



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**Mínimos cuadrados: orígenes,
fundamentos, alternancia, extensión y
aplicación**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

Gerardo Quintero Pérez



**DIRECTOR DE TESIS:
Mat. Margarita E. Chávez Cano**

2013



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Datos del alumno:

Gerardo

Quintero

Pérez

Teléfono: 21-57-06-85

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

Número de cuenta: 30519360-7

Datos del tutor:

Matemática

Margarita Elvira

Chávez

Cano

Datos del sinodal 1:

Doctora

María del Pilar

Alonso

Reyes

Datos del sinodal 2:

Maestro en Ciencias

José Antonio

Flores

Díaz

Datos del sinodal 3:

Doctora

Ruth Selene

Fuentes

García

Datos del sinodal 4:

Maestro

Alejandro

Mina

Valdés

Datos del trabajo escrito:

Título: *Mínimos cuadrados: orígenes, fundamentos, alternancia, extensión y aplicación*

Número de páginas: 232

Año: 2013

Agradecimientos

Profesora Margarita:

Me siento muy agradecido con usted por toda su paciencia y confianza que ha depositado en mí y en este trabajo. Gracias por su empeño en el momento de estructurarlo y corregirlo pese a todas las actividades que tiene de por sí a su cargo dentro de nuestra amada Facultad.

Gracias por todas sus enseñanzas durante los cursos en que tuve el privilegio de ser su alumno durante la carrera, por abrirme las puertas y dado la oportunidad de conocerla no sólo como profesora, sino también como persona. Nuevamente gracias por todo, le estaré siempre agradecido.

Facultad de Ciencias:

Gracias por permitirme entrar a tus salones y ser una fuente de conocimiento inagotable. Gracias a tí he conocido a grandes académicos comprometidos contigo y con nuestra máxima casa de estudios que enorgullecen a la comunidad científica de este país. De todos ellos he adquirido las mejores enseñanzas. También gracias a tí he conocido a grandes amigos con los he compartido los mejores años de mi vida. Contento y satisfecho, te dejo este trabajo para demostrarte que como profesional y persona me has formado.

A mis sinodales:

Aún sin conocerlos a algunos de ustedes, me mostraron confianza y calidez humana desde el primer momento. Gracias por el valioso tiempo que se tomaron para leer y corregir este trabajo. Estoy convencido de toda su capacidad y compromiso con nosotros. Gracias por toda su dedicación y apoyo.

A mi madre:

Estoy consciente de todo el esfuerzo que tuviste que hacer en los momentos más difíciles, y de todo lo que pasaste a causa de mis problemas de salud aun siendo muy pequeño. Como siempre lo he pensado, la mejor manera de agradecerte es a través de mis logros escolares y personales. Este logro en particular te lo dedico a tí, ya que sin tí, yo no estaría aquí. Gracias por estar siempre conmigo y por toda tu motivación. Te quiero.

A mi padre:

Me consta el esfuerzo que has hecho durante todos estos años para darnos a mi hermana y a mí una educación, y el compromiso que siempre has tenido con nosotros. Gracias por ser un modelo a seguir y por haberme conducido de la mejor manera posible. Gracias por inculcarme todos los valores que se necesitan para afrontar la vida como un hombre de bien. Gracias por todo. Te quiero.

A mi hermana:

A tí que te he visto crecer, te agradezco por todo el apoyo y cariño que siempre me has mostrado de cualquier manera. Como hermano mayor, espero poder darte siempre la mejor imagen y servirte de ejemplo. Si alguna vez tienes en mente seguir mi camino, no dudes que estaré ahí para ayudarte y motivarte como tú lo has hecho conmigo. Te quiero.

A otros:

Agradezco a todos los matemáticos -o científicos con corazón de matemáticos- cuyas aportaciones en el tema aparecen aquí. Fue un verdadero reto comprender y sintetizar sus escritos. Agradezco también a todos aquellos cuya música se convirtió en el *soundtrack* e inspiración de la escritura de este trabajo –y en general de mi vida-, en especial a esos cuatro muchachos de Liverpool. Gracias a todos en donde quiera que estén.

Índice general

INTRODUCCIÓN	5
Estructura y objetivos	9
PARTE I. ORÍGENES	13
1.1. Orígenes y antecedentes de los mínimos cuadrados.....	13
1.1.1. La Regla de Cotes	14
1.1.2. Tobias Mayer y la libración de la Luna	15
1.1.3. Saturno, Júpiter y Euler	21
1.1.4. El rescate del Sistema Solar de Laplace	25
1.1.5. Boscovich y la figura de la Tierra	30
1.1.6. Laplace y el Método de Situación	37
1.2. La invención de los mínimos cuadrados.....	41
1.2.1. Legendre y la invención del método	41
1.2.2. Gauss y su primera prueba	48
1.2.3. La Síntesis de Gauss-Laplace.....	51
PARTE II. FUNDAMENTOS	57
2. El modelo de regresión lineal tradicional	57
2.1. Estimación de los parámetros por mínimos cuadrados.....	58
2.1.1. Estimación de los parámetros	58
2.1.2. Propiedades de los estimadores y del modelo ajustado de regresión.....	63
2.1.3. Estimación de σ^2	66
2.2. Pruebas de hipótesis para la pendiente y la intercepción.....	67
2.2.1. Desarrollo de la prueba de hipótesis sobre la pendiente.....	68
2.2.2. Prueba de significancia.....	69
2.2.3. El análisis de varianza.....	71
2.3. Intervalos de confianza en la regresión lineal simple.....	73
2.3.1. Intervalos de confianza para β_0 , β_1 y σ^2	73
2.3.2. Intervalo de confianza para la respuesta media	74
2.4. Predicción de nuevas observaciones	75

ÍNDICE GENERAL

2.5.	Coefficiente de determinación	77
2.6.	Estimación por máxima verosimilitud	77
2.7.	Caso cuando la regresora x es aleatoria	79
2.8.	Un ejemplo numérico en R	80
2.8.1.	Los cohetes de la NASA	80
2.8.2.	Resultados en R para el ejemplo de los cohetes	86
PARTE III.	ALTERNANCIA	97
3.	Método alternativo de regresión lineal por mínimos cuadrados	97
3.1.	Estimación de los parámetros por mínimos cuadrados	97
3.2.	Ventajas y desventajas del método alternativo	107
3.3.	Comparación de precisión	108
PARTE IV.	EXTENSIÓN	111
4.	Regresión lineal múltiple	111
4.1.	Modelos de regresión múltiple	111
4.2.	Estimación de los parámetros del modelo	116
4.2.1.	Estimación de los coeficientes de regresión mediante mínimos cuadrados	116
4.2.2.	Una interpretación geométrica de los estimadores por mínimos cuadrados en el modelo de regresión lineal múltiple	121
4.2.3.	Propiedades de los estimadores por mínimos cuadrados	122
4.2.4.	Estimación de σ^2	123
4.2.5.	Estimación por máxima verosimilitud	124
4.3.	Pruebas de hipótesis en la regresión lineal múltiple	125
4.3.1.	Prueba de significancia	125
4.3.2.	Coefficientes de determinación	128
4.3.3.	Prueba de los coeficientes individuales de la regresión	128
4.4.	Intervalos de confianza en la regresión lineal múltiple	130
4.4.1.	Intervalos de confianza para los coeficientes de la regresión	130
4.4.2.	Intervalo de confianza para la respuesta media	131
4.5.	Predicción de nuevas observaciones	131
4.6.	Un ejemplo numérico en R	132
4.6.1.	Los discos de vinilo de la BBC	132
4.6.2.	Resultados en R para el ejemplo de los vinilos	140

ÍNDICE GENERAL

PARTE V. APLICACIÓN	145
5. Optimización en ingeniería financiera: El método Monte Carlo de mínimos cuadrados de Longstaff-Schwartz	145
5.1. Contexto	145
5.2. Opciones americanas.....	146
5.2.1. Valuación.....	147
5.3. El método LSM de Longstaff-Schwartz.....	148
5.3.1. Marco teórico.....	148
5.3.2. Calibración en LSM.....	151
5.3.3. Resultados de convergencia.....	153
5.4. Ejemplos numéricos y resultados del algoritmo LSM.....	155
5.4.1. ¿Por qué opciones Put americanas?.....	156
5.4.2. Ejemplo con aproximación mediante monomios.....	156
5.4.3. Ejemplo con aproximación mediante polinomios de Laguerre.....	162
5.4.4. Un ejemplo más real	165
CONCLUSIONES	169
BIBLIOGRAFÍA	173
APÉNDICES	175
Apéndice de la Parte I	175
Apéndice A.....	175
A.1. Breve recuento sobre la controversia Gauss - Legendre	175
Apéndice de la Parte III	184
Apéndice A.....	184
A.1. Programa del modelo general de la línea de regresión por el método alternativo y tradicional sobre Maple.....	184
A.2. Ejemplo numérico 1	186
A.3. Ejemplo numérico 2	189
Apéndice B.....	192
B.1. Comparación de precisión para datos con error presente en ambas variables	192
B.2. Comparación de precisión para datos con error presente en ambas variables (Versión de iteración múltiple).....	195
B.3. Comparación de precisión para datos con error presente sólo en la variable y	196

ÍNDICE GENERAL

B.4. Comparación de precisión para datos con error presente sólo en la variable y (Versión de iteración múltiple)	199
B.5. Resultados de precisión entre ambos métodos	200
Apéndice de la Parte IV	201
Apéndice A	201
A.1. Distribuciones Centrales	201
A.2. Distribuciones no centrales.....	203
Apéndice B	204
B.1. Definiciones básicas para matrices	204
B.2. Derivadas de las matrices	206
B.3. Esperanzas	207
B.4. Teoría de distribución	207
Apéndice C	209
C.1. Resultados importantes de SC_R	209
C.2. Resultados importantes de SC_{Res}	211
C.3. La prueba F	213
Apéndice D	215
D.1. El teorema de Gauss – Markov cuando $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$	215
Apéndice de la Parte V	217
Apéndice A	217
A.1. Black-Scholes PDE (Ecuación diferencial parcial de Black-Scholes).....	217
Apéndice B	219
B.1. Demostración de las proposiciones.....	219
Apéndice C	221
C.1. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana (con monomios como funciones base)	221
C.2. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana (con polinomios de Laguerre como funciones base)	224
C.3. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana expedida por PEMEX.....	227

INTRODUCCIÓN

Uno de los problemas fundamentales de la estadística es aquel en donde se pretende encontrar la mejor ecuación de regresión lineal o no lineal que exprese la relación entre una variable dependiente y una o más variables independientes. Dadas ciertas observaciones, la variable dependiente está sujeta a un error aleatorio. Estas observaciones son más grandes en número que los parámetros en la ecuación de regresión sobre la variable dependiente y los valores correspondientes de las variables independientes, los cuales deberían ser conocidos de forma exacta o también estar sujetos al error aleatorio.

Algunos problemas relacionados son aquellos en donde se escogen las mejores medidas de tendencia central y de dispersión de las observaciones. Las mejores soluciones a estos tres problemas dependen de la distribución de los errores aleatorios. Si se supone que los valores de las variables independientes son conocidos de forma exacta y que los errores en las observaciones de la variable dependiente están normalmente distribuidos, entonces es bien conocido que la media aritmética es la mejor medida de tendencia central, la desviación estándar es la mejor medida de dispersión y el *método de mínimos cuadrados* es el mejor método para ajustar una ecuación de regresión. Otros supuestos conducen a diferentes cosas.

El nombre de *mínimos cuadrados* deriva del hecho de que este método consiste en minimizar una suma de cuadrados -o de alguna otra forma cuadrática- de las observaciones y las cantidades -por ejemplo, coeficientes de la regresión- que están siendo estimadas. En el caso degenerado, la cantidad estimada es el valor verdadero α de una cantidad física sobre la cual se tienen distintas muestras o mediciones y_i ($i=1,2,\dots,n$), cada una de ellas sujetas al error. Éste en la i -ésima observación será denotado por ε_i . Se tienen entonces las siguientes ecuaciones:

$$y_i = \alpha + \varepsilon_i, \quad \hat{y}_i = a, \quad e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n, \quad (1)$$

donde a y e_i son los valores estimados de α y ε_i , respectivamente. Resolviendo para e_i , elevando al cuadrado ambos lados, sumando los n valores de i y ocultando los subíndices, se obtiene:

INTRODUCCIÓN

$$\sum e^2 = \sum (a - y)^2 \quad (2)$$

Para encontrar el valor de a que minimiza $\sum e^2$, se toma la derivada parcial con respecto a a e igualando a cero $\partial \sum e^2 / \partial a$ produce:

$$a = \sum y / n = \bar{y}, \quad (3)$$

donde \bar{y} es el promedio de las y_i . Nótese que el valor estimado por medio de mínimos cuadrados del valor verdadero α es el promedio de las observaciones, dado que la suma de los cuadrados de las desviaciones de la media es menor que de cualquier otro valor. El valor estimado de la dispersión -correspondiente al error- es la desviación estándar, llamada también la raíz del cuadrado de la desviación.

El problema de regresión no degenerado más simple, es en el cual se pretende estimar la relación lineal $y = \alpha + \beta x$, dado un conjunto de puntos de la forma (x_i, y_i) , $i = 1, 2, \dots, n$, donde los valores x_i son exactos y los valores y_i están sujetos al error. El error en la i -ésima observación se denotará nuevamente por ε_i . Se tienen entonces las siguientes ecuaciones:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \hat{y}_i = a + b x_i, \quad e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n \quad (4)$$

donde a , b y e_i son los valores estimados de α , β y ε_i respectivamente. Resolviendo para e_i , elevando al cuadrado ambos lados, sumando los n valores de i y ocultando los subíndices, se obtiene:

$$\sum e^2 = \sum (a + b x - y)^2 \quad (5)$$

Para encontrar los valores de los coeficientes de regresión a y b que minimizan (5), se toman las derivadas parciales con respecto a a y b , y se igualan a cero $\partial \sum e^2 / \partial a$ y $\partial \sum e^2 / \partial b$, lo cual lleva a:

$$\begin{aligned} na + b \sum x &= \sum y \\ a \sum x + b \sum x^2 &= \sum xy \end{aligned} \quad (6)$$

Las ecuaciones (6) son las llamadas *ecuaciones normales*. Resolviendo estas ecuaciones de forma simultánea para a y para b se tiene:

INTRODUCCIÓN

$$b = \frac{[n \sum xy - \sum x \sum y]}{[n \sum x^2 - (\sum x)^2]}$$
$$a = \frac{[\sum x^2 \sum y - \sum x \sum xy]}{[n \sum x^2 - (\sum x)^2]} = \bar{y} - b\bar{x} \quad (7)$$

Los valores de los coeficientes b y a dados en (7) son, por supuesto, la pendiente y la intercepción con el eje y , respectivamente, la recta obtenida por mínimos cuadrados.

El método de mínimos cuadrados también puede ser usado para estimar los coeficientes de ecuaciones de regresión lineal en donde están involucradas dos variables independientes (planos) o tres o más (hiperplanos). También es aplicado para encontrar los coeficientes de una regresión no lineal, donde está involucrada una variable independiente (curvas), dos (superficies), o tres o más (hipersuperficies).

En una ecuación de regresión, las funciones cuyos coeficientes se desean encontrar no son precisamente lineales. Éstas pueden ser trascendentales, por ejemplo: exponenciales, logarítmicas o trigonométricas. En cada caso, las ecuaciones normales son encontradas tomando las derivadas parciales de $\sum e^2$ con respecto a cada uno de los coeficientes desconocidos que se van a estimar. De aquí que el número de ecuaciones normales resultantes es igual al de coeficientes desconocidos, siendo posible su resolución de manera simultánea.

El método de mínimos cuadrados parece por sí solo bastante sencillo, y con la ayuda de la labor computacional, es ahora mucho menos fatigante de lo que fue antes, anterior al desarrollo de computadoras y lenguajes de programación.

El método de mínimos cuadrados fue desarrollado de forma independiente por Gauss en Alemania y Legendre en Francia. La historia estipula que Legendre fue el primero en publicarlo. Sin embargo, los orígenes del método se vuelven oscuros y controvertidos, como se verá en el apéndice de la *Parte I*. Legendre comenzó con la forma lineal $E = a + bx + cy + \dots$ y derivó las ecuaciones normales sin el uso explícito del cálculo. El principio fundamental de Legendre consistió en multiplicar esta forma lineal de cantidades desconocidas, por cada uno de los coeficientes de las incógnitas, sumando sobre todas las observaciones e igualando después cada suma a cero.

