenta valor y_i ; es decir, es punto palanca.

Estos afectan a normalidad de errores, pero pueden tener influencia alta o baja en estimaciones de parámetros tal que para constatar su influencia se elimina el punto y se estiman nuevamente parámetros, si hay cambios importantes indica que es un punto atípico influyente y requiere tratamiento “especial”.

La influencia de un punto cualquiera se mide mediante estadístico de Cook: sea \hat{y}_i predicción de y en $x_i = (1, x_{i2}, x_{i3}, x_{i4}, \dots, x_{ik})^t$, con todas las observaciones mientras $\hat{y}_{i(i)}$ predicción de y en x_i , cuando se elimina (x_i, y_i) en estimación de recta de regresión. El estadístico de Cook es:

$$D_i = \left[\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{k(\widehat{\text{Var}}(\hat{y}_i))} \right] = \left[\frac{(\hat{y}_i - \hat{y}_{i(i)})^2}{k(s_R^2)(h_{ii})} \right] = \left[\left(\frac{r_i^2}{k} \right) \left(\frac{h_{ii}}{1 - h_{ii}} \right) \right]$$

Sin embargo, no existen reglas precisas para decidir qué puntos son influyentes, pero se considera que un punto lo es moderadamente si $D_i > 0.5$ e influyente si $D_i > 1.0$, equivalente a:

$$\frac{|\hat{y}_i - \hat{y}_i(i)|}{\sqrt{\text{Var}(\hat{y}_i)}} > 1, \frac{|\hat{y}_i - \hat{y}_i(i)|}{\sqrt{\text{Var}(\hat{y}_i)}} > \sqrt{2}$$

La discrepancia estandarizada es superior a 1 en primer caso y superior a $\sqrt{2}$ en el segundo. Un punto palanca puede ser influyente o no. Su influencia se mide con estadístico D. Si un punto es influyente existes dos opciones:

- El modelo es correcto y el punto es producto de un error en datos.
- No hay error en datos, pero el modelo está mal especificado por:
 - Existe tercera variable, no incluida en modelo, que produce el cambio de comportamiento.
 - Si a más de influyente es palanca, la relación no es lineal o varianza no es constante.

Un punto atípico puede ser producto de un error en datos, producto de factores no tomados en cuenta por el investigador o datos con características particulares tal que pueden contener información muy valiosa, su justificación y conocimiento puede constituirse en resultado más importante de la investigación. Grandes descubrimientos son producto de datos atípicos, como un ingeniero que investiga modos de obtener poli fenoles con base en especies de maíz negro autóctono de x lugar, pero encontró otra sustancia más efectiva que contrarresta el cáncer en consumidores regulares.

¿Un punto singular influyente del que no se conoce la causa sería recomendable eliminarlo? La respuesta es no necesariamente, pues depende el nivel de su influencia. Se calculan los residuos estudentizados mediante:

$$\hat{t}_i = \frac{\hat{u}_i}{\sqrt{(S_{R(i)}^2)(1 - h_{ii})}}$$

Tal que $S_{R(i)}^2$ (Suma de Cuadrados por error puro) estimada sin incluir el punto i. Si $|\hat{t}_i|$ es mayor que valor de tabla, se recomienda eliminarlo:

Tabla. 2. 8. Un punto atípico puede ser producto de un error en datos, de factores no tomados en cuenta por el investigador o datos con características particulares

n	20	30	40	50	60	70	80	100
Fractil	3.14	3.25	3.30	3.32	3.38	3.42	3.46	3.51

Si desea no eliminar estos puntos se puede introducir variables indicatrices que los identifiquen. También, se puede recurrir a mínimos cuadrados ponderados tal que a estos puntos se les asigna un peso bajo.

2.12. Nomenclatura matricial de ANOVA

Según (Grossman & Flores, 2012), $V_{(\text{Espacio Vectorial})} = \mathbb{R}^n = \left\{ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{matrix} : X_j \in \mathbb{R} \text{ para } j = 1, 2, 3, 4 \dots n \right\}$ tal que cada vector \mathbb{R}^n es una matriz $n \times 1$,

siendo \mathbb{R}^n e símbolo que denota al conjunto de todos los vectores de dimensión n :