Legendre ofreció una prueba no matemática del método de mínimos cuadrados, sin embargo, puso un especial énfasis en la superioridad que poseía en relación a otros métodos. Poco después del trabajo de Legendre, Gauss reclamó su invención, afirmando

INTRODUCCIÓN

que él ya lo había usado durante doce años, prometiendo a la comunidad científica publicar su resultado más tarde.

Adrain, un matemático americano, desconociendo aparentemente el trabajo de Legendre y el aún no publicado trabajo de Gauss, desarrolló de forma independiente el método de mínimos cuadrados usándolo en diversos problemas. Gauss dedujo la hoy conocida *ley gaussiana del error*, ante el hecho de que cuando se tiene cualquier número de observaciones directas igualmente buenas de una cantidad desconocida x , el valor más probable de esta cantidad es la media aritmética. Gauss entonces mostró lo que se conocería como su primera prueba: que el método de mínimos cuadrados, usado por él desde 1795, pero bautizado por Legendre, se sigue como consecuencia de la ley gaussiana del error.

Laplace dio una rudimentaria forma del teorema del límite central y lo usó en dos de sus trabajos posteriores para mostrar que la media aritmética es “más ventajosa” sólo cuando la distribución del error es normal (gaussiana) y es “desventajosa” cuando el número de observaciones es muy grande o cuando uno toma el promedio de resultados, cada uno de los cuales está basado en un número grande de observaciones y, por lo tanto, en cada uno de los casos se debería usar el método de mínimos cuadrados. Laplace, en su trabajo monumental sobre la teoría analítica de probabilidades, defendió el uso de los mínimos cuadrados, aunque de ninguna forma excluyó los métodos antecesores.

Gauss comparó su formulación de 1809 (primera prueba) del método de mínimos cuadrados con los trabajos de Laplace y concluyó que ninguno de los dos era completamente satisfactorio. Él ofreció otra exposición (segunda prueba) basada en minimizar una función de pérdida proporcional a la suma de los cuadrados de los errores y sobre esta base justificó el uso del método de mínimos cuadrados, sin importar el número de observaciones ni la distribución de sus errores. Sus argumentos fueron aparentemente convincentes para la mayoría de sus contemporáneos, desde el hecho de que la literatura estadística de muchas de las siguientes décadas incluye numerosos escritos sobre mínimos cuadrados, pero muy pocos sobre sus métodos rivales o alternativos.

Diversos historiadores matemáticos han aportado trabajos de gran valía. En 1877, Merriman compiló una extensa bibliografía de escritos sobre mínimos cuadrados que existían hasta la fecha. Excelentes libros sobre la teoría de mínimos cuadrados y sus aplicaciones fueron publicados a finales del siglo XIX y a principios del XX. Particularmente, son notables aquellos trabajos de Merriman y Helmert, cuyas aplicaciones recaen en la geodésica. Plackett ha examinado en detalle, con la ayuda de la correspondencia, las circunstancias en las cuales tomó lugar el descubrimiento del método de mínimos cuadrados y los orígenes de la controversia entre Gauss y Legendre.

INTRODUCCIÓN

Muchos programas de estadística computacional han incluido bibliotecas en donde el método de mínimos cuadrados es aplicado a la regresión lineal simple, a la regresión lineal múltiple (incluyendo su desarrollo mediante subconjuntos) y a la regresión no lineal, destacando por ejemplo: BMDP, IMSL, R, SPSS, etc.

A pesar de que el método de mínimos cuadrados ha sido criticado a través de los años y que diversos métodos alternativos han sido propuestos, el método de mínimos cuadrados ha sido el método predominante durante más de 200 años, debido a su fácil derivación matemática y a su simplicidad computacional. Hoy en día, este método es enseñado en prácticamente todos los libros de estadística e incluso ha excluido a todos los demás métodos.

Si la distribución de los errores fuera mesocúrtica (de cola mediana) y no existieran *outliers* en las observaciones, el método de mínimos cuadrados debería ser el ideal. Sin embargo, éste no es robusto ante la presencia de *outliers*, o cuando la distribución de los errores es leptocúrtica (de cola pesada). De cualquier forma, cuando se aprecian *outliers* dentro de las observaciones, éstos pueden ser rechazados antes de usar el método de mínimos cuadrados, o cuando uno mucho más robusto puede ser utilizado. Los resultados arrojados por el método de mínimos cuadrados cuando se tiene una distribución platicúrtica (de cola ligera) no son precisamente desastrosos, sin embargo, diversas mejoras pueden ser hechas por la aplicación del método más apropiado.

Estructura y objetivos

Este trabajo lleva por nombre *Mínimos cuadrados: orígenes, fundamentos, alternancia, extensión y aplicación* y se encuentra conformado por cinco partes o capítulos.

En la *Parte I* se dan a conocer y se estudian los orígenes del método de mínimos cuadrados, culminando con su oportuna invención y primeros estudios de sus propiedades. En esta sección se analizarán los trabajos hechos por grandes matemáticos de los siglos XVIII y XIX, todos encaminados sobre un claro objetivo: encontrar un método general para el ajuste de ecuaciones observacionales. Se sigue con la invención del método general de mínimos cuadrados, situación que desencadena uno de los más grandes conflictos en la historia de las matemáticas por la atribución del magnífico descubrimiento. Por último, se estudian los primeros resultados en la historia del método, los cuales le dieron la credibilidad que éste reclamaba. A lo largo de todo el capítulo aparecen nombres concretos, pero sobre todo, está el trabajo de muchos matemáticos que hicieron evolucionar el conocimiento humano. Este capítulo está conformado de

INTRODUCCIÓN

historia matemática como tal, pero en ningún momento se descuida el rigor teórico-matemático que la sustenta.

En la *Parte II* se presentan los fundamentos del modelo de regresión lineal tradicional, haciendo uso del principio de mínimos cuadrados, esto es, usando una sola variable regresora. Se desarrollan las partes fundamentales del análisis de la regresión lineal, desde la estimación de los parámetros, pasando por las pruebas de significancia, intervalos de confianza y de predicción, así como la interpretación de los resultados. Se presenta la teoría clásica de una forma clara y estructurada. Uno de los objetivos a lo largo de este trabajo también es mostrar la importancia que representan los distintos tipos de software matemático, usando el paquete más adecuado en cada capítulo a lo largo de la tesis. Es por eso que en este capítulo, se comienza con R, cuando se verifican los cálculos obtenidos directamente de la teoría durante un ejemplo numérico, en un intento por identificar una forma práctica y precisa para realizar un análisis de regresión simple. Todos los códigos del programa utilizados aquí, son explicados.

En la *Parte III* se presenta la teoría de un método de regresión lineal alternativo de mínimos cuadrados. Se destacan también las ventajas y desventajas que este método representa en los aspectos teóricos y computacionales. Se desarrolla también un criterio para medir y comparar la precisión entre este método y el tradicional, presentado en el capítulo anterior. El software utilizado aquí es *Maple*. Se presentan diversos programas elaborados en este paquete, destinados a encontrar la línea ajustada por ambos métodos de forma simultánea, así como también para encontrar cuál de los dos es más exacto de acuerdo al criterio de precisión estudiado, tanto para errores presentes sólo en la variable de respuesta, así como para el caso especial en donde el error también se encuentra en la variable regresora.

En la *Parte IV* se desarrolla el modelo de regresión lineal múltiple por mínimos cuadrados, siguiendo el mismo planteamiento de la *Parte II*, en una clara extensión del modelo lineal simple. También se presenta la teoría concreta del uso de matrices, indispensable para entender algunas pruebas y resultados. Un ejemplo de este modelo es desarrollado de la misma forma que en la *Parte II*, haciendo primero algunos cálculos directamente de la teoría y después, utilizando nuevamente R para mostrar lo sencillo que se vuelven. Ante ello, será posible observar lo verdaderamente útil y práctico que resulta la implementación de software, en donde los cálculos a mano pueden resultar tediosos al tratar ahora con matrices. Los códigos del programa son explicados.

En la *Parte V* se identifica la teoría del algoritmo de Monte Carlo de mínimos cuadrados (abreviado *LSM* por sus siglas en inglés), implementado por Longstaff y Schwartz para la valuación de opciones Put americanas, presentando sus diferentes etapas. Se comienza

INTRODUCCIÓN

con una explicación de conceptos básicos de opciones financieras para situarse en contexto, así como sus diferentes maneras de valorar estos instrumentos financieros y la complicación que existe en la toma de decisiones cuando se trabaja con estos derivados de ejercicio continuo. Se presenta un ejemplo numérico básico, pero muy ilustrativo, que reafirma la teoría de forma clara. Otro ejemplo con datos reales es presentado para la valuación de una Put americana expedida por Pemex, en la venta del crudo mexicano. Esta vez, se implementa *MATLAB* en la construcción de programas que facilitan los cálculos, mostrando también los asombrosos resultados de convergencia estudiados previamente.

PARTE I

ORÍGENES

1.1. Orígenes y antecedentes de los mínimos cuadrados

El método de mínimos cuadrados fue sin duda el tema predominante dentro del siglo XIX de la estadística matemática. En muchos aspectos, la estadística se convirtió en el centro de estudio de grandes intelectuales de la época, tal y como lo había sido el estudio del cálculo un siglo antes. Las pruebas del método de mínimos cuadrados dieron una firme dirección al desarrollo de la estadística. Los trabajos explicando su uso, guiados por su aplicación de una forma sobresaliente, además de las disputas relativas a su descubrimiento, trajeron consigo el reconocimiento y valor del método por parte de la comunidad intelectual. Al igual que el cálculo en matemáticas, este “cálculo de observaciones” cuenta con sus propios antecedentes. La exploración de su potencial y sutileza tuvo lugar durante casi un siglo.

En los comienzos del siglo XVIII, una versión generalizada de asignar pesos a las observaciones fue introducida por Cotes. Subsecuentemente, los desarrollos en el análisis de datos involucraron métodos de estimación para modelos lineales cuando el número de variables eran dos o más. Otros, desarrollados de manera independiente por Mayer y Euler, consistían en dividir las ecuaciones de condición en grupos, tantos como parámetros hubiesen, resolviendo para cada uno y comparando los resultados con los obtenidos mediante observaciones. Un método más de subdivisión fue empleado por Laplace. La subdivisión, considerada a veces como arbitraria y subjetiva, fue removida por Boscovich, haciendo uso de un complicado análisis netamente geométrico aplicado a mediciones de arcos. A pesar de esto, cabe destacar sus condiciones y primeras nociones del comportamiento del error. El trabajo de Boscovich fue inspiración para Laplace, al igual que el algoritmo conocido como “Método de Situación” que, además de ser una versión analítica del desarrollo de Boscovich, muestra su ingeniosa idea de asignar pesos a las observaciones mediante ponderaciones de sus magnitudes. Ambos trabajos fueron aplicados a la Geología.

Después de todos estos desarrollos admirables, el método de mínimos cuadrados apareció impreso por primera vez en 1805. Este hecho marcó el principio de una lucha por la atribución del fabuloso descubrimiento entre dos grandes matemáticos, incluso comparada por algunos, a aquella controversia por la invención del cálculo entre Newton y Leibniz. Los nombres: Adrien Marie Legendre y Carl Friedrich Gauss.

En esta sección (1.1), se analizarán los trabajos anteriormente descritos, abordando todo aquello que en cierta forma dio origen al método de mínimos cuadrados; continuando en la siguiente sección (1.2) con su invención oportuna. Se inicia entonces con las primeras señales de vida del método.

1.1.1. La Regla de Cotes

La técnica de combinar observaciones independientes, sintetizadas a una única cantidad formando lo que se conoce como media aritmética, había aparecido al final del siglo XVII. Diversos astrónomos consideraban que las observaciones recabadas debían tener la misma precisión, debido al hecho de que éstas eran realizadas por el mismo observador, en el mismo tiempo, en el mismo lugar, y con el mismo instrumento. Excepciones: instancias en las cuales las observaciones no eran consideradas con equivalente precisión, eran muy raras antes de 1750. La primera excepción a este hecho es una regla encontrada en el trabajo de Roger Cotes (10 de julio de 1682 - 5 de junio de 1716) publicado de manera póstuma en 1722. El principio de la Regla de Cotes es el siguiente:

“Sea p el lugar de algún objeto definido por la observación; q, r, s , los lugares del mismo objeto en las observaciones subsecuentes. Sean también P, Q, R y S los pesos recíprocamente proporcionales a los desplazamientos, los cuales deben surgir debido a los errores en cada una de las observaciones. Estos errores están definidos en un cierto límite establecido. Entonces, los pesos P, Q, R y S son colocados en las posiciones p, q, r , y s , y su centro de gravedad Z es obtenido, siendo el punto Z la posición más probable del objeto en cuestión.” (Cotes, 1722, p.22)

La Regla de Cotes puede ser interpretada como una media con pesos no equitativos -a cada observación se le asigna un cierto peso-, o incluso, como una muy temprana aparición del método de mínimos cuadrados. Sin embargo, esta última conjetura no cuenta con el sustento suficiente para considerarla como verdadera. En ninguna parte de los trabajos de Cotes aparecen ejemplos que sustenten dicha afirmación. La Regla de Cotes tuvo muy poca o nula influencia en los estudios inmediatos posteriores. Una ligera revisión sobre esta regla apareció en un escrito de Laplace en el año de 1812.

Para entender el origen del método de mínimos cuadrados hace falta adentrarse, de cierta forma, a los trabajos realizados por grandes astrónomos y matemáticos de la época, antes de la aparición formal del método por parte de Legendre en 1805.

1.1.2. Tobias Mayer y la libración de la Luna

El desarrollo del método de mínimos cuadrados estuvo fuertemente asociado con tres de los más grandes problemas científicos del siglo XVIII: (1) determinar y representar matemáticamente el movimiento de la Luna; (2) estudiar un aparente movimiento no periódico que había sido observado en los planetas Júpiter y Saturno; y (3) determinar la forma de la figura de la Tierra. Todos estos problemas involucraban observaciones astronómicas, así como la teoría de atracción gravitacional y todos ellos presentaron retos intelectuales que llamaron la atención de los científicos más capaces del período.

Existen dos grandes trabajos que merecen ser señalados: *Recherches sur la question des inégalités du mouvement de Saturne et Jupiter* (Investigaciones sobre la cuestión de las desigualdades del movimiento de Saturno y Júpiter), escrito por Leonhard Euler y publicado en 1749; y *Abhandlung über die Umwälzung des Mondes um seine Axe und die scheinbare Bewegung der Mondsflecken* (Tratado sobre la Luna, su eje, y el movimiento aparente en relación con la Tierra) escrito por Tobias Mayer y publicado en 1750. A pesar de que estos tratados no fueron los primeros en ser publicados en relación a sus respectivos problemas (las desigualdades en el movimiento de Júpiter y Saturno y la libración de la Luna), se convirtieron en los más valiosos en relación a sus objetivos. En cierta forma, fueron de gran influencia para escritos posteriores y, desde un punto de vista estadístico, establecieron una forma única e innovadora en el manejo de las observaciones. Cuentan la historia de un éxito estadístico -por un dedicado astrónomo, Mayer- y un fracaso estadístico -por un líder matemático, Euler-. Estos dos trabajos muestran por qué el descubrimiento del método de mínimos cuadrados no era posible en el contexto intelectual de 1750; destacando las barreras conceptuales que habían sido finalmente cruzadas, prediciendo de algún modo los posteriores avances de Legendre, Gauss y Laplace.