$\left\{ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{matrix} \right\}$ donde cada a_i es un número real (\mathbb{R}). Según la definición de matrices, una

matriz A de $m \times n$ es un arreglo rectangular de mn números dispuestos en m

renglones y n columnas tal que $A = \left\{ \begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \dots & a_{2j} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \dots & a_{3j} & \dots & a_{3n} \\ a_{41} & a_{42} & a_{43} & a_{44} & \dots & a_{4j} & \dots & a_{4n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & a_{i4} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} & \dots & a_{mj} & \dots & a_{mn} \end{matrix} \right\}$,

$\mathbf{x} + \mathbf{y}$ es una matriz de $n \times 1$ si \mathbf{x} y \mathbf{y} son matrices $n \times 1$. Haciendo $\mathbf{0} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$ y $-\mathbf{x} =$

$\begin{Bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{Bmatrix}$, 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 $m \times n$ y sean $\alpha - \beta$ dos escalares implica los siguientes puntos:

- viii) $A + \mathbf{0} = A$
- ix) $\mathbf{0}A = \mathbf{0}$
- x) $A + B = B + A$ (ley conmutativa para suma de matrices)
- xi) $(A + B) + C = A + (B + C)$ (ley asociativa para suma de matrices)
- xii) $\alpha(A + B) = \alpha A + \alpha B$ (ley distributiva para multiplicación por un escalar)
- xiii) $1A = A$
- xiv) $(\alpha + \beta)A = \alpha A + \beta A$

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:

$$\left. \begin{array}{l} 1. \vec{\mu} + \vec{v} = \vec{v} + \vec{\mu} \\ 2. \vec{\mu} + (\vec{v} + \vec{w}) = (\vec{\mu} + \vec{v}) + \vec{w} \\ 3. \vec{w} + \vec{0} = \vec{w} \\ 4. \vec{\mu} + (-\vec{\mu}) = \vec{0} \end{array} \right\} \forall \vec{\mu}, \vec{v}, \vec{w} \in \mathbb{R}^2$$

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

Se llama espacio vectorial V sobre \mathbb{R} a la terna $(V, +, *) \Rightarrow V = \{\vec{\mu}, \vec{v}, \vec{w}, \dots\}$ donde sus elementos son vectores y se cumplen los 8 axiomas anteriores. Entonces, un sub espacio vectorial $(H, +, *) \subset (V, +, *)$ tal que H es un sub espacio vectorial de V si H es un subconjunto no vacío 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 V .

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 $\vec{v} \in H \Rightarrow \vec{\mu} + \vec{v} \in H$ 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 $H \subset V \Leftrightarrow \left. \begin{array}{l} \vec{\mu}, \vec{v} \in H \\ \alpha, \beta \in \mathbb{R} \end{array} \right\} \Rightarrow (\alpha\vec{\mu} + \beta\vec{v}) \in H$; 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 $(0, 0)$.

Además, combinación lineal se demuestra a partir que sean $v_1, v_2, v_3, v_4, \dots, v_n$ vectores en un espacio vectorial V tal que cualquier vector de la forma $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$ se denomina **combinación lineal de $v_1, v_2, v_3, v_4, \dots, v_n$** , 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 $v_1, v_2, v_3, v_4, \dots, v_n$ 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, $\forall v \in V_{(\text{Espacio Vectorial})} \exists$ escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \dots, \alpha_n \Rightarrow v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_n v_n$.

Además, un espacio generado por un conjunto de vectores se genera a partir que sea $v_1, v_2, v_3, v_4, \dots, v_k$ k vectores de un espacio vectorial V . El espacio generado por $\{v_1, v_2, v_3, v_4, \dots, v_k\}$ es el conjunto de combinaciones lineales $v_1, v_2, v_3, v_4, \dots, v_k$; es decir, $\text{gen}\{v_1, v_2, v_3, v_4, \dots, v_k\} = \{v | v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 \dots + \alpha_k v_k\}$.

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:

- xi) Si $X \in V$ y $Y \in V$, entonces $X + Y \in V$ (**cerradura bajo suma**)
- xii) Para $\forall x, y$ y z en V , $(x + y) + z = x + (y + z)$ (**ley asociativa de suma de vectores**)
- xiii) \in un vector $0 \in V \Rightarrow \forall X \in V, X + 0 = 0 + X = X$ (**0 es llamado vector cero o idéntico aditivo**)
- xiv) Si $X \in V$, existe un vector $-X$ en $\in V$ tal que $X + (-X) = 0$ (**-x se llama inverso aditivo de X**)
- xv) Si X y Y están en V , entonces $X + Y = Y + X$ (**Ley conmutativa de suma de vectores**)
- xvi) Si $X \in V$ y α es una escalar, entonces $\alpha X \in V$ (**cerradura bajo multiplicación por un escalar**)
- xvii) Si X y Y están en V y α es una escalar, entonces $\alpha(X + Y) = \alpha X + \alpha Y$ (**primera ley distributiva**)
- xviii) Si $X \in V$ y $\alpha - \beta$ son escalares, entonces $(\alpha + \beta)X = \alpha X + \beta X$ (**segunda ley distributiva**)
- xix) Si $X \in V$ y $\alpha - \beta$ son escalares, entonces $\alpha(\beta X) = (\alpha\beta)X$ (**ley asociativa de multiplicación por escalares**)
- xx) Para \forall vector $X \in V, 1X = X$

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

Demostración:

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

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

Por lo tanto, \forall vector en V es vector en \mathbb{R}^2 y \mathbb{R}^2 es V , como se muestra enseguida:

$$V_{(\text{Espacio Vectorial})} = \mathbb{R}^n = \left\{ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_n \end{matrix} : X_j \in \mathbb{R} \text{ para } j = 1, 2, 3, 4 \dots n \right\} \text{ tal que cada vector } \mathbb{R}^n$$

es una matriz $n \times 1$, siendo \mathbb{R}^n e símbolo que denota al conjunto de todos los

vectores de dimensión n : $\left\{ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{matrix} \right\}$ 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 $V \Rightarrow (X_1, Y_1, Z_1) + (X_2, Y_2, Z_2) = (X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2) \in V$ 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 $\Pi \mathbb{R}^3$ 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 $V = \{(X, Y): Y = 2X + 1, X \in V\}$, 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 $X \in V$ y $Y \in V$, entonces $X + Y \in V$).

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 $Y_2 = 2X_2 + 1$, de manera que $Y_1 + Y_2 = (2X_1 + 1) + (2X_2 + 1) = 2X_1 + 2X_2 + 2$. Por lo tanto, $(X_1 + X_2, Y_1 + Y_2) \notin V \Leftrightarrow (X_1, Y_1) \in V$ y $(X_2, Y_2) \in V$.

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 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 $p(X) = a_n X^n + a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + a_{n-3} X^{n-3} + a_{n-4} X^{n-4} + \dots + a_1 X + a_0 \forall a_i \in \mathbb{R}$. La suma de $p(X) + q(X)$

está definida usualmente, si $q(X) = b_n X^n + b_{n-1} X^{n-1} + b_{n-2} X^{n-2} + b_{n-3} X^{n-3} + b_{n-4} X^{n-4} + \dots + b_1 X + b_0 \forall b_j \in \mathbb{R} \Rightarrow p(X) + q(X) = (a_n + b_n)X^n + (a_{n-1} + b_{n-1})X^{n-1} + (a_{n-2} + b_{n-2})X^{n-2} + (a_{n-3} + b_{n-3})X^{n-3} + (a_{n-4} + b_{n-4})X^{n-4} + \dots + (a_1 + b_1)X + (a_0 + b_0)$ 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 $\mathbf{0} = 0X^n + 0X^{n-1} + 0X^{n-2} + 0X^{n-3} + 0X^{n-4} + \dots + 0X + 0 \Rightarrow 0 \in P_n$ 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 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 $m \times n$ 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 $m \times n$. 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 3×3 . 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 $m \times n \Rightarrow AB$ 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 puede 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 $B(AC) = (AB)C \Rightarrow B = BI = (AB)C = IC = C \Rightarrow B = C$ 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 $m \times n$, $B = (b_{ij})$ matriz de $m \times p$ y $C = (c_{ij})$ matriz de $p \times q$, 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 $n \times q$.

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:

- v) $\alpha \mathbf{0} = \mathbf{0}$ para $\forall \alpha_{(\text{Escalar})}$
- vi) $\mathbf{0} * \mathbf{X} = \mathbf{0}$ para $\forall \mathbf{X} \in V$
- vii) $\alpha \mathbf{X} = \mathbf{0} \Rightarrow \alpha = 0, \mathbf{X} = \mathbf{0}$ o ambos
- viii) $(-1)\mathbf{X} = -\mathbf{X}$ para $\forall \mathbf{X} \in V$

Demostración:

- v) 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 $\alpha\mathbf{0} + (-\alpha\mathbf{0}) = [\alpha\mathbf{0} + \alpha\mathbf{0}] + (-\alpha\mathbf{0}) \Rightarrow \mathbf{0} = (\alpha\mathbf{0}) + [\alpha\mathbf{0} + (-\alpha\mathbf{0})] = \alpha\mathbf{0} + \mathbf{0} = \alpha\mathbf{0}$.

vi) Esencialmente, se usa 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 $0\mathbf{X} + (-0\mathbf{X}) = 0\mathbf{X} + [0\mathbf{X} + (-0\mathbf{X})] = \mathbf{0} = 0\mathbf{X} + \mathbf{0} = 0\mathbf{X}$

vii) 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 $\mathbf{X} = \mathbf{0}$

viii) 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.

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**:

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

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

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), $(-1)\mathbf{X} \in H$ 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.

2.12.1. Análisis de varianza *ANOVA*

El procedimiento del análisis de la varianza (ANOVA) es una metodología estadística que permite la comparación de dos o más medias poblaciones midiendo la variación dentro de las muestras. En nuestro caso emplearemos el análisis de la varianza para realizar pruebas sobre los parámetros estimados del modelo y así conocer su exactitud.

Anteriormente establecimos que se cumple la siguiente igualdad:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{t=1}^n (y_t - \tilde{y}_t)^2 + \sum_{L=1}^n (\tilde{y}_L - \bar{y})^2$$

Significa que la suma de los cuadrados corregida es igual a la suma de cuadrados de los errores más la suma de cuadrados debida a la regresión. La suma de los cuadrados corregida tiene $(n - 1)$ grados de libertad, la suma de los cuadrados de los errores tiene $(n - 2)$ grados de libertad y la suma de los cuadrados debida a la regresión tiene 1 grado de libertad. Es decir, la igualdad correspondiente a los grados de libertad de la ecuación es:

$$n - 1 = (n - 2) + 1$$

Con base en ecuaciones anteriores, se tiene la tabla de análisis de la varianza:

Tabla. 2. 9. Análisis de varianza (ANOVA)

Tabla de Análisis de la Varianza				
Fuente de variación	Grados de libertad (g.l.)	Suma de Cuadrados (SC)	Cuadrado medio (MC)	F
Regresión Error o residual	$n - 2$	SCR o SCE	$MCR = \frac{SCR}{1}$ $s^2 = SCE/(n - 2)$	$F_{(Calculada)} = \frac{MCR}{s^2}$
Total corregido	$n - 1$	SC_{yy}		

La columna del «cuadrado medio» (MC) se obtiene al dividir cada una de las sumas de los cuadrados entre sus correspondientes grados de libertad. El valor de $F_{(Calculada)}$ resulta de la división del cuadrado medio de regresión para el cuadrado medio residual $F_{(Calculada)} = \frac{MCR}{s^2}$. Una vez elaborada la tabla de análisis de varianza, el valor de $F_{(Calculada)}$ se emplea para realizar una prueba de hipótesis sobre la razón de dos varianzas, que sirve para probar si $\beta_1 = 0$. La prueba es la siguiente:

1. Hipótesis Nula. $H_0: \beta_1 = 0$.
2. Hipótesis Alternativa. $H_1: \beta_1 \neq 0$.
3. Estadístico de Prueba. $F_{(Calculada)} = \frac{MCR}{s^2}$.

Región de rechazo. No se acepta H_0 si $F_{(Calculada)} > F_{(Tablas;(1,n-2))}$.

2.12.2. Transformaciones de variables

En la práctica de regresión frecuentemente se halla con el problema que la relación entre variables X, Y no es lineal y/o no es homocedástica o la distribución de residuos no es simétrica tal que los errores no provienen de una distribución normal. Muchas ocasiones, estos problemas pueden solucionarse mediante transformación de variables. Se busca una relación de forma:

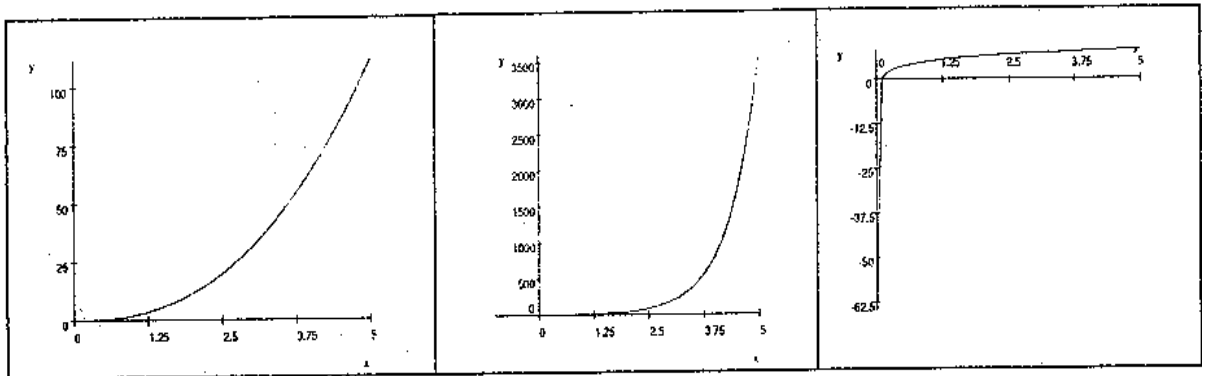
$$y'_i = \beta_1 + \beta_2 x'_i + u_i$$

Con $y'_i = g(y_i)$, $x'_i = f(x_i)$ donde g y f son funciones conocidas.

2.12.3. Usuales para linealizar relación

Tabla. 2. 10. Ecuaciones usuales para linealizar una relación

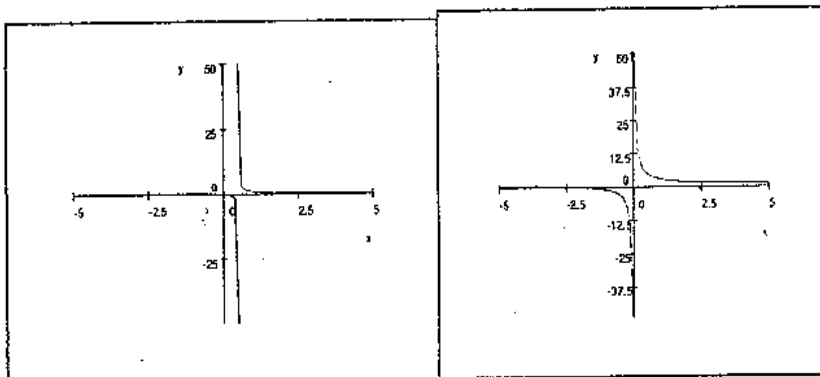
Gráfico	Función	Transformación
A	$y = \alpha x^{\beta_2}$	$y' = \text{Ln}(y), x' = \text{Ln}(x)$
B	$y = \alpha x^{\beta_2 x}$	$y' = \text{Ln}(y)$
C	$y = \beta_1 + \beta_2 \text{Ln}(x)$	$x' = \text{Ln}(x)$
D	$y = \left(\frac{x}{\beta_1 x + \beta_2} \right)$	$y' = \left(\frac{1}{y} \right), x' = \left(\frac{1}{x} \right)$
E	$y = \beta_1 + \left(\frac{\beta_1}{x} \right)$	$x' = \left(\frac{1}{x} \right)$



A

B

C



D

E

Figura. 2. 25. Gráfica de ecuaciones usuales para linealizar una relación

Interpretación

En modelo $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2(x)$ tal que $\hat{\beta}_2$ indica la magnitud de cambio de X respecto a Y; es decir, si X aumenta o disminuye una unidad Y también lo hará en la misma magnitud. También, esto puede expresarse como:

$$d(\hat{y}) = \hat{\beta}_2(dx)$$

Tal que si considera individualmente a cada uno de los modelos:

$$\text{Ln}(\hat{y}) = \hat{\beta}_1 + \hat{\beta}_2(x), \hat{y} = \hat{\beta}_1 + \hat{\beta}_2(\text{Ln}[x]), \text{Ln}(\hat{y}) = \hat{\beta}_1 + \hat{\beta}_2 \text{Ln}[x]$$

La cuestión es si todos los casos son equivalentes y cómo varía X respecto a Y. El modelo $\text{Ln}(\hat{y}) = \hat{\beta}_1 + \hat{\beta}_2(x)$ si:

$$\text{Ln}(\hat{y}_1) = \hat{\beta}_1 + \hat{\beta}_2(x) \quad \text{Ln}(\hat{y}_2) = \hat{\beta}_1 + \hat{\beta}_2(x + 1)$$