Se expone primero el éxito estadístico logrado por Mayer. A pesar del supuesto movimiento “mensual” y regular de la Luna alrededor de la Tierra, el estudio detallado de éste es extremadamente complejo. En el siglo XVIII, la precisión para medir las pequeñas perturbaciones en el movimiento de la Luna, ya sea mediante el uso de alguna fórmula matemática o a partir de una tabla construida empíricamente describiendo su posición futura, tenía una gran importancia científica, comercial y también militar. La primera recae en el deseo general de demostrar que la *teoría gravitacional de Newton* puede influir en el movimiento de cuerpos celestes que se encuentran más próximos a uno, si se tiene en cuenta la atracción de otros externos -tales como el Sol-, trayendo consigo cambios

PARTE I. ORÍGENES

periódicos en las órbitas de la Tierra y la Luna, además de desviaciones en la esfericidad de estos dos cuerpos celestes. Sin embargo, se debe decir que el interés en el movimiento de la Luna era principalmente comercial y militar. Durante los nueve siglos anteriores, desde Hipparchus y Ptolomeo, hasta Newton y Flamsteed, la astronomía práctica y teórica había jugado un papel muy importante en la navegación de barcos mercantiles y militares. El desarrollo de mejores instrumentos náuticos, así como una mejor comprensión de las teorías astronómicas, incrementaron los viajes en altamar a través de mares desconocidos. Para el año de 1700 ya era posible determinar la latitud en el mar con relativa precisión gracias a las estrellas. Sin embargo, determinar la longitud del mar no era una tarea sencilla. En 1714, Inglaterra estableció a “los comisionarios para el descubrimiento de la longitud del mar”, un grupo que, para 1815, había ya desembolsado una cantidad de 101,000 libras en premios y subvenciones buscando alcanzar este objetivo. El método más prometedor estaba basado en tablas del movimiento lunar, permitiendo el descubrimiento del tiempo de Greenwich. Este método comparaba la posición de la luna con respecto a las estrellas.

Johann Tobias Mayer (17 de febrero de 1723 – 20 de febrero de 1762) ya se había hecho de nombre por sí solo como un cartógrafo y astrónomo para el año de 1747, momento en el cual llevó a cabo un estudio sobre la Luna. Su trabajo publicado en 1750, abordaba el fenómeno de libración. Se denomina libración lunar al conjunto de movimientos de oscilación que presenta el disco de la Luna con respecto a un observador ubicado en la Tierra. La noción popular de que el satélite presenta siempre la misma cara hacia nuestro planeta, no es precisamente cierta. Su cara visible desde la Tierra varía, de tal modo que durante un extenso período de tiempo es posible observar alrededor del 60% de la superficie lunar. Aspectos importantes sobre la libración fueron encontrados por Galileo: la aparente libración diurna, dada por la rotación de la Tierra y una libración en latitud gracias al hecho de que el eje de rotación de la Luna no es perpendicular a la órbita de la Tierra cuando gira alrededor del Sol. En los tiempos de Mayer, era ya conocido que la ubicación de la Tierra era un foco, no el centro, de la órbita elíptica lunar. Entonces, la rotación de la Luna a una velocidad uniforme producía un tercer tipo de libración, uno de longitud.

Durante el período de abril de 1748 a marzo de 1749, Mayer hizo numerosas observaciones de la posición de distintos rasgos lunares y, dentro de su trabajo de 1750, mostró como estos datos podían ser usados para determinar varias características de la órbita lunar. La manera con la cual manejó estos datos fue novedosa. Se revisa a continuación de manera general los conceptos de Mayer:

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

Su método para la solución de sistemas de ecuaciones inconsistentes -obtenidas gracias a las observaciones- fue desarrollado con base en la discusión sobre la posición del cráter de Manilius. La siguiente figura representa a la Luna, la cual Mayer la considera como una esfera.

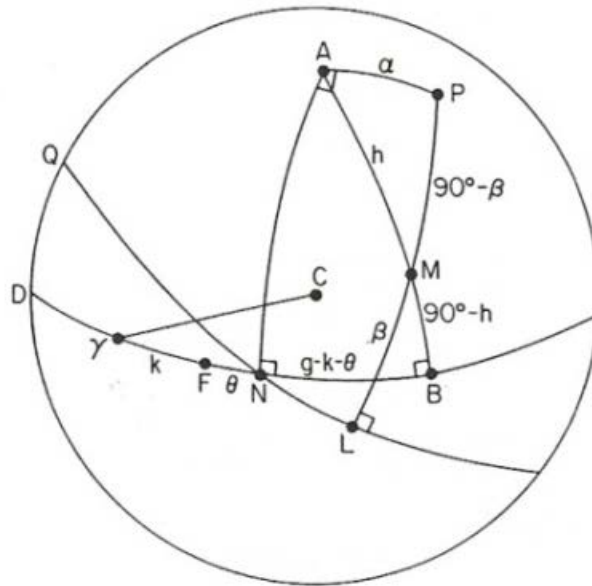


Figura 1.1. La Luna de Mayer. [17] pp. 20

El gran círculo QNL representa el verdadero ecuador de la Luna, siendo P su polo con relación a éste. El otro gran círculo DNB es el aparente ecuador visto desde la Tierra. A es el polo de la Luna con respecto a DNB . El punto γ fue tomado como referencia, y representa la trayectoria del centro de la Luna hacia el equinoccio. El círculo DNB y el polo A varían con el tiempo, consecuencia inmediata de la libración de la Luna, sin embargo, forman parte del sistema natural de coordenadas en un tiempo dado. Tanto el ecuador QNL como el polo P son fijados, pero no son observables desde la Tierra. El objetivo de Mayer era determinar la relación entre las variables, pudiendo así obtener de manera precisa las ubicaciones de QNL y P . Éste llevó a cabo repetidas observaciones del cráter de Manilius, representado en la figura mediante el punto M . Por otro lado, PL y AB son las perpendiculares a través de M con respecto a los dos sistemas coordenados de los polos. Mayer fue capaz de observar la posición de M en diversas ocasiones con respecto al constante cambio del sistema coordenado determinado por DNB y A ; esto es, pudo encontrar la medida de los arcos $AM \equiv h$ y $\gamma B \equiv g$, todo esto sujeto al error observacional.

Para determinar la relación entre el sistema coordenado, Mayer intentó buscar la longitud desconocida del arco fijo $AP \equiv \alpha$, la verdadera latitud de Manilitus $\beta \equiv ML$ y la distancia

PARTE I. ORÍGENES

θ entre el punto desconocido de intersección N de los dos círculos y el punto conocido F , el cual representa la intersección del plano de la órbita de la Luna y el círculo DNB . Entonces, g , h y k eran observables y variables en cada observación como resultado del movimiento de la Luna -y también del error observacional-; mientras que α , θ y β eran fijas pero desconocidas, buscando ser determinadas mediante las observaciones. Debido a que NAP forma un ángulo recto, una identidad de la trigonometría esférica relaciona a estas cantidades dentro de una ecuación no lineal.

$$\text{sen } \beta = \cos \alpha \cos h + \text{sen } \alpha \text{ sen } h \text{ sen}(g - k - \theta)$$

Mayer sabía que tanto α como θ eran pequeños -alrededor de 2 a 3 grados-, procediendo entonces, vía identidades trigonométricas, a derivar una aproximación “casi lineal” a esta ecuación, bajo la suposición de que $\cos \alpha$, $\cos \theta$, y $\cos(\beta - 90^\circ + h)$ eran aproximadamente 1.0 y $\text{sen}(\beta - 90^\circ + h) \cong \beta - 90^\circ + h$. Esto llevó a la ecuación:

$$\beta - (90^\circ - h) = \alpha \text{ sen}(g - k) - \alpha \text{ sen } \theta \cos(g - k),$$

la cual muestra una relación entre las variables observables y las variables desconocidas.

Mayer estaba a punto de alcanzar su objetivo. Solamente necesitaba hacer observaciones de tres días diferentes y resolver las tres ecuaciones resultantes para α , β y $\alpha \text{ sen } \theta$ -resolviendo después para θ -. Sin embargo, Mayer fue más allá. Estuvo observando a Manilius durante veintisiete días, obteniendo un total de veintisiete ecuaciones. Las ecuaciones de Mayer llegaron a ser nombradas como *ecuaciones de condición*, debido a que expresaban una relación que se mantendría si no se consideraban los errores. En términos modernos es posible introducir el error como:

$$(h - 90^\circ) = -\beta + \alpha \text{ sen}(g - k) - \alpha \text{ sen } \theta \cos(g - k) + E,$$

tratando a $h - 90^\circ$ como la variable dependiente y a β , α y $-\alpha \text{ sen } \theta$ como los parámetros de un modelo de regresión lineal. ¿Cómo hacer frente a este conjunto de ecuaciones? El objetivo de Mayer consistió en reducir este sistema a uno más simple, en el cual un lector del presente hubiera llegado a la errónea opinión de que este procedimiento no es del todo notable. Dividió sus ecuaciones en tres grupos de nueve, incorporando mediante suma, cada uno de los tres sistemas en una sola ecuación por separado, y resolviendo el conjunto resultante de tres ecuaciones lineales para α , β y $\alpha \text{ sen } \theta$ -resolviendo después para θ -.

La elección de cuál ecuación debería pertenecer a cada grupo, estuvo basada en los coeficientes de las variables α , β y $\alpha \text{ sen } \theta$. El primer grupo consistía en aquellas nueve

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

con el coeficiente positivo más grande para α , llamadas: 1, 2, 3, 6, 9, 10, 11, 12, y 27. El segundo grupo estaba compuesto por las ecuaciones con el coeficiente negativo más grande para la misma variable: 8, 18, 19, 21, 22, 23, 24, 25 y 26. El resto formaban el tercer grupo, el cual parece tener el valor más grande para el coeficiente de $\alpha \sin \theta$.

A continuación se muestran las veintisiete ecuaciones condición de Mayer:

No.	Ecuación	Grupo
1	$\beta - 13^{\circ}10' = +0.8836\alpha - 0.4682\alpha \sin \theta$	I
2	$\beta - 13^{\circ}8' = +0.9996\alpha - 0.0282\alpha \sin \theta$	I
3	$\beta - 13^{\circ}12' = +0.9899\alpha + 0.1421\alpha \sin \theta$	I
4	$\beta - 14^{\circ}15' = +0.2221\alpha + 0.9750\alpha \sin \theta$	III
5	$\beta - 14^{\circ}42' = +0.0006\alpha + 1.0000\alpha \sin \theta$	III
6	$\beta - 13^{\circ}1' = +0.9308\alpha - 0.3654\alpha \sin \theta$	I
7	$\beta - 14^{\circ}31' = +0.0602\alpha + 0.9982\alpha \sin \theta$	III
8	$\beta - 14^{\circ}57' = -0.1570\alpha + 0.9876\alpha \sin \theta$	II
9	$\beta - 13^{\circ}5' = +0.9097\alpha - 0.4152\alpha \sin \theta$	I
10	$\beta - 13^{\circ}2' = +1.0000\alpha + 0.0055\alpha \sin \theta$	I
11	$\beta - 13^{\circ}12' = +0.9689\alpha + 0.2476\alpha \sin \theta$	I
12	$\beta - 13^{\circ}11' = +0.8878\alpha + 0.4602\alpha \sin \theta$	I
13	$\beta - 13^{\circ}34' = +0.7549\alpha + 0.6558\alpha \sin \theta$	III
14	$\beta - 13^{\circ}53' = +0.5755\alpha + 0.8178\alpha \sin \theta$	III
15	$\beta - 13^{\circ}58' = +0.3608\alpha + 0.9326\alpha \sin \theta$	III
16	$\beta - 14^{\circ}14' = +0.1302\alpha + 0.9915\alpha \sin \theta$	III
17	$\beta - 14^{\circ}56' = -0.1068\alpha + 0.9943\alpha \sin \theta$	III
18	$\beta - 14^{\circ}47' = -0.3363\alpha + 0.9418\alpha \sin \theta$	II
19	$\beta - 15^{\circ}56' = -0.8560\alpha + 0.5170\alpha \sin \theta$	II
20	$\beta - 13^{\circ}29' = +0.8002\alpha + 0.5997\alpha \sin \theta$	III
21	$\beta - 15^{\circ}55' = -0.9952\alpha - 0.0982\alpha \sin \theta$	II
22	$\beta - 15^{\circ}39' = -0.8409\alpha + 0.5412\alpha \sin \theta$	II
23	$\beta - 16^{\circ}9' = -0.9429\alpha + 0.3330\alpha \sin \theta$	II
24	$\beta - 16^{\circ}22' = -0.9768\alpha + 0.2141\alpha \sin \theta$	II
25	$\beta - 15^{\circ}38' = -0.6262\alpha - 0.7797\alpha \sin \theta$	II
26	$\beta - 14^{\circ}54' = -0.4091\alpha - 0.9125\alpha \sin \theta$	II
27	$\beta - 13^{\circ}7' = +0.9284\alpha - 0.3716\alpha \sin \theta$	I

Tabla 1.1. Ecuaciones de condición de Mayer. [17] pp. 22

PARTE I. ORÍGENES

Los coeficientes de α y $\alpha \sin \theta$ son $\sin(g-k)$ y $\cos(g-k)$, los cuales son relacionados mediante la identidad $[\sin(g-k)]^2 + [-\cos(g-k)]^2 = 1$. El primer grupo consiste de aquellas nueve ecuaciones cuyos coeficientes para α son cercanos a 1.0, el segundo con coeficientes de α cercanos a -1.0, dejando al tercer grupo con $\sin(g-k)$ “cerca” del cero, es decir, con $-\cos(g-k)$ “grandes”. La forma de seleccionar ecuaciones mediante técnicas de agregación, sujetas a la restricción de cada uno de los tres grupos, maximiza el contraste entre los coeficientes de α y produce buenos estimadores para α , y la correcta presentación del sistema coordinado arroja también buenos estimadores para $\alpha \sin \theta$. Parece que Mayer entendió esto, pues escribió: “Estas ecuaciones (Tabla 1.2) pueden tomar totalmente el lugar de las anteriores, debido a que cada una de estas tres fue construida en la más ventajosa de las formas. La ventaja del método consiste en que la división en estas tres clases, hace que las diferencias entre las sumas sean lo más grande posible. La más grande de estas diferencias determinaría los valores de α , β y θ ” (Mayer, 1750, p. 154).

Grupo	Ecuación
I	$9\beta - 118^{\circ}8' = +8.4987\alpha - 0.7932\alpha \sin \theta$
II	$9\beta - 140^{\circ}17' = -6.1404\alpha + 1.7443\alpha \sin \theta$
III	$9\beta - 127^{\circ}32' = +2.7977\alpha + 7.9649\alpha \sin \theta$

Tabla 1.2. Ecuaciones agregadas. [17] pp. 23

Mayer resolvió las tres ecuaciones que había construido a partir de las veintisiete obteniendo: $\alpha = 89'.90 \cong 1^{\circ}30'$, $\theta = -3.45'$ y $\beta = 14^{\circ}33'$, siguiendo con determinar la precisión de estos valores. A pesar de que no hizo algún tipo de intento sobre el análisis no estadístico del error, mostró un estudio empírico sobre la precisión. Antes de haber encontrado estos valores, demostró como α , β y θ eran calculados basándose en sólo tres de las ecuaciones observacionales (9, 16 y 19 de la Tabla 1.1). El valor encontrado para α en este caso fue $\alpha = 1^{\circ}40'$. Entonces señaló: “Debido a que los últimos valores -basados en las veintisiete ecuaciones- fueron derivados de nueve veces más observaciones, uno puede concluir que son nueve veces más precisos; por lo tanto, el error en cada una de las constantes se encuentra en relación inversa al número de sus observaciones.” (Mayer, 1750, p. 154)

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

Mayer convirtió esta declaración en una descripción de la variable α introduciendo el símbolo $\pm x$ para el error, tomando $\alpha = 1^\circ 30'$; siendo el error de $10 \pm x$ cuando se considera a $\alpha = 1^\circ 40'$. Debido a que el resultado de $1^\circ 30'$ estaba basado en nueve veces más observaciones que el de $1^\circ 40'$ -y por ende, nueve veces más preciso-, supuso que el error estaría determinado por la ecuación $\pm x \cdot 27 = (10 \pm x) \cdot 3$, señalando que ambos lados deberían tener el mismo signo, habiendo considerado $e = \pm x$. Procedió entonces a resolver $e \cdot 27 = (10 + e) \cdot 3$ y $|e| \cdot 27 = |10 + e| \cdot 3$, obteniendo $e = 30/24 = 1.25$ y $e = -1$ respectivamente. Tomó en forma deliberada el valor más grande, para dar en cierta forma, la noción de un error acotado.

Hoy en día se sabe que el juicio de Mayer sobre la relación inversa entre el número de ecuaciones usadas y la precisión de los resultados fue muy optimista. Ésta incrementa máximo hasta la raíz cuadrada del número de observaciones, todo bajo diversas suposiciones sobre las condiciones en las cuales son hechas. Fue hasta muchos años después que esta relación emergió en los trabajos de Laplace y Gauss.

Se puede expresar el razonamiento de Mayer con notación moderna, en el caso especial de determinar la media como sigue: Sea e el límite de precisión (análogo a la notación de Mayer $\pm x$) para una media \bar{X} (similar a la obtenida de $1^\circ 30'$), sea X_1 una de las medias encontradas (análogo a $1^\circ 40'$), y sea n la razón del número de observaciones entre \bar{X} y X_1 (similar a $9 = 27/3$). Entonces $|X_1 - \bar{X}|$ es análogo al valor obtenido de 10, pudiendo escribir entonces:

$$e = |X_1 - \bar{X}| / (n-1) = 10/8 = 1.25$$

Por supuesto, lo último representa una extrapolación formal de la intención original de Mayer, sin embargo, se puede apreciar la calidad de su trabajo hoy en día. El punto no es destacar que encontró un método particular para combinar veintisiete ecuaciones, sino que le dio un importante uso e interpretación. La aproximación de Mayer se aprecia de mejor forma cuando se le compara con el trabajo de Euler, publicado un año antes.

1.1.3. Saturno, Júpiter y Euler

Leonhard Euler (15 de abril de 1707 - 18 de septiembre de 1783) es conocido sobre todo por sus trabajos en análisis puro; sin embargo, también intervino en cada área de las matemáticas puras y aplicadas, conocidas hasta sus días. Euler fue el matemático más

PARTE I. ORÍGENES

prolífico de la historia. Fue premiado en varias ocasiones por diversas academias gracias a sus soluciones sobresalientes de problemas. En 1748, La Academia de Ciencias de Paris anunció un premio concerniente al segundo más grande problema científico mencionado al principio de este capítulo. Matemáticos de diversas áreas fueron invitados a preparar trabajos dando “Una teoría de Saturno y una de Júpiter, en la que cada una pueda explicar las irregularidades que causa cada planeta en el movimiento del otro, principalmente cerca de su tiempo de conjunción”. Este reto se le presentó a Euler cuando se encontraba en Berlín.

En 1676, Halley había verificado su sospecha de que los movimientos de Júpiter y Saturno estaban sujetos a una ligera perturbación. Cuando estas suposiciones eran comparadas con observaciones tabuladas de varios siglos antes, parecía que el movimiento principal de Júpiter era acelerado, mientras que el de Saturno era ligeramente retrasado. Halley fue capaz de mejorar la precisión de las tablas mediante un ajuste empírico, especulando que el movimiento irregular se presentaba debido a la atracción mutua de ambos planetas. Sin embargo, no fue capaz de dar una teoría matemática que pudiera comprobar esta suposición.

Este problema, a diferencia de aquel del movimiento lunar, representaba uno de atracción entre tres cuerpos. Contrariamente al problema de la Luna, éste no era comercialmente motivador. Más bien, su mayor ímpetu vino de nociones filosóficas sobre un sistema solar inestable. El problema establecido por la Academia fue interpretado como una extensión a las teorías existentes de atracción de tres cuerpos, con la *teoría gravitacional newtoniana* como punto de partida.

El trabajo de Euler sobre este difícil problema, fue considerado como el ganador del premio, aun cuando se quedó lejos de brindar una solución completa. Su escrito de ciento veintitres páginas titulado “Investigaciones sobre la cuestión de las desigualdades del movimiento de Saturno y Júpiter” fue publicado en Paris en el año de 1749 y sigue siendo utilizado hoy en día como modelo de una clara y ordenada exposición matemática. Comenzó suponiendo que estos planetas siguen órbitas circulares alrededor del Sol, y que éstas recaían sobre un mismo plano. Encontrando que la simple teoría que resultaba de esta suposición era insuficiente para resolver el problema, consideró hipótesis más complicadas. Primero, permitió que la órbita de Saturno fuese una elipse, continuando con la misma hipótesis para Júpiter. Finalmente, introdujo el hecho de que ambas no se encontraban sobre el mismo plano, formando entonces una ligera inclinación a ser encontrada con los cálculos.

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

Después de haber completado su análisis matemático, bastaba compararlo con su resultado empírico. Para hacer esta comparación, desarrolló una fórmula para la longitud heliocéntrica φ de Saturno:

$$\begin{aligned}\varphi = & \eta - 23525'' \text{sen } q + 168'' \text{sen } 2q - 32'' \text{sen } 2\omega \\ & - 257'' \text{sen}(\omega - q) - 243'' \text{sen}(2\omega - p) + m'' \\ & - x'' \text{sen } q + y'' \text{sen } 2q - z'' \text{sen}(\omega - p) \\ & - u(\alpha + 360v + p) \cos(\omega - p) + Nn'' \\ & - 0.11405k'' \cos q + (1/600)k'' \cos 2q\end{aligned}$$

En esta impresionante ecuación, los símbolos φ , η , q , ω , p , N y v estaban dados por observaciones y variaban en cada una. El resto de las variables: x , y , m , z , α , k , n y u eran correcciones fijas pero desconocidas, cuyos valores no habían sido especificados por la teoría. En esencia, la expresión representa la longitud heliocéntrica observada de Saturno (φ), siendo igualada a su media η (φ debería ser idéntica a η si se ignoran las perturbaciones y se supone una órbita elíptica) y a los términos de corrección. Éstos dependen de la diferencia de las longitudes entre los dos planetas (ω), el número de años desde 1582 (N), entre otras características (p y q eran las anomalías excéntricas, mientras que v era el número de órbitas completas de Júpiter desde 1582).

El problema que Euler enfrentaba, era que tenía setenta y cinco conjuntos completos de observaciones de φ , η , q , ω , p , N y v , hechas entre los años de 1582 a 1745. Primero obtuvo valores de n y u en los cuales tenía confianza. Restaba determinar las seis correcciones x , y , m , z , α y k para verificar si, cuando sus valores eran sustituidos en la ecuación de φ , los resultados obtenidos concordaban lo suficiente con los observados de φ , tal que le permitieran concluir que su teoría explicaba los movimientos de Saturno.

El problema era extraordinariamente difícil en su tiempo, y los intentos de Euler son a lo más admirables. Su trabajo es, en comparación con el de Mayer un año después, un fracaso estadístico. Después de haber encontrado valores para n y u , tenía los datos suficientes para producir setenta y cinco ecuaciones, todas lineales en x , y , m , z , α y k . Encontró la manera de conformar todas en seis sistemas, sin establecer combinaciones específicas sobre las ecuaciones -a diferencia de Mayer-.

Para entender cómo Euler trabajó con sus datos, se explica primero la forma en la que encontró dos incógnitas, n y u . Euler se dio cuenta que los coeficientes de todos los términos, con excepción a aquellos que involucraban valores de n y u , eran aproximadamente periódicos, en un lapso igual a 59 años. Entonces sustrajo la ecuación

para 1703 de aquella de 1585 (2 x 59 años), haciendo lo mismo con la de 1732 de la de 1673 (59 años), consiguiendo así, dos ecuaciones lineales únicamente para n y u ; resolvió éstas y comparó sus resultados con los de otros sistemas construidos de la misma forma, utilizando cuatro ecuaciones originales, todas ellas similares en espacio y tiempo.

Euler intentó evaluar otros factores de corrección por el mismo método, es decir, observando pequeños sistemas de ecuaciones formados con condiciones astronómicas similares, creando entonces una situación en donde muchos de los coeficientes eran aproximadamente iguales y la diferencia de ecuaciones anularía la mayoría de los términos. Sin embargo, no tuvo éxito en la construcción de otras situaciones -como la de n y u - donde diferentes conjuntos de ecuaciones dieran los mismos resultados, conformándose únicamente con dos incógnitas resueltas de esa manera. Una vez habiendo derivado sus seis sistemas inconsistentes con sólo dos incógnitas, declaró que: "Ahora, de estas ecuaciones no podemos concluir nada; por la razón de que he estado intentando satisfacer las observaciones de forma exacta, mientras que debería hacerlo sólo de manera aproximada; y este error mío se hizo más grande por sí solo." (Euler, 1749, p. 136)

Después de esto, Euler presentó veintiun ecuaciones involucrando las seis incógnitas, sin poder dar una solución al problema. Lo más lejos que llegó fue reducir a cero, cinco de las incógnitas -todas menos la variable cuyo coeficiente era unitario en todas las ecuaciones, m -, ajustó este término restante para que estuviera entre las más grandes y las más pequeñas constantes en las veintiun ecuaciones, haciendo la máxima discrepancia tan pequeña como fuese posible. En una perspectiva moderna, el lector quizá reflexione sobre la idea de datos contradictorios, si se considera que m tomaba veintiun valores diferentes. Sin embargo, se puede decir que la media denotada por \bar{m} , es simultáneamente (1) el estimador por mínimos cuadrados de m , y (2) la solución del sistema sumando las veintiun ecuaciones y resolviendo para m . La aserción (2), es un caso especial del *método de promedios*, cuando se trabaja con un sistema lineal inconsistente, encontrando un nuevo sistema lineal gracias a combinaciones lineales de las ecuaciones originales. Debido a que en este primer capítulo se desea dar un panorama histórico sin descuidar lo teórico, se dará una breve explicación de la aserción (1). Considérese que se tienen los valores m_1, m_2, \dots, m_{21} , todos ellos para m . Sea entonces:

$$L(m) = (m - m_1)^2 + (m - m_2)^2 + \dots + (m - m_{21})^2$$

la suma de los cuadrados de los errores. Tomando derivadas, se encuentra que $L''(m) = 42$ y $L'(m) = 0$ sólo cuando $m = \bar{m}$. En el siguiente capítulo se dará una

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

exposición detallada del método. A pesar de que Euler no resolvía su problema, implícitamente daba un buen punto de partida para la historia de los mínimos cuadrados.

La comparación entre las aproximaciones de Euler y Mayer es dramática. En 1750, Mayer se enfrentó a un conjunto de veintisiete ecuaciones inconsistentes con tres incógnitas, divisando un método sensible, combinándolas todas en tres ecuaciones y resolviendo para las variables. En 1749, Euler trabajó con pequeños conjuntos de ecuaciones -tantos como variables tuviese-, y sólo aceptaba resultados numéricos cuando diferentes conjuntos le daban esencialmente los mismos. Su problema fue similar al de Mayer, siendo este último el único capaz de dar una solución estadística a su problema. Una combinación estratégica de ecuaciones que Euler no pudo ver.

Existió una esencial diferencia conceptual en sus aproximaciones que hizo imposible para Euler adoptar una actitud estadística. Mayer trató su problema como un astrónomo práctico, trabajando con observaciones que él mismo había realizado considerando similares condiciones observacionales. Euler por su parte, abordó su problema como un matemático puro, trabajando con observaciones hechas por otros a través de los siglos bajo condiciones desconocidas. Mayer le dio un sentido de aleatoriedad a los errores, aproximando el problema con la firme convicción de que distintas combinaciones incrementarían la precisión de sus resultados en proporción al número de ecuaciones combinadas. Euler desconfiaba de las combinaciones de ecuaciones, basándose en un punto de vista matemático donde el error se incrementa cuando se agregan ecuaciones; contrario al sentido estadístico, en que los errores aleatorios tienden a cancelarse entre sí. Fue hasta 1750 que el avance conceptual de combinar ecuaciones observacionales comenzó a aparecer como punto de partida; y con ello, de forma inconsciente, un ligero esbozo de los mínimos cuadrados. Mayer no hizo el intento por describir sus cálculos como un método que sería aplicable a otros problemas. Fue Laplace el encargado de desarrollar en forma, lo que sería conocido durante el siglo XIX como “El Método de Mayer”.

1.1.4. El rescate del Sistema Solar de Laplace

Pierre Simon de Laplace (23 de marzo de 1749 - 5 de marzo de 1827) fue el científico más ilustre en la historia de Francia, nombrado por Poisson como “*Le Newton de la France*”. Fue miembro de la Academia de Ciencias, desarrollando distintos cargos importantes en diversas universidades. Cerca de la mitad de su trabajo está dedicado a la *Mecánica Celeste*, cerca de un cuarto está basado exclusivamente en la *Teoría de Probabilidad* y el

PARTE I. ORÍGENES

resto es una mezcla de Probabilidad y Física. Es mejor conocido en la historia de la ciencia por sus dos más grandes tratados: “*Traité de Mécanique Céleste*” (cuatro volúmenes, 1799-1805, con un quinto suplementario en 1825) y “*Théorie analytique des probabilités*” (1812).

En 1787, mediante un trabajo sobre los movimientos de Saturno y Júpiter, Laplace propuso lo equivalente a una extensión del método de Mayer de reciclar ecuaciones lineales inconsistentes. Finalmente puso fin al problema de las desigualdades en el movimiento de estos planetas mostrando que éstas eran, en efecto, periódicas (de un lapso muy grande). En el transcurso de su demostración se enfrentó a ecuaciones similares a aquellas que impidieron a Euler llegar a una solución. El éxito de Laplace recae en que su análisis de datos sugirió mediante la teoría; hipótesis que iban más lejos del desarrollo teórico y la confirmación observacional de los resultados. Buscó encontrar las respuestas a estos movimientos dentro de los límites de la teoría gravitacional de Newton y de la precisión en las observaciones. Notó también que las perturbaciones periódicas en el movimiento de los planetas correspondían a los términos que involucraban expresiones iguales a $5m - 2n$, debiendo tener un período de 900 años. Esencialmente, m y n representan los movimientos anuales promedio medidos en grados de Saturno y Júpiter respectivamente. Una propiedad establecida por la teoría gravitacional sugería que estas expresiones estarían cerca del valor $5/2$; en efecto, hoy en día se sabe que $5m - 2n = 0.40758$.

Los resultados de este análisis fueron dados a conocer en un trabajo de ciento veintisiete páginas llamado “*Théorie de Jupiter et de Saturne*” por la Academia de Ciencias en 1787. La clave de la investigación de Laplace fue la comparación de su teoría con las observaciones. Hizo uso de los mejores datos disponibles sobre los movimientos de los planetas; sin embargo, su teoría requirió cuatro variables extras, que Laplace intuyó con base en los mismos datos. Estas cantidades expresadas con la notación original de Laplace son: $\delta\varepsilon^I$, δn^I , δe^I , $\delta\tilde{\omega}$ y $\delta\tilde{\omega}$; ellas indican necesariamente correcciones (tasas de cambio) para, respectivamente, la longitud media de Saturno en 1750, su movimiento anual promedio, su excentricidad, y la posición de afelio (su posición más distante del Sol). Laplace seleccionó veinticuatro observaciones de Saturno, hechas en el momento de oposición (cuando el Sol, la Tierra, y Saturno están alineados) durante un período de 200 años. En cada caso, Laplace expresó la diferencia entre la longitud observada de Saturno y la que estaba dada por su teoría como una *ecuación de condición*. Por ejemplo, la correspondiente para el año de 1672 estaba dada por:

$$0 = -3'32.8'' + \delta\varepsilon^I - 77.28\delta n^I - 2\delta e^I 0.98890 - 2e^I (\delta\tilde{\omega} - \delta\varepsilon^I) 0.14858$$

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

Una excentricidad especial denotada como e^l fue estudiada en otra parte de su trabajo en el cual no se entrará en detalles. El conjunto completo de ecuaciones está dado en la Tabla 1.3, en la cual $-a_i$ es el primer término de cada ecuación ($a_i = -3'32.8''$ para la ecuación anterior), y b_i , c_i , d_i son los coeficientes de las variables δn^l , $2\delta e^l$, y $2e^l(\delta\tilde{\omega}^l - \delta\varepsilon^l)$. Los residuales son la diferencia entre los valores ajustados menos los observados. Laplace se dio cuenta que sus coeficientes estaban fuertemente relacionados con aquellos de la investigación de Mayer, debido a que $c_i^2 + d_i^2 = 1$. En efecto, $c_i = -\text{sen}(\varphi_i - \omega_i)$ y $d_i = \text{cos}(\varphi_i - \omega_i)$, donde $\varphi_i - \omega_i$ es la diferencia entre la longitud observada de Saturno y su afelio en el i -ésimo año; b_i es el número de años de 1750 al tiempo en que fue hecha la i -ésima observación.