Tal que:

$$\hat{\beta}_2 = \text{Ln}(\hat{y}_2) - \text{Ln}(\hat{y}_1) = \text{Ln}\left(\frac{\hat{y}_2}{\hat{y}_1}\right) = \text{Ln}\left(1 + \frac{\hat{y}_2 - \hat{y}_1}{\hat{y}_1}\right)$$

Con base en desarrollo de Taylor, si $\left(\frac{\hat{y}_2 - \hat{y}_1}{\hat{y}_1}\right) \rightarrow 0$:

$$\text{Ln}\left(1 + \frac{\hat{y}_2 - \hat{y}_1}{\hat{y}_1}\right) \approx \left(\frac{\hat{y}_2 - \hat{y}_1}{\hat{y}_1}\right)$$

Tal que $\hat{\beta}_2$ es tasa de variación de \bar{Y} . Puede afirmarse que si X incrementa o decrementa en una unidad, Y lo hace en promedio $100\% \hat{\beta}_2$ unidades; por ejemplo: si $\hat{\beta}_2 = 0.15$ el incremento promedio de Y = 15%/Unidad de incremento de X. También, se puede calcular mediante:

$$\left(\frac{dy}{dx}\right) = e^{(\hat{\beta}_1 + \hat{\beta}_2(x))\hat{\beta}_2 = y\hat{\beta}_2} = \left(\frac{dy}{y}\right) = \hat{\beta}_2(dx)$$

Modelo. $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2(\text{Ln}[x])$ tal que suponga que X cambia de x_1 a x_2 mientras que \hat{Y} cambia de \hat{y}_1 a \hat{y}_2 tal que:

$$\hat{y}_1 = \hat{\beta}_1 + \hat{\beta}_2(\text{Ln}[x_1]) \quad \hat{y}_2 = \hat{\beta}_1 + \hat{\beta}_2(\text{Ln}[x_2])$$

Tal que, el incremento \bar{Y} será:

$$\hat{y}_2 - \hat{y}_1 = \hat{\beta}_2(\text{Ln}[x_2] - \text{Ln}[x_1]) = \hat{\beta}_2 \text{Ln}\left(\frac{x_2}{x_1}\right) = \hat{\beta}_2 \text{Ln}\left(1 + \frac{x_2 - x_1}{x_1}\right) \approx \hat{\beta}_2 \left(\frac{x_2 - x_1}{x_1}\right)$$

Si la variación de X es 1% $\left(\frac{x_2 - x_1}{x_1} = 0.01\right)$, el incremento o decremento de $\bar{Y} = 0.01\hat{\beta}_2$ unidades.

Por ejemplo: si $\hat{\beta}_2 = -15.00$, el decremento de $\bar{Y} = 0.15$ unidades tal que su derivada será:

$$\left(\frac{d\hat{y}}{dx}\right) = \left(\frac{\hat{\beta}_2}{x}\right) \text{ o } d(\hat{y}) = \hat{\beta}_2 \left(\frac{dx}{x}\right)$$

Modelo. $\text{Ln}(\hat{y}) = \hat{\beta}_1 + \hat{\beta}_2 \text{Ln}[x]$ tal que, como en caso pasado:

$$\text{Ln}(\hat{y}_2) - \text{Ln}(\hat{y}_1) = \hat{\beta}_2 [\text{Ln}(\hat{y}_2) - \text{Ln}(\hat{y}_1)]$$

$$\text{Ln}\left(\frac{y_2}{y_1}\right) = \hat{\beta}_2 \text{Ln}\left(\frac{x_2}{x_1}\right)$$

$$\text{Ln}\left(1 + \frac{y_2 - y_1}{y_1}\right) = \hat{\beta}_2 \text{Ln}\left(1 + \frac{x_2 - x_1}{x_1}\right)$$

$$\left(\frac{y_2 - y_1}{y_1}\right) \approx \hat{\beta}_2 \left(\frac{x_2 - x_1}{x_1}\right)$$

Tal que, a una variación porcentual de X corresponde a una variación porcentual de Y. En consecuencia, si X incrementa en 1% la variación media de Y es $\hat{\beta}_2$ %; por ejemplo: si $\hat{\beta}_2 = 2$ indica si X incrementa en 1% implica que Y aumenta en 2%. Su derivada es:

$$\left(\frac{d\hat{y}}{\hat{y}}\right) = \hat{\beta}_2 \left(\frac{dx}{x}\right)$$

En econometría $\hat{\beta}_2$ es elasticidad media de Y respecto a X. Por ejemplo: Modelo Cobb-Douglas señala que cantidad de producción (Q_t) depende de mano de obra o cantidad de trabajo (N_i) según:

$$\text{Ln}(Q_t) = \beta_1 + \beta_2 \text{Ln}(N_i) + u_i$$

Se sabe que $\hat{\beta}_2$ es elasticidad de producción respecto a trabajo o elasticidad-trabajo de producción tal que si trabajo se incrementa en 1% implica que producción será en promedio, $\hat{\beta}_2 > 0$, $\hat{\beta}_2$ %.