Eq. No. (i)	Año	$-a_i$	b_i	c_i	d_i	Residuales		
						Laplace	Halley	Mínimos C.
1	1591	1'11.9"	-158.0	0.22041	-0.97541	+1'33"	-0'54"	+1'36"
2	1598	3'32.7"	-151.78	0.99974	-0.02278	-0.07	+0.37	+0.05
3	1660	5'12.0"	-89.67	0.79735	0.60352	-1.36	+2.58	-1.21
4	1664	3'56.7"	-85.54	0.04241	0.99910	-0.35	+3.20	-0.29
5	1667	3'31.7"	-82.45	-0.57924	0.81516	-0.21	+3.50	-0.33
6	1672	3'32.8"	-77.28	-0.98890	-0.14858	-0.58	+3.25	-1.06
7	1679	3'9.9"	-70.01	0.12591	-0.99204	-0.14	-1.57	-0.08
8	1687	4'49.2"	-62.79	0.99476	0.10222	-1.09	-4.54	-0.52
9	1690	3'26.8"	-59.66	0.72246	0.69141	+0.25	-7.59	+0.29
10	1694	2'4.9"	-55.52	-0.07303	0.99733	+1.29	-9.00	+1.23
11	1697	2'37.4"	-52.43	-0.66945	0.74285	+0.25	-9.35	+0.22
12	1701	2'41.2"	-48.29	-0.99902	-0.04435	+0.01	-8.00	-0.07
13	1731	3'31.4"	-18.27	-0.98712	-0.15998	-0.47	-4.50	-0.53
14	1738	4'9.5"	-11.01	0.13759	-0.99049	-1.02	-7.49	-0.56
15	1746	4'58.3"	-3.75	0.99348	0.11401	-1.07	-4.21	-0.50
16	1749	4'3.8"	-0.65	0.71410	0.70004	-0.12	-8.38	+0.03
17	1753	1'58.2"	3.48	-0.08518	0.99637	+1.54	-13.39	+1.41
18	1756	1'35.2"	6.58	-0.67859	0.73452	+1.37	-17.27	+1.35
19	1760	3'14.0"	10.72	-0.99838	-0.05691	-0.23	-22.17	-0.29
20	1767	1'40.2"	17.98	0.03403	-0.99942	+1.29	-13.12	+1.34
21	1775	3'46.0"	25.23	0.99994	0.01065	+0.19	+2.12	+0.26
22	1778	4'32.9"	28.33	0.78255	0.62559	-0.34	+1.21	-0.19
23	1782	4'4.4"	32.46	0.01794	0.99984	-0.23	-5.18	-0.15
24	1785	4'17.6"	35.56	-0.59930	0.80053	-0.56	-12.07	-0.57

Tabla 1.3. Ecuaciones de condición de Laplace. [17] pp. 34

PARTE I. ORÍGENES

Laplace se enfrentó con veinticuatro ecuaciones de condición inconsistentes, cada una de ellas lineal en cuatro variables. Se encontraba en una situación similar en donde Euler había fallado, y en donde Mayer había resuelto lo suyo dividiendo las ecuaciones en grupos con similares características para, finalmente, reducirlos mediante sumas y resolver su sistema de tres ecuaciones con tres incógnitas. Laplace trató su problema como una semblanza superficial a la aproximación de Mayer; sin embargo, difería en aspectos sutiles que marcan un importante avance hacia los mínimos cuadrados. Laplace no dio una descripción algebraica de su solución, pero sí sobre los pasos que siguió. Lo que hizo, fue reducir sus veinticuatro ecuaciones lineales a cuatro ecuaciones: (i) la suma de las ecuaciones 1-24; (ii) la diferencia entre la suma de las ecuaciones 1-12 y la suma de las ecuaciones 13-24; (iii) la combinación lineal de las ecuaciones $-1+3+4-7+10+11-14+17+18-20+23+24$; y (iv) la combinación lineal de las ecuaciones $+2-5-6+8+9+12-13+15+16-19+21+22$. Laplace resolvió estas cuatro ecuaciones y comparó sus resultados con aquellos obtenidos de las ecuaciones ajustadas de las observaciones; todo esto trabajando con cada residual, definido como el exceso del valor ajustado sobre el observado -el negativo de la definición moderna de residual-.

Laplace no dio una explicación del porqué de esta combinación, pero parece haberla realizado con base en el efecto que los coeficientes tenían sobre las variables de las ecuaciones de la (i) a la (iv). En efecto, (i) y (ii) son combinaciones lineales naturales a considerar: en (i) se maximiza el coeficiente del término constante, mientras que (ii) lo elimina. Exactamente lo mismo para el coeficiente del término $\delta\varepsilon^l$ y casi inversamente para el del término δn^l debido a la naturaleza en que los años fueron medidos. Evidentemente, la elección sobre que ecuaciones deberían ser incluidas en (iii) y cuales en (iv), requirió de un profundo análisis algebraico usando el hecho de que si $|c_i| < |d_i|$ o $|c_i| > |d_i|$. Una vez que las veinticuatro ecuaciones fueron divididas entre (iii) y (iv), los signos $+$ y $-$ de acuerdo a los signos de d_i [para la ecuación (iii)] y c_i [para la ecuación (iv)], intentando maximizar el contraste entre los coeficientes de las dos últimas variables.

El sutil avance de Laplace fue que, mientras Mayer había sumado ecuaciones de condición en grupos disjuntos, combinó las mismas ecuaciones juntas en distintas formas. La relación entre el método de Mayer y Laplace puede ser comprendida de mejor manera si se trasladan todos los términos a una notación moderna. Escribiendo una ecuación de condición involucrando cuatro variables incluyendo una constante se tiene:

$$0 = a_i + \omega + b_i x + c_i y + d_i z, \quad i = 1, \dots, n$$

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

donde a_i , b_i , c_i y d_i son observables y ω , x , y y z son variables. Entonces se puede escribir la j -ésima ecuación agregada (combinación mediante sumas y restas de n ecuaciones), $j=1,2,3,4$, como:

$$0 = \sum_i k_{ij} a_i + \sum_i k_{ij} \omega + \sum_i k_{ij} b_i x + \sum_i k_{ij} c_i y + \sum_i k_{ij} d_i z$$

donde $\{k_{ij}\}$ forma un sistema de multiplicadores.

Mayer trató el caso $n=27$ y $d_i=0$ para toda i (es decir, cuando sólo tres incógnitas aparecen) y determinó los multiplicadores $\{k_{ij}\}$ para $j=1,2,3$ con base en la naturaleza de los coeficientes que se denotarán como π_i , de la variable α ($b_i x = \pi_i \alpha$). Naturalmente, el objetivo de esta comparación no es argumentar que Laplace y Mayer casi llegan a descubrir el método de los mínimos cuadrados, sino preparar al lector a lo que Legendre tiene reservado. El método de Legendre a diferencia de los anteriores, fue formalmente derivado de un criterio explícito para un mejor ajuste. Mayer separó el conjunto de datos en grupos disjuntos, con base en el coeficiente de una de sus variables; en esto fue más lejos que Euler, quien mantuvo la postura de que éstos tendrían que ser parecidos antes de ser combinados. Laplace fue más allá, combinando todas sus ecuaciones de condición y, lo más importante, dejó que todos los coeficientes influyeran en la elección de estas combinaciones. Otro punto importante a destacar, es que Mayer tuvo éxito debido a que su trabajo estaba enfocado a una aplicación particular. Otro investigador imitando su procedimiento podría no tener tanta suerte. Por otro lado, la generalización del método de Laplace no sufre de este inconveniente.

Los esfuerzos de Mayer y Laplace son hoy valorados en términos de precisión. Por ejemplo, cuando se compara la raíz cuadrada de la suma de los cuadrados de los residuales de Mayer -para sus datos de la libración de la Luna-, con aquella obtenida mediante mínimos cuadrados, se encuentra que el valor por el método de Mayer es sólo 6% más grande que el arrojado por los mínimos cuadrados. Para los datos de Saturno por parte de Laplace es hecha la misma comparación, siendo tan sólo 5.5% más grande el valor por el método de éste último que aquel dado por mínimos cuadrados.

En esta descripción, en donde se intenta explicar la extensión del método de Laplace con respecto al de Mayer, existe una suposición implícita sobre el cambio de los signos en los coeficientes, como debería ser el caso si estos se centran en sus medias. En efecto, estas primeras aplicaciones lo son, debido a que cada una de las variables representa una corrección de alguna media. Muchos autores hicieron mención del método de Laplace a lo largo del siglo XIX, siempre describiéndolo como una extensión a la formulación de Mayer.

1.1.5. Boscovich y la figura de la Tierra

Hasta ahora, se ha visto cómo los problemas que envolvían a la Luna, Júpiter y Saturno, introdujeron y desarrollaron métodos para combinar ecuaciones lineales inconsistentes, medio siglo antes de la aparición de los mínimos cuadrados. El método de Mayer fue fácil de usar, y sobre todo, arroja resultados aceptables. Sin embargo, carecía de una cualidad de la cual se encontró en abundancia en el método de mínimos cuadrados. El primero no era “mejor”, debido a que apareció como solución a un problema matemático específico, intentando encontrar la “mejor” combinación de ecuaciones inconsistentes. Las tablas de Mayer del movimiento lunar y su mapa de la cara de la Luna, fueron comúnmente vistos como el más preciso de los trabajos en el manejo de datos observacionales del siglo XVIII. Fue Laplace el primero en desarrollar el método de Mayer como una herramienta más general. Conforme pasaron los años, la impresión de estos hechos se fue desvaneciendo, siendo remplazada por los triunfos de los mínimos cuadrados.

Los mínimos cuadrados fueron el más exitoso de los primeros métodos para combinar ecuaciones inconsistentes, y los hechos de que estuvieron basados en principios fáciles de entender y contar con un criterio objetivo, fueron las razones de su éxito. Sin embargo, para entender su génesis, hace falta revisar al más famoso de sus predecesores: “El método de Boscovich”. Se debe considerar entonces el tercer problema científico más grande del siglo XVIII, aquel sobre la figura de la Tierra.

La primera pista de que la Tierra no era una esfera perfecta, parece haber sido descubierta por Richer en 1672, observando que un péndulo cerca del ecuador no se ve afectado de la misma manera que el mismo en Paris. Newton en su *Principia* (1687), mostró cómo la rotación de la Tierra podría ser causante de un achatamiento en los polos y alargar el ecuador, una forma conocida como esferoide oblato que, en la actualidad, su parametrización en coordenadas cartesianas está dada por $\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1$, con $a > c$.

Sin embargo, Newton estaba casi seguro de que $a \approx c$. Domenico Cassini, director del Observatorio Real de Paris, aseguraba que los achatamientos estaban presentes en el ecuador, no en los polos.

Los dos métodos principales para determinar la forma de la Tierra eran mediante péndulos o medidas de arco. La realización de éstas requirió la cooperación de un equipo de científicos y meses de labor bajo condiciones adversas; fue un reto perfecto para la creciente comunidad científica francesa. La idea consistía en medir longitudes de arco del

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

tamaño de un grado de latitud -o dos a lo más- entre dos puntos contiguos de la superficie. Si una medición cerca del ecuador era más pequeña que una cerca del polo, entonces la esfera de la Tierra estaría achatada en los polos; y la diferencia entre estas dos mediciones podría ser usada para calcular el grado de achatamiento.

La latitud de un punto sobre la superficie de la Tierra no es el ángulo formado entre dos rayos del centro de la Tierra, uno hacia el punto dado y el otro siendo la intersección del plano ecuatorial con el plano meridiano del punto en cuestión; sino es tomada como el ángulo entre el rayo del cenit del punto dado y el plano ecuatorial, o en forma alternativa, el complemento del ángulo entre el cenit y la Estrella Polar. El cenit es el punto más alto en el cielo con relación al observador que se encuentra justo sobre su cabeza (90°). La Figura 1.2 muestra arcos de 10 grados de latitud en una Tierra oblatada exagerada.

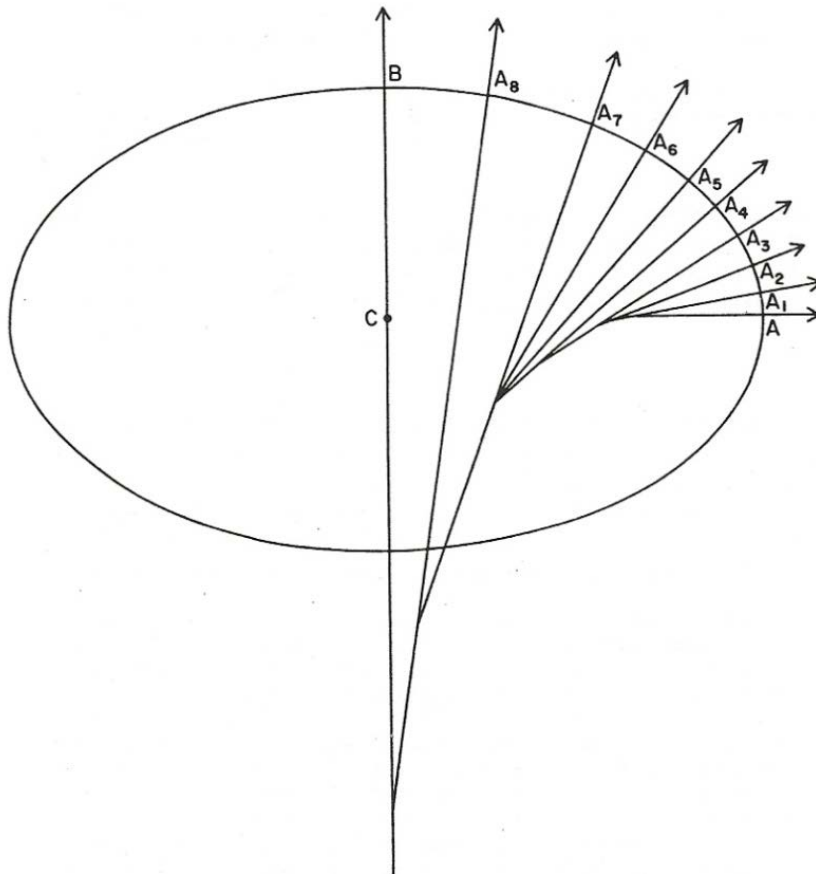


Figura 1.2. La forma oblatada de la Tierra. [17] pp. 41

La relación aproximada entre la longitud de un arco y la latitud puede ser derivada haciendo uso de secciones cónicas; la relación exacta está dada por una integral elíptica de la cual no se darán detalles en este trabajo. Sin embargo, se hace una aproximación

PARTE I. ORÍGENES

simple: Si a es la longitud de arco de un grado centrado en la latitud θ , medido a través de su meridiano, entonces una buena aproximación es $a = z + y \text{sen}^2 \theta$, donde z es la longitud de arco centrado en el ecuador, mientras que y es el exceso -o deficiencia- de la longitud del arco de un grado de latitud medido con respecto al Polo Norte sobre otro medido con respecto al ecuador. Otra notación expresa esto como: $a = z + y \cdot \frac{1}{2} \cdot \text{versoseno}(2\theta)$, donde $\text{versoseno}(2\theta) = 2 \text{sen}^2 \theta$.

En 1735, La Academia Francesa lanzó expediciones con el fin de realizar medidas de arco cerca del ecuador y hasta lugares de alrededor de 66° de latitud, para compararlas con las realizadas en Paris. Los resultados refutaron de inmediato la hipótesis de Cassini, dándole la razón a Newton: La Tierra era oblata. La única cuestión restante era calcular la elipticidad, debido a que distintos pares de mediciones daban resultados diferentes. En 1755, éstos fueron publicados por Roger Boscovich (18 de mayo de 1711 – 13 de febrero de 1787) nacido en la República de Ragusa (hoy Dubrovnik, Croacia). A continuación se analiza cómo obtuvo éxito en la resolución de inconsistencia en las medidas de arco, así como la introducción a un procedimiento estadístico que es lo que concierne en este apartado.