2.12.4. Homocedasticidad

Un grupo de transformaciones que sirven para linealizar el modelo como para estabilizar la varianza son las transformaciones de Box – Cox. Transformaciones Box-Cox. La familia de transformaciones Box – Cox es:

$$y^{(\lambda)} = \begin{cases} \left(\frac{y^{(\lambda)} - 1}{\lambda} \right) & \text{Si } \lambda \neq 0 \\ \text{Ln}(y) & \text{Si } \lambda = 0 \end{cases}$$

Casos atípicos son:

Tabla. 2. 11. Casos atípicos de transformación

λ (Letra griega Lambda)	Nombre	Transformación
-1	Inversa o recíproca	$y^{(-1)} = 1 - \left(\frac{1}{y}\right)$
$-\frac{1}{2}$	Inversa de la raíz	$y^{(-\frac{1}{2})} = 2 - \left(\frac{2}{\sqrt{y}}\right)$
0	Logarítmica	$\text{Ln}(y)$
$\frac{1}{2}$	Raíz cuadrada	$y^{(\frac{1}{2})} = 2\sqrt{y} - 2$
1	—	Transformación \notin

Para determinación aproximada de λ es:

- Si varianza crece con x , se ordenen puntos (x_i, y_i) en función de primera componente. Si varianza crece con \hat{y} , se ordenan observaciones y_i en función de \hat{y} .
- Se forma grupos de 4 o 5 observaciones y_i .
- Para cada grupo se calcula la media \bar{y}_k y desviación estándar s_k . Además, en pocas observaciones s_k es, aproximadamente, la amplitud.
- Se gráfica puntos (\bar{y}_k, s_k) .
- Si puntos se distribuyen al azar en una franja horizontal, la varianza es constante y no requiere transformación. Si no, si existe alguna tendencia se “estima visualmente” el parámetro α de relación $s_k = c\bar{y}_k^\alpha$ tal que $y = 1 - \alpha$.

Si varianza crece con una función $g(x)$; es decir, si $\text{Var}(u_i) = \sigma^2 g(x_i)$, el modelo se estima:

$$\left(\frac{y_i}{\sqrt{g(x_i)}} \right) = \left(\beta_1 * \frac{1}{\sqrt{g(x_i)}} \right) + \left(\beta_2 * \frac{x_i}{\sqrt{g(x_i)}} \right) + \left(\frac{u_i}{\sqrt{g(x_i)}} \right)$$

Donde, la varianza del nuevo error es:

$$\varepsilon_i = \frac{u_i}{\sqrt{g(x_i)}}$$

Es σ^2 para $\forall i$. Un caso frecuente es $g(x) = x^2$, el modelo que se estima es:

$$\left(\frac{x_i}{y_i} \right) = \beta_2 + \beta_1 \left(\frac{1}{x_i} \right)$$

Tal que, la pendiente del modelo original es el intercepto.

BIBLIOGRAFÍA

Castro, B. A. (2008). *Regresión lineal. Monografías de matemática y estadística*. Quito, Ecuador: Facultad de Ciencias. Departamento de Matemáticas. Escuela Politécnica Nacional.

Galindo, E. (2006). *Estadística Métodos y Aplicaciones*. Distrito Federal, México: Prociencia Editores.

González, B. G. (1985). *Métodos Estadísticos y Diseño Experimental*. Quito, Ecuador: Editorial Universitaria. Universidad Central del Ecuador.

Grossman, S. S., & Flores, G. J. (2012). *Álgebra Lineal*. Colonia Desarrollo Santa Fé. Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores S. A. de C. V.

Gujarati, D. N., & Porter, D. C. (2010). *Econometría*. Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores S. A. de C. V.

Lind, D. A., Marchal, W. G., & Mason, R. D. (2006). *Estadística para Administración y Economía*. Col. Valle. México, D. F.: Alfaomega Grupo Editor S. A. de C. V.

Navidi, W. (2006). *Estadística para Ingenieros*. Colonia Desarrollo Santa Fe, Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores, S.A. DE C.V.

Triola, M. F. (2004). *Estadística*. Naucalpan de Juárez, estado de México. México: Pearson Educación de México S. A. de C. V.

Wackerly, D. D., Mendehall, W. I., & Scheaffer, R. L. (2010). *Estadística Matemática con Aplicaciones*. Col. Cruz Manca, Santa Fe. México, D.F.: Achen, C. H. (1982). *Interpreting and using regression*. London: Sage.

- Aiken, L., AND West, S. (1991). *Multiple regression: Testing and interpreting Interactions*. London: Sage
- Amon, J. (1990). *Estadística para psicólogos (1)*. Estadística Descriptiva. Madrid: Pirámide.
- Amon, J. (1990). *Estadística para psicólogos (2)*. Probabilidad. Estadística Inferencial. Madrid: Pirámide.
- Castro, B. A. (2008). *Regresión lineal. Monografías de matemática y estadística*. Quito, Ecuador: Facultad de Ciencias. Departamento de Matemáticas. Escuela Politécnica Nacional.
- D.G. Kleinbaum, L.L. Kupper, K.E. Muller. *Applied Regression Analysis and Other Multivariable Methods*. PWS-KENT Publishing Company. 1988.
- Galindo, E. (2006). *Estadística Métodos y Aplicaciones*. Distrito Federal, México: Prociencia Editores.
- González, B. G. (1985). *Métodos Estadísticos y Diseño Experimental*. Quito, Ecuador: Editorial Universitaria. Universidad Central del Ecuador.
- Grossman, S. S., & Flores, G. J. (2012). *Álgebra Lineal*. Colonia Desarrollo Santa Fé. Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores S. A. de C. V.
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría*. Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores S. A. de C. V.
- Jaccard, J., Lee Teitel, Turrisi, R., Wan, C. (1990). *Interaction effects in multiple regression*. Sage University Paper series on Quantitative Applications in the Social Sciences. Newbury Park, CA:Sage
- James, L. R. (1982). *Causal analysis: assumptions, models and data*. Beverly Hills: Sage.
- Jañez, L. (1980). *Fundamentos de psicología matemática*. Madrid: universidad Complutense.
- Lind, D. A., Marchal, W. G., & Mason, R. D. (2006). *Estadística para Administración y Economía*. Col. Valle. México, D. F.: Alfaomega Grupo Editor S. A. de C. V.

- Navidi, W. (2006). *Estadística para Ingenieros*. Colonia Desarrollo Santa Fe, Delegación Álvaro Obregón. México, D. F.: McGraw-Hill/Interamericana Editores, S.A. DE C.V.
- Olmo, J. y Frias, D. (2000). «Capítulo 6. Métodos de dependencia. Regresión Lineal». En *Técnicas de análisis de datos en investigación de mercados*, 247-80. Madrid: Ediciones Pirámide.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research. Explanation and prediction* (2nd ed.). New York: Halt, Rinehart and Winston.
- Peña, D. (1987).: *Estadística, modelos y métodos*. 2. Modelos lineales y series temporales Alianza Universidad.
- Sánchez Vizcaino, G. (2000). «Capítulo 10. Métodos de dependencia. Regresión Logística». En *Técnicas de análisis de datos en investigación de mercados*, 431-67. Madrid: Ediciones Pirámide.
- Shoeder et al. (1982). *Understanding regression analysis: an introductory guide*. Beverly Hills: Sage.
- Triola, M. F. (2004). *Estadística*. Naucalpan de Juárez, estado de México. México: Pearson Educación de México S. A. de C. V.
- V. Abaira, A. Pérez de Vargas. *Métodos Multivariantes en Bioestadística*. Ed. Centro de Estudios Ramón Areces. 1996.
- Wackerly, D. D., Mendehall, W. I., & Scheaffer, R. L. (2010). *Estadística Matemática con Aplicaciones*. Col. Cruz Manca, Santa Fe. México, D.F.: Cengage Learning Editores, S. A. de C. V., una Compañía de Cengage Learning, Inc.
- Wonnacott, T. H. and Wonnacott, R. J. (1981). *Regression: a second course in statistics*. New York: Wiley.
- Cengage Learning Editores, S. A. de C. V., una Compañía de Cengage Learning, Inc.