Boscovich estaba consciente de que, para determinar la figura de la Tierra, sería necesario comparar medidas independientes en latitud; tomando en cuenta que aún pequeños errores -debido a la aproximación a sus longitudes de arco- podrían exagerar de forma drástica los resultados obtenidos para cada una de sus combinaciones. Entonces puso su atención en sólo cinco de las medidas que fueron hechas en diferentes lugares y que eran las más confiables en cuanto a precisión. Estos datos aparecen en la Tabla 1.4.

Sitio	Latitud (θ)	Longitud Del arco (toesas)	$\text{sen}^2 \theta \times 10^4$ de Boscovich
(1) Quito	$0^\circ 0'$	56,751	0
(2) Cape of Good Hope	$33^\circ 18'$	57,037	2,987
(3) Rome	$42^\circ 59'$	56,979	4,648
(4) Paris	$49^\circ 23'$	57,074	5,762
(5) Lapland	$66^\circ 19'$	57,422	8,386

Tabla 1.4. Datos de Boscovich. [17] pp. 43

Boscovich no dio una descripción analítica para el manejo de estos datos, sino que siguió una tradicional descripción geométrica atribuida a Newton. A pesar de ello, debe ser fácil

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

para el lector comprender su trabajo si todo se traduce a una descripción analítica moderna. En términos analíticos, Boscovich se enfrentó a cinco ecuaciones observacionales -o *de condición*- equivalentes con la siguiente forma:

$$a_i = z + y \operatorname{sen}^2 \theta_i,$$

donde a_i y θ_i son la longitud de un arco medido en toesas por grado (1 toesa \cong 6.39 pies) y la latitud del punto medio del arco, ambos en el sitio i . Las incógnitas y y z son, respectivamente, el exceso de un arco de un grado en el polo sobre otro medido en el ecuador, y la longitud de éste último.

En principio, dos de los cinco sitios -cualquiera que éstos sean-, podrían ser usados para resolver las ecuaciones y obtener el valor de las variables y y z , o equivalentemente, el exceso y y la elipticidad definida por Boscovich como $1/\text{elipticidad} = 3z/y$. De hecho, eso fue lo que Boscovich hizo: calculó y y la elipticidad tomando cada una de las $\binom{5}{2} = 10$ parejas y presentando los resultados mostrados en la Tabla 1.5. Esto le dio no una, sino diez soluciones al mismo problema. Lo que hizo entonces fue obtener el valor promedio de los diez valores de y y encontrar, usando el valor de z en Quito ($z = 56,751$), el valor de $1/155$ para la elipticidad. Este valor debió haberle parecido demasiado grande, por lo que hizo un nuevo cálculo. Esta vez rechazando los pares (2,4) y (2,3), definiéndolos como “diferentes a los otros”. El nuevo resultado arrojó una elipticidad de $1/198$, siendo aún insatisfactorio para Boscovich. En vez de comprometerse con uno u otro, centró su atención en la discrepancia que existía entre la media y sus componentes originales; discrepancia que lo llevó a pensar que sus datos eran inconsistentes con la hipótesis elipsoidal de Newton.

Pareja	Exceso polar		Pareja	Exceso polar	
	(y , en toesas)	Elipticidad		(y , en toesas)	Elipticidad
1, 5	800	1/213	2, 4	133	1/128
2, 5	713	1/239	3, 4	853	1/200
3, 5	1,185	1/144	1, 3	491	1/347
4, 5	1,327	1/128	2, 3	-350	-1/486
1, 4	542	1/314	1, 2	957	1/78

Tabla 1.5. Cálculos de Boscovich. [17] pp. 45

Boscovich sintió una tremenda inquietud de sus propias conclusiones, situación que le impidió finalizar su análisis de 1755. Dos años más tarde publicó una sinopsis que incluía una breve descripción de un nuevo principio radical para la combinación de medidas de arco inconsistentes. En 1760, dio una detallada descripción de esto, y una explicación de como podía ser usado en la práctica trabajando sobre los cinco grados que él había considerado en 1755. En este trabajo de 1760, Boscovich comenzó introduciendo su principio como sigue:

“Dado un cierto número de grados de latitud, para encontrar las correcciones que deberían hacerse en cada uno de ellos, supongamos que se deben cumplir tres condiciones: la primera, que sus diferencias deberían ser proporcionales a las diferencias de los *senoversos* de dos veces sus latitudes; la segunda, que la suma de las correcciones positivas debería ser igual a la suma de las negativas; la tercera, que la suma de todas las correcciones sin signo, debería ser mínima.” (Boscovich, 1760, p.501)

La primera condición se deriva de una propiedad de las figuras elípticas llamada ley del equilibrio. La segunda, del hecho de que los errores del observador aumentan o disminuyen con la misma probabilidad. Y la tercera, es con la intención de acercarse tanto como sea posible a los valores reales. Para comprender mejor estas tres condiciones, nuevamente se lleva todo a una forma analítica.

Introduciendo los símbolos a_1, a_2, \dots para las longitudes de arco en toesas por grado medidas en las latitudes $\theta_1, \theta_2, \dots$, y sea δa_i la “corrección” introducida por Boscovich para a_i , entonces la primera condición establece que las medidas corregidas $a_i + \delta a_i$ satisfacen:

$$a_i + \delta a_i - (a_j + \delta a_j) \propto \text{senoverso}(2\theta_i) - \text{senoverso}(2\theta_j)$$

Debido a que $\text{senoverso}(2\theta) = 2\text{sen}^2 \theta$, y recordando como se había definido a la variable a_i en su trabajo de 1755, se puede escribir entonces $a_i + \delta a_i = z + y\text{sen}^2 \theta_i$ o equivalentemente $\delta a_i = z + y\text{sen}^2 \theta_i - a_i$, para alguna elección de z y y . La segunda condición estaba relacionada a la noción de que los errores eran igualmente probables, intuyendo que la suma de las correcciones positivas debería ser igual a la de las negativas. Analíticamente, esto puede ser expresado como: $\sum_{\forall i} \delta a_i = 0$. La tercera condición de

Boscovich era que la suma de las correcciones sin signo debería ser un mínimo; esto es: $\sum_{\forall i} |\delta a_i|$ es minimizado.

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

Se debe decir ahora que algunas de estas formulaciones analíticas no son encontradas en el trabajo de Boscovich. La explicación de su principio fue dada sólo de manera verbal, tratando su problema desde un punto de vista principalmente geométrico. A pesar de ello, Boscovich se dio cuenta que podría aproximarse a su primera condición de forma analítica, involucrando las funciones *senoverso* originales. La segunda expresión podría ser expresada haciendo uso de la primera, involucrando los coeficientes desconocidos y y z . Su declaración fue nuevamente verbal, pero claramente él tenía la ecuación

$$\sum_{\forall i} (z + y \operatorname{sen}^2 \theta_i - a_i) = 0$$

en mente. Sin embargo, sintió equivocadamente, que una solución analítica podría ser encontrada sólo diferenciando esta expresión con respecto a los coeficientes desconocidos, viendo que esto sería absurdo: “Con suponer que $dz = 0$ no llegaremos a nada, la ecuación desaparecerá por completo, junto con la esperanza del matemático”. Fue entonces cuando buscó la manera de encontrar una solución geométrica.

Esta extensión de 1760, contaba ahora con un diagrama mostrado aquí en la Figura 1.3. Se considera a AF como una unidad de longitud; A es el origen; y A, B, C, D y E representan los valores de $\operatorname{sen}^2 \theta_i$, viviendo todos en el intervalo unitario. Los cinco arcos son indicados por a, b, c, d , y e . G es el centro de gravedad. Las longitudes Aa, Bb, Cc, Dd y Ee representan los valores -en toesas por grado- de las correspondientes medidas de arco. El problema era encontrar la línea recta $A'H$ tal que las correcciones aA', bO, cK, dL y eM cumplieran la segunda y tercera condición. Boscovich comenzó observando que la segunda, interpretada también como una condición de equilibrio mecánico, significaba que la línea debía pasar por el centro de gravedad de los puntos (G). Esto reducía el problema a encontrar esa línea a través de G tal que la tercera condición se cumpliera.

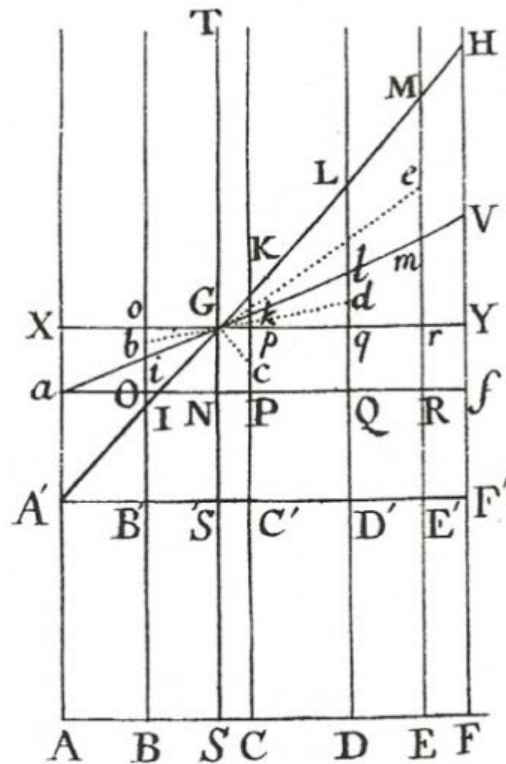


Figura 1.3. Diagrama de Boscovich. [17] pp. 48

Después de una serie de procedimientos geométricos, los cuales rebasan el objetivo de este trabajo, Boscovich encontró que la solución era la línea que pasaba por a y G , representada en el diagrama como aV . Después de esto, encontró los valores correspondientes de $y = 692$ y $z = 56,751$, con los cuales obtuvo una elipticidad de $1/248$. Aquí, en 1760, Boscovich usó la fórmula $1/\text{elipticidad} = 3z/y + 2$. Satisfecho quizá por su cercana aproximación al valor de Newton de $1/230$, y del hecho de que su nuevo valor emergió de su propio nuevo método de combinar observaciones, Boscovich no repitió sus primeras dudas de 1755 sobre la figura elíptica de la Tierra.

Después de estas conclusiones de 1760, Boscovich agregó una discusión actualizada sobre su método en su extensión de 1770. Esta vez utilizó nueve medidas de arco haciendo su análisis después de haber omitido las tres mediciones más discordantes, llegando a la conclusión de que sus resultados soportaban la hipótesis de que la Tierra estaba constituida sobre la forma elipsoidal. Boscovich no dio más desarrollo de su método, estudio de sus propiedades, formulación analítica, o alguna aplicación a otros problemas que no fuesen la figura de la Tierra; nunca intentó generalizar su método. En efecto, su principio permaneció en la oscuridad durante mucho tiempo, hasta que llamó la atención de Laplace.

1.1.6. Laplace y el Método de Situación

En 1789, Laplace retomó la cuestión de la figura de la Tierra por segunda ocasión, aunque ya había considerado el problema de comparar medidas de arcos con una hipótesis elipsoidal en 1783, pero en una primera instancia, ignorando el trabajo de Boscovich, planteó sólo cuatro de las cinco mediciones estudiadas por éste (descartando la realizada en Roma). Habiéndose contentado con determinar el mejor ajuste para una figura elíptica de la cual había minimizado la máxima corrección involucrada, en 1789 introdujo un algoritmo de su propia invención teniendo a Boscovich como figura principal.

Primero, derivó e implementó sin prueba una primera versión de su algoritmo, mostrando su uso con nueve medidas de arco (ahora incluía la de Roma); después, se dirigió al procedimiento que más tarde fue conocido como el *método de Boscovich*. Laplace sintió que la figura elíptica que minimizaba la máxima corrección debía cumplir las siguientes dos condiciones:

“... 1° que la suma de los errores sea cero; 2° que la suma de los errores tomados todos con el signo positivo sea mínima. Boscovich ha dado un ingenioso método en el que se basa esto; y es explicado en su versión francesa *Voyage astronomique et géographique* (1770); pero desde que ha complicado el método de forma innecesaria con diagramas, lo que presentaré aquí es su forma analítica más simple.” (Laplace, 1973, p.32)

Laplace procedió a hacer justamente eso, dando una precisa declaración del algoritmo de Boscovich acompañando su descripción con una demostración analítica rigurosa. También dio dos ejemplos numéricos, uno involucrando nueve medidas de arco y el otro considerando trece movimientos de péndulos realizados en diferentes latitudes. En este tratado de 1789, Laplace agregó una formulación analítica a la primera presentación de Boscovich. Este paso fue crucial para él en los estudios posteriores de las propiedades estadísticas del método -las que le daría el nombre de “Método de Situación”-, sin embargo, sólo se trataba de un pequeño avance conceptual, una transición del lenguaje de Newton al de Euler. A pesar de ello, diez años después Laplace dio un paso adicional, un sutil desarrollo de la idea de Boscovich que fue un claro síntoma de su creciente intuición estadística.

El renovado interés de Laplace por la figura de la Tierra dio origen al segundo volumen de su *Mécanique Céleste*. Después de desarrollar una teoría matemática de las figuras de cuerpos celestiales, Laplace regresó al problema que había enfrentado una década antes, aquel de comparar mediciones de arcos con la hipótesis de que la figura de la Tierra era

PARTE I. ORÍGENES

elipsoidal. Su análisis consideraba sólo siete arcos; omitiendo una vieja medida realizada en Holanda y sustituyéndola por una hecha en Francia en 1795.

Las ecuaciones de condición en su notación original describen la siguiente relación:

$$a_k - z - p_k y = x_k ,$$

donde a_k es la longitud del k -ésimo arco -en toesas por grado-, z es la longitud de un grado en el ecuador, y es el exceso polar, p_k el cuadrado del seno de la k -ésima latitud, y x_k el error, debido a la medición o al no cumplimiento de la hipótesis elipsoidal. El valor de a_k fue determinado midiendo la longitud del arco por tierra -sólo de una décima a una quinta parte fue medida con cadenas; la longitud total era obtenida a través de procedimientos geométricos- y mediante observaciones astronómicas -en grados- de la Estrella Polar desde los puntos extremos del arco, siendo el cociente de estas dos medidas el valor a_k en toesas por grado. Las longitudes de los arcos variaban considerablemente, desde tan sólo un grado hasta cerca de diez. Seguramente estas longitudes afectarían la precisión, sin embargo, deberían ser incorporadas al análisis; por lo tanto, en su *Mecánica Celeste (vol. 2)*, Laplace modificó las primeras condiciones de Boscovich de la siguiente forma:

“Primero, la suma de los errores involucrados en las medidas de arcos enteros debería ser cero. Segundo, la suma de todos estos errores, tomados positivamente, debería ser un mínimo. Considerando de esta manera los arcos completos, en vez de los grados que había deducido de ellos, deberíamos darle a cada uno de estos grados una mucha mayor influencia, calculando la figura elíptica de la Tierra de acuerdo a la correspondiente extensión del arco medido en grados.” (Laplace, 1799-1805, vol. 2, p. 134)

Siguiendo a Laplace, i_k representa entonces la longitud, en grados, del k -ésimo arco -la cantidad $i_k a_k$ es la longitud total del arco medido en toesas-. Entonces, los arcos completos satisfacen $i_k a_k - i_k z - i_k p_k y = i_k x_k$, siendo entonces las nuevas condiciones de Laplace:

$$(i) \quad \sum_k i_k x_k = 0$$

$$(ii) \quad \sum_k |i_k x_k| = \min .$$

Laplace derivó la solución del problema mediante una simple modificación. La condición (i) es equivalente a $A - z - yP = 0$, donde

1.1. ORÍGENES Y ANTECEDENTES DE LOS MÍNIMOS CUADRADOS

$$A = \frac{\sum i_k a_k}{\sum i_k}, \quad P = \frac{\sum i_k p_k}{\sum i_k}$$

Sustrayendo esta ecuación de cada una de las de condición, el resultado es: $b_k - yq_k = x_k$, donde:

$$b_k = a_k - A, \quad q_k = P - p_k$$

Entonces, Laplace siguió la idea de Boscovich, intentando encontrar la línea que pasara por un punto $G^* = (P, A)$, haciendo uso de la condición (i) para reducir el problema a encontrar una sola variable, la pendiente y de la línea que pasaría por el centro de gravedad de los pesos, denotado por G^* .

El siguiente paso en el algoritmo fue suponer que las ecuaciones son indexadas con respecto a la sucesión decreciente b_k / q_k , esto es:

$$\frac{b_1}{q_1} \geq \frac{b_2}{q_2} \geq \dots \geq \frac{b_n}{q_n}$$

Finamente, tomando $h_k = |i_k q_k|$ y $F = h_1 + \dots + h_n$, entonces si r es el índice tal que

$$h_1 + \dots + h_{r-1} < \frac{1}{2}F \quad \text{y} \quad h_1 + \dots + h_r > \frac{1}{2}F,$$

la solución del problema fue tomar $y = b_r / q_r$ y $z = A - Py$. El algoritmo es una versión analítica del procedimiento geométrico de Boscovich, pero esta vez, considerando la longitud del arco completo. Laplace aplicó su método para siete mediciones, concluyendo con una elipticidad de $1/312$. Sin embargo, un error tal grande de 172.52 toesas en el arco de Lapland (Finlandia) era evidencia de que la Tierra no era elipsoidal (Laplace, 1799-1805, vol.2, pp. 138-141).

El análisis ponderado de Laplace parece ahora un pequeño pero significativo avance en la técnica estadística. Los trabajos anteriores habían asignado diferentes pesos a las observaciones de acuerdo a la longitud de sus valores. James Short por ejemplo, en 1763, había disminuido sus mediciones solares conforme a qué tan lejos se encontraban de la media. Otros científicos habían descartado sus medidas discordantes. Análisis como los de Mayer y Boscovich daban la impresión de asignar grandes pesos a observaciones tomadas bajo condiciones extremas, cerca del polo y del ecuador. La Regla de Cotes, explicada al comienzo de este capítulo, es otro ejemplo de asignación de pesos. Diferente a todos los anteriores, Laplace otorgó pesos a las mediciones de acuerdo a la longitud de los arcos.

PARTE I. ORÍGENES

Una pregunta que se puede plantear es: ¿La forma de asignar pesos por parte de Laplace es correcta desde una moderna perspectiva? Primero, recuérdese que dos fuentes de error entran en cada medición a_k : en la medición por tierra y en las observaciones astronómicas, compuestas estas últimas por un conjunto de tan sólo dos mediciones por arco (una en cada extremo). Sin considerar su longitud, se podría ignorar esta última fuente de error en el esquema de asignar pesos, debido a que no hay más valores a considerar; sin embargo, actualmente esto no es del todo correcto. Las observaciones astronómicas de los arcos afectan a a_k como un divisor, siendo que los errores en arcos pequeños tienen un mayor efecto que los mismos en arcos grandes. Es aquí donde se aprecia el análisis de Laplace. El error fue relativo e igualmente probable gracias a su ponderación. Los errores en las mediciones por tierra tendrían una varianza aproximadamente proporcional a i_k -si se asume una fracción constante de la longitud medida con cadenas para cada arco-, por lo que se debería esperar una varianza para a_k inversamente proporcional a i_k . Ahora, el trabajo original de Boscovich asignaba pesos a las mediciones tomando en cuenta varianzas iguales para las mediciones de a_k , mientras que Laplace lo hacía considerando que a_k tenía varianzas inversamente proporcionales i_k^2 -por su numerador y su divisor-. La forma en que Laplace trató su problema, considerando la época, es más que interesante.

El método de Boscovich, generalizado por Laplace, continuó aplicándose ocasionalmente desde la publicación de la *Mécanique Céleste*. Prony lo aplicó en detalle a problemas de fluidos en 1804 (Prony, 1804, pp. 21-32). Tres años más tarde Puissant (1807, p.63) presentó en detalle un uso recomendado en *supervivencia*. En 1809 y 1815, Bowditch publicó dos generalizaciones del método con comentarios. Éstas fueron particularmente interesantes, debido a que combinaban los métodos de Mayer y Boscovich. Por ejemplo, enfrentando cincuenta y seis ecuaciones de condición involucrando cinco incógnitas, aplicó la condición de que la suma de ecuaciones debería ser cero, para cuatro diferentes conjuntos separados de ecuaciones; eliminando cuatro de las variables y resolviendo la última usando el algoritmo de Laplace. En 1832, recomendaba el método de Boscovich sobre los mínimos cuadrados debido a que el primero otorgaba un menor peso a observaciones defectuosas (Bowditch, 1832, p.434).

1.2. La invención de los mínimos cuadrados

Ahora, casi se ha llegado al método de mínimos cuadrados, tanto cronológica como conceptualmente. Se ha visto cómo para 1800, el principio de combinar ecuaciones de condición a través de los trabajos de Mayer y Laplace, había sido involucrado para producir procedimientos especiales en muchas situaciones generales. También, cómo la idea de comenzar con un criterio matemático había dejado, en los trabajos de Boscovich y Laplace, una solución elegante y adecuada para relaciones simples lineales involucrando sólo dos variables. Las primeras de estas aproximaciones fueron desarrolladas en torno a problemas astronómicos; las segundas, fueron -al menos en los primeros años- exclusivamente empleadas en conexión con intentos de determinar la figura de la Tierra. Estas dos líneas llegaron juntas en el trabajo de un hombre que, como Laplace, fue un excelente matemático trabajando en problemas de ambas áreas: Legendre. Éste llegó a tratar con problemas astronómicos empíricos y geodésicos, al mismo tiempo que los procedimientos que se han discutido eran desarrollados de forma separada en los dos campos. Fue un tiempo en donde estos métodos habían dado un cambio revolucionario en el manejo de observaciones, desde los intentos de Euler a la teoría de Laplace. Llegó entonces el turno de Legendre, a quien le tocaría dar el paso más importante.

1.2.1. Legendre y la invención del método

En marzo de 1805 la Europa científica miraba hacia Francia; ahí, la disciplina era intelectual, no marcial, como lo era en la mayoría del territorio europeo. Laplace celebraba su cumpleaños número cincuenta y seis con la preparación del cuarto volumen de su *Traité de Mécanique Céleste* y, tal vez, comenzaba con sus ideas sobre la publicación de un libro de probabilidad en el cual había trabajado durante casi veinte años. También en marzo de 1805 otro científico matemático francés, Adrien Marie Legendre, terminaba sus últimas páginas de un largo tratado que contenía la primera publicación de la que hasta hoy es la técnica no-trivial más extensamente utilizada en la estadística: el método de mínimos cuadrados.

Legendre (18 de Septiembre de 1752 – 10 de enero de 1833) fue un matemático de gran amplitud y originalidad. Era tres años más joven que Laplace, sucediendo a éste último como profesor de matemáticas en la Escuela Militar y en la Escuela Normal. Su trabajo matemático más conocido fue uno sobre integrales elípticas -pionero en esta área cuarenta años antes que Abel y Jacobi-, descubridor de la ley de reciprocidad cuadrática, además de ser un gran geómetra. Su *Éléments de géométrie* fue uno de los tratados más

exitosos del siglo XIX. Además, escribió importantes trabajos sobre la teoría de atracción gravitacional. Fue miembro de dos comisiones francesas, una en 1795 reuniéndose en los observatorios de Paris y Greenwich, y la otra destinada a medir el arco a través de su meridiano desde Barcelona a Dunkirk, el arco en el cual el metro fue aplicado como unidad de medida. Este proyecto fue el nexo con sus trabajos posteriores en astronomía y geodesia en donde el método de mínimos cuadrados apareció.

El 6 de marzo de 1805 Legendre publicó el trabajo al cual se hace referencia, llamado *Nouvelles méthodes pour la détermination des orbites des comètes*; un libro de ocho páginas, pero con un extenso suplemento de cincuenta y cinco páginas en enero de 1806 y un segundo suplemento de ocho páginas en agosto de 1820. El apéndice presentaba el método de mínimos cuadrados ocupando las nueve primeras páginas tituladas “*Sur la méthode des moindres quarrés*” (Sobre el método de los mínimos cuadrados). Esta clara y rígida exposición debió ser una de las introducciones más elegantes de un nuevo método estadístico en la historia de esta disciplina. Para hacer referencia a una suma, Legendre usó \int en vez de \sum , comenzando con una declaración de su objetivo:

Sobre el método de los mínimos cuadrados

“En muchas investigaciones, donde el objetivo es deducir el resultado más preciso posible desde mediciones observacionales, nos lleva a un sistema de ecuaciones de la forma

$$E = a + bx + cy + fz + \&c.,$$

en las cuales, $a, b, c, f, \&c.$ son coeficientes conocidos, variando de una ecuación a otra, y $x, y, z, \&c.$ son cantidades desconocidas, a ser determinadas bajo la condición de que cada valor de E se reduzca, para cada ecuación, a una cantidad nula o muy pequeña.” (Legendre, 1805, p. 72)

Hoy se puede escribir esto como

$$E_i = a_i + b_i x + c_i y + f_i z + \dots$$

o como

$$a_i = -b_i x - c_i y - f_i z - \dots + E_i,$$

siendo entonces los E_i los errores. Cuando el número de ecuaciones es el mismo que el número de variables, Legendre no encontró problema. Pero cuando había más ecuaciones que incógnitas, llegó a ser imposible escoger valores para las variables que pudieran eliminar todos los errores. Notó que había un grado de arbitrariedad de cualquier forma

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

en la “distribución de los errores entre las ecuaciones”, pero esto no lo detuvo para proponer la única mejor solución:

“De todos los principios que pueden ser propuestos para este propósito, pienso que no hay ninguno más general, más exacto, o más fácil de aplicar, que del que haremos uso en este trabajo; éste consiste en reducir la suma de los cuadrados de los errores a un mínimo. Por este método, un tipo de equilibrio es establecido a lo largo de los errores, el cual es apropiado para revelar el estado del sistema que se aproxima más a la verdad.” (Legendre, 1805, pp. 72-73)

¡Minimizar la suma de los cuadrados de los errores! Tan simple, pero ¿sería práctico? Legendre no perdió tiempo en escribir las ecuaciones derivadas por la diferenciación de la suma de los cuadrados de los errores $(a+bx+cy+fz+\&c.)^2 + (a'+b'x+c'y+f'z+\&c.)^2 + (a''+b''x+c''y+f''z+\&c.)^2 + \&c.$ con respecto a cada una de las variables x, y, z, \dots , de la siguiente forma:

$$0 = \int ab + x \int b^2 + y \int bc + z \int bf + \&c.$$

$$0 = \int ac + x \int bc + y \int c^2 + z \int fc + \&c.$$

$$0 = \int af + x \int bf + y \int cf + z \int f^2 + \&c.$$

“... donde por $\int ab$ entendemos la suma de los productos similares $ab + a'b' + a''b'' + \&c$; y por $\int b^2$ la suma de los cuadrados de los coeficientes de x , esto es $b^2 + b'^2 + b''^2 + \&c.$, y así sucesivamente.” (Legendre, 1805, p.73)

Para que no hubiera duda de como formar estas “ecuaciones normales” -un nombre introducido más tarde por Gauss-, Legendre repitió la regla con palabras cursivas:

“En general, para formar las ecuaciones del mínimo con respecto a una de las variables, es necesario multiplicar todos los términos de cada ecuación por el coeficiente de la variable en esa ecuación, tomado con su propio signo, y entonces encontrar la suma de todos estos productos.” (Legendre, 1805, p.73)

Las primeras páginas en la historia del método de mínimos cuadrados son mostradas en la Figura 1.4, en su francés original.

Las ecuaciones, tantas como variables haya, pueden ser resueltas mediante “métodos ordinarios”. Entonces, la practicidad del principio de Legendre era evidente; todo lo que requería eran simples multiplicaciones y sumas, así como la voluntad de resolver un

sistema de ecuaciones lineales. De una forma más convincente, Legendre agregó los siguientes puntos:

- (1) Si un ajuste perfecto fuese posible, su método debería encontrarlo.
- (2) Si fuera necesario descartar consecuentemente una ecuación (debido a que el "error" fuese demasiado grande), sería sólo cuestión de corregir las ecuaciones sustrayendo los términos apropiados.
- (3) La media aritmética es un caso especial del método, encontrada cuando hay una sola variable con coeficientes constantes $b = b' = \dots = 1$.
- (4) Igualmente, encontrar el centro de gravedad de varias masas iguales en el espacio es una consecuencia.

Seguramente el lector relacionará el caso especial del punto (3) con la extrapolación analítica hecha anteriormente para el trabajo de Euler.

Legendre cierra su introducción del método con estas palabras sobre el punto (4), "Vemos, por lo tanto, que el método de mínimos cuadrados revela, a manera de palabras, el centro alrededor del cual las observaciones se concentran por si mismas, a condición de que las desviaciones con respecto al centro sean tan pequeñas como sea posible." (Legendre, 1805, p.75)

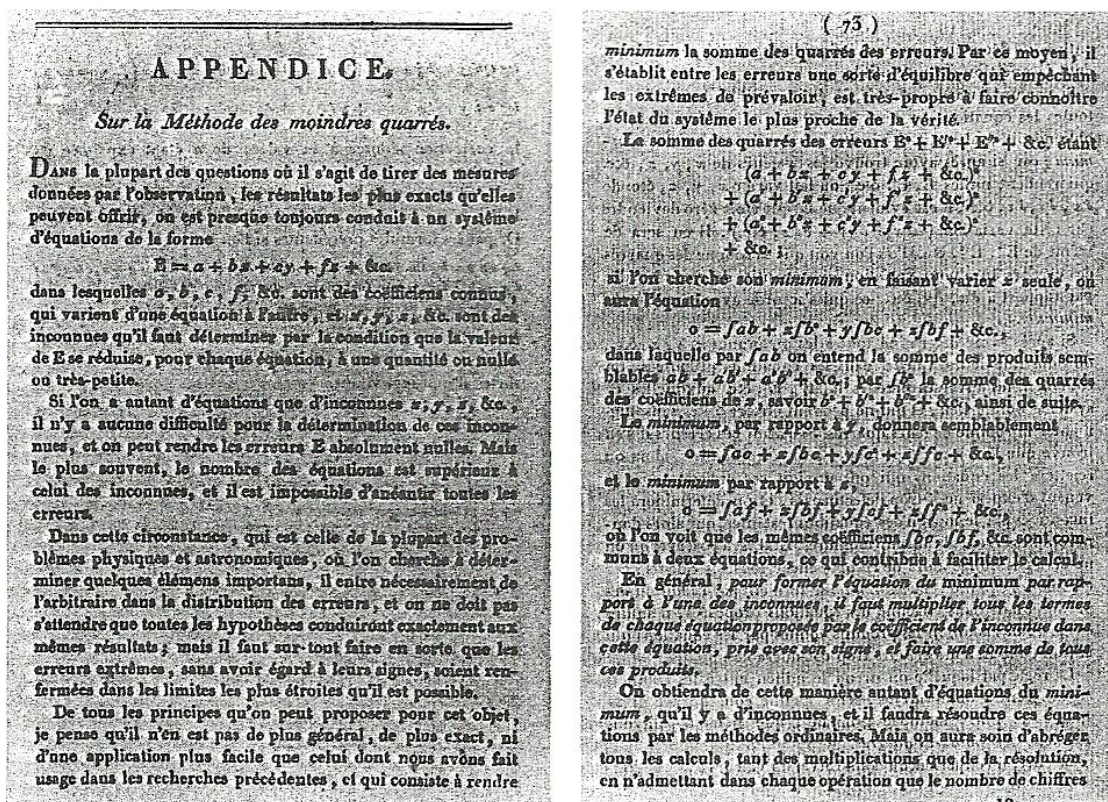


Figura 1.4. Texto original en francés de Legendre. (Legendre, 1805, pp. 72, 73)

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

Legendre siguió con un ejemplo, usando datos de 1795 sobre un estudio de arcos que involucraba tres variables en cinco ecuaciones; ejemplo que se verá a continuación.

Para 1795, el antiguo sistema de medición cambió como una simbólica adaptación al nuevo orden impulsado por la Revolución Francesa. La base del nuevo sistema fue el *metro*, definido como la fracción $1/10,000,000$ del cuadrante de un meridiano, medido entre el Polo Norte y el ecuador. Restaba entonces determinar la longitud de este arco. Manteniendo el nacionalismo que inspiró este objetivo, esta determinación estuvo basada en mediciones realizadas sólo en tierras francesas; lo que le dio el nombre a este arco de: *Arco Francés*. Para los datos (Tabla 1.6), Legendre desarrolló la relación entre la longitud de arco en grados y en toesas de una forma diferente de lo que se ha visto hasta ahora. Siendo L y L' latitudes astronómicas determinadas en los puntos extremos de un arco (columna dos de la Tabla 1.6) y S la longitud medida del arco (columna tres de la Tabla 1.6 en módulos, donde un módulo son sólo dos toesas), Legendre escribió:

$$L' - L = \frac{S}{D} + \frac{3}{2} \cdot \alpha \cdot \frac{180}{\pi} \text{sen}(L' - L) \cos(L' + L)$$

$$= \frac{S}{28,500} + \zeta \cdot \frac{S}{28,500} + \alpha \cdot \frac{270}{\pi} \text{sen}(L' - L) \cos(L' + L)$$

Lugar de la Observación	Latitud L	Longitud de del arco S	$L' - L$	$L' + L$
Dunkirk	51°2'10''50	62,472.59	2°11'20''75	99°53'0''
Pantheon (Paris)	48°50'49''75	76,145.74	2°40'7''25	95°1'32''
Evaux	46°10'42''50	84,424.55	2°57'48''10	89°23'37''
Carcassonne	43°12'54''40	52,749.48	1°51'9''60	84°34'39''
Montjoux	41°21'44''80			

Tabla 1.6. Latitudes del arco francés. [17] pp. 59

Aquí D es la longitud en módulos de 1° centrado en una latitud de 45° , α es la elipticidad de la Tierra, y ζ es definido en la relación $D^{-1} = (1 + \zeta) / 28,500$. A primera vista esto parece ser muy diferente de la relación anterior $a = z + y \text{sen}^2 \theta$; pero no lo es. Nótese que $\text{sen}^2 45^\circ = 0.5$, $D = z + y / 2$; y desde que $\theta = (L + L') / 2$, se tiene $2 \text{sen}^2 \theta = 1 - \cos 2\theta = 1 - \cos(L + L')$. Entonces, debido a que $3\alpha = y / D$ y $a = S / (L' - L)$, se puede ver que las fórmulas son equivalentes, salvo que Legendre emplea $180 \text{sen}(L' - L) / \pi$ en vez de su aproximación local $L' - L$. Su formulación es -excepto por la aproximación local de $\text{sen}(L' - L)$ - exactamente la misma a la usada por Laplace para su

PARTE I. ORÍGENES

análisis ponderado: en la notación previa, la ecuación de Legendre es equivalente a $i_k a_k = i_k z + i_k y p_k$. La introducción de ζ en vez de D es sólo una reparametrización, basada en que se espera que D es aproximadamente 28,500; siendo más fácil y más preciso trabajar con números pequeños. De esta forma, el error es concebido de forma análoga a la ecuación de Legendre $i_k a_k - i_k z - i_k y p_k = i_k x_k$.

Legendre nombró como E^i al error hecho en la determinación de la i -ésima latitud en vez de $i_k x_k$; y con base en los datos de la Tabla 1.6, obtuvo cuatro ecuaciones:

$$E^I - E^{II} = 0.002923 + \zeta(2.192) - \alpha(0.563) \quad (1)$$

$$E^{II} - E^{III} = 0.003100 + \zeta(2.672) - \alpha(0.351) \quad (2)$$

$$E^{III} - E^{IV} = -0.001096 + \zeta(2.962) + \alpha(0.047) \quad (3)$$

$$E^{IV} - E^V = -0.001808 + \zeta(1.851) + \alpha(0.263) \quad (4)$$

Para formar estas ecuaciones, Legendre supuso que el efecto de estos errores sobre $\text{sen}(L' - L)\cos(L' + L)$ era despreciable. Ahora, pudo haber aplicado el método de mínimos cuadrados directamente haciendo un cambio de variable $E^j = E^{i-1} - E^i$; pero en vez de eso, pensó que era necesario considerar a los errores en forma separada. Para ello introdujo una quinta ecuación, $E^{III} = E^{III}$, la cual le permitió construir el siguiente sistema:

$$E^I = E^{III} + 0.006023 + \zeta(4.864) - \alpha(0.914) \quad (i)$$

$$E^{II} = E^{III} + 0.003100 + \zeta(2.672) - \alpha(0.351) \quad (ii)$$

$$E^{III} = E^{III} \quad (iii)$$

$$E^{IV} = E^{III} + 0.001096 - \zeta(2.962) - \alpha(0.047) \quad (iv)$$

$$E^V = E^{III} + 0.002904 - \zeta(4.813) - \alpha(0.310) \quad (v)$$

de donde la ecuación (i) es la suma de (1) y (2), mientras que (v) es la resta de (4) menos (3). La ecuación (ii) permaneció igual a (2), lo mismo con (iv) y (3). Entonces Legendre resolvió las ecuaciones por el método de mínimos cuadrados, tratando a E^{III} del lado derecho como variable, junto con ζ y α . Esto lo hizo escribiendo

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

$$\begin{aligned} \sum (E^i)^2 = & [E''' + 0.006023 + \zeta(4.864) - \alpha(0.914)]^2 \\ & + [E''' + 0.003100 + \zeta(2.672) - \alpha(0.351)]^2 \\ & + [E''']^2 \\ & + [E''' + 0.001096 - \zeta(2.962) - \alpha(0.047)]^2 \\ & + [E''' + 0.002904 - \zeta(4.813) - \alpha(0.310)]^2 \end{aligned}$$

Después de hacer el desarrollo, derivar esta suma con respecto a cada una de las incógnitas para obtener las ecuaciones normales y resolver el nuevo sistema de tres ecuaciones con el mismo número de variables por algún método, Legendre encontró que $\alpha = 0.00675 = 1/148$ y $\zeta = 0.0000778$. Entonces $D = 28500 / (1 + \zeta) = 28,497.78$, obteniendo después la longitud total del cuadrante como $90 \cdot D = 2,564,800.20$ módulos, un valor que deja un metro de 0.256480 módulos = 0.512960 toesas $\cong 3.280$ pies. El *metro actual*, tal y como se conoce, está basado en el valor de D encontrado por Laplace anteriormente en su *Mécanique Céleste* (vol. 2, 1799-1805, bk.3 p.465) en su intento por minimizar el error más grande, obteniendo $D = 28,504.11$; lo cual dio un cuadrante a través del meridiano de $2,565,370$ módulos y un metro de 0.256537 módulos = 0.513074 toesas $\cong 3.281$ pies. La determinación de Laplace incorporó la medida de un arco en Perú para calcular la elipticidad; después usó ésta para encontrar el valor de D del *Arco Francés* con base en su propio algoritmo, previamente visto en este trabajo.

Una observación en el ejemplo de Legendre entra en cuestión: hoy esta aproximación es incorrecta desde un punto de vista correlacional. No introduce un sesgo considerable en los resultados, pero ignora la variabilidad de la incógnita artificialmente construida E''' ; esto produce un peso en las observaciones que no es de máxima eficiencia para la estructura estocástica de los errores. Esta observación no debe considerarse como una crítica ni mucho menos, considerando que esto sucedió más de un siglo antes de que métodos eficientes fueran desarrollados para tratar el tipo de correlaciones que Legendre enfrentó, además de que actualmente, la diferencia en los resultados es despreciable. Es remarcable que, carentes de cualquier formulación probabilística explícita, Laplace y Legendre hicieron lo posible para manejar sus problemas, mostrando una temprana noción de dependencia cuando los mismos errores aparecían en diferentes ecuaciones. Medio siglo después, el concepto de probabilidad fue introducido en el análisis de diversos problemas, en donde la idea de dependencia fue más evidente que en el primer ejemplo publicado del método de mínimos cuadrados.

Con la introducción de los mínimos cuadrados por parte de Legendre, se llega al final de las líneas de desarrollo que comenzaron en forma separada por Mayer y Boscovich cerca de medio siglo antes. La idea de combinar ecuaciones diferentes evolucionó lentamente desde el trabajo astronómico de Mayer, la idea de un criterio objetivo del ajuste en el trabajo geodésico de Boscovich y la síntesis de las dos áreas con el tratado de Legendre. Pero un elemento clave seguía faltando: En todos estos trabajos, no había una llamada formal a la probabilidad ni alguna manera de cuantificar la precisión de los valores estimados. Todo esto es un misterio, debido a que Laplace se había introducido en el desarrollo de la probabilidad desde 1774, teniendo un papel muy importante en esta área. A continuación se verá cómo el siguiente paso fue completado en las dos décadas posteriores al análisis de Legendre, por el mismo Laplace y por un alemán de nombre Carl Friedrich Gauss.

1.2.2. Gauss y su primera prueba

La publicación del método de mínimos cuadrados en 1805 marcó sólo el fin de lo anteriormente iniciado. Legendre había propuesto un procedimiento que fue fácilmente aplicable a problemas de astronomía y geodesia, basado en un criterio sencillo de comprender e intuitivamente razonable. Pero desde un punto de vista moderno algo estaba faltando: una consideración formal de la probabilidad y su relación con los mínimos cuadrados. Ésta fue crucial para la utilidad y posterior desarrollo del método, ya que sin ella, ninguna valoración relativa a la precisión de los resultados hubiera sido posible. El método de mínimos cuadrados produce resultados que podrían ser llamados “los mejores”, ya que minimiza la suma de cuadrados de los errores y produce un atractivo equilibrio en las observaciones; pero conforme la naturaleza estocástica de los errores y observaciones pasara desapercibida, la cuantificación de la incertidumbre y la pregunta de qué tanto son “los mejores resultados”, eran cuestiones imposibles de abordar.

La siguiente etapa del desarrollo fue producida por la publicación en 1809 de *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientium* (La teoría del movimiento de cuerpos celestes moviéndose alrededor del Sol en secciones cónicas) por Carl Friedrich Gauss (30 de abril de 1777 – 23 de febrero de 1855). Gauss fue un matemático, astrónomo, geodésico y físico alemán que contribuyó significativamente en muchos campos, incluida la teoría de números, el análisis matemático, la geometría diferencial, la estadística, el álgebra, la geodesia, el magnetismo y la óptica. Considerado “el príncipe de las matemáticas” y “el matemático más grande desde la antigüedad”, Gauss ha tenido una influencia notable en muchos campos de la matemática y de la ciencia. La estadística no ha sido la excepción.

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

En su libro de 1809, el objetivo principal era una investigación matemática de las órbitas planetarias. Al final, sin embargo, agregó una sección sobre combinación de observaciones. En esta sección enfrentó esencialmente el mismo problema que Mayer, Euler, Boscovich, Laplace y Legendre habían enfrentado antes, pero con una gran diferencia: Gauss redactó su problema en términos probabilísticos explícitos. Supuso que había μ funciones lineales

$$V = ap + bq + cr + ds + etc.$$

$$V' = a'p + b'q + c'r + d's + etc.$$

$$V'' = a''p + b''q + c''r + d''s + etc.$$

etc.

con valores observables a, b, c, \dots y con v coeficientes desconocidos p, q, r, \dots ; y que éstos son encontrados también mediante observaciones directas formando los valores $M, M', M'',$ etc. Gauss supuso que los posibles valores de los errores definidos como $\Delta = V - M, \Delta' = V' - M', \Delta'' = V'' - M'',$ etc., tenían probabilidades dadas por una curva $\varphi(\Delta), \varphi(\Delta'), \varphi(\Delta''),$ etc. Los ingredientes para esta formulación fueron todos tomados de trabajos anteriores, en particular de varios escritos de Laplace. Y entonces, éste fue el enfoque que Gauss eligió para enfrentar el dilema de resolver un sistema con más ecuaciones lineales que incógnitas: usó la forma de Laplace para el teorema de Bayes. Supuso *a priori*, que todos los valores para las incógnitas eran igualmente probables; con esta suposición, sólo restaba encontrar las soluciones para $p, q, r,$ etc., que maximizaran

$$\Omega = \varphi(\Delta)\varphi(\Delta')\varphi(\Delta'')etc.$$

siendo éste el conjunto “más probable” de soluciones. Ahora, como Gauss había notado, estos valores podían ser encontrados derivando Ω con respecto a $p, q, r,$ etc., igualando a cero y resolviendo en forma simultánea el sistema resultante. Sin embargo, estas ecuaciones involucraban a la curva $\varphi(\Delta)$; antes de que Gauss pudiera continuar, necesitaba una fórmula para la distribución de los errores. Notó que sólo podía establecer condiciones generales para $\varphi(\Delta)$: debería tener un máximo en $\Delta = 0$, ser simétrica y valer cero fuera del rango de los errores posibles. En vez de imponer más condiciones en forma directa, él supuso la conclusión: Gauss adoptó como axioma el principio de que el valor más probable de una variable, observada varias veces bajo las mismas circunstancias, es la media aritmética de las observaciones y probó además en este caso (esto es, donde $V = V' = V'' = \dots = p$) que, tomando $p = (1/\mu)(M + M' + M'' + etc.)$ se maximiza a Ω sólo cuando

PARTE I. ORÍGENES

$$\varphi(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2},$$

para alguna constante positiva h , la cual podría ser vista como una medida de precisión de las observaciones. Entonces, Gauss demostró cómo en la más general de las situaciones, esta distribución de los errores permitía al método de mínimos cuadrados obtener valores para p , q , r , etc., que maximizaran Ω . Es evidente que $\varphi(\Delta)$ se trata de una *distribución normal*.

Visto desde una perspectiva actual –de dos siglos más tarde–, se pueden encontrar algunos aspectos a considerar en el argumento de Gauss. Su razonamiento cae en un círculo de ideas lógicas que quizá hoy en día no tengan mucho sentido. ¿Por qué? Tres cuestiones aquí están involucradas: La media aritmética –un caso especial del método de mínimos cuadrados que ya se ha considerado previamente en la introducción de este trabajo– es sólo “más probable” si los errores están normalmente distribuidos; la media aritmética era “en general reconocida” como una excelente forma de combinar observaciones, por lo que los errores deberían ser normalmente distribuidos –como si los anteriores estadísticos ya hubieran leído a Gauss–. Finalmente, como se sabe, suponer que los errores son normalmente distribuidos le da entrada al método de mínimos cuadrados. Incluso el mismo Gauss, buscó doce años después la salida a esta cadena de razonamientos cerrados. Excepto por una circunstancia, su argumento pudo haber pasado inadvertido, ya que se hizo de construcciones a modo para explicar el problema de una forma más limpia, pero menos convincente. Sin embargo, la circunstancia por la que su argumento es crucial, radica en que Gauss le dio un primer sentido probabilístico a la combinación de observaciones y con ello, permitió al método de mínimos cuadrados seguir con su desarrollo. Su razonamiento tuvo un gran impacto en la persona de Laplace, a pesar de que el método estaba lejos de ser completado.

El libro de Gauss llegó a París en mayo de 1809. A pesar de ello, parece ser que Laplace no leyó este trabajo hasta después de haber presentado el *teorema del límite central* a la Academia en abril de 1810. Ninguna referencia de las ideas de Gauss es encontrada en este escrito y, menos aún, alguna pista de que Laplace fuera capaz de pensar en la posibilidad de e^{-t^2} como curva para los errores. Incluso con esto, contaba con una poderosa herramienta que aparentaba no tener mayor aplicación que en la inclinación media de los cometas. Pero para finales de 1810, las cosas cambiarían de una forma dramática e irreversible.

1.2.3. La Síntesis de Gauss-Laplace

Laplace encontró el trabajo de Gauss en abril de 1810 y de inmediato mostró su desacuerdo. Por supuesto, debió haber pensado que su derivación no tenía mucho sentido. En cambio, él contaba con una alternativa en mano que si lo tenía. Antes de ver el libro de Gauss, no había encontrado alguna conexión entre el *teorema del límite central* y la *estimación lineal*, pero casi de inmediato, pudo ver cómo se conjuntaban. Una secuela de este teorema fue publicada por Laplace en donde cita a Gauss, reexpresando el teorema en una forma más útil, y demostró cómo éste podía proveer una mejor formulación en la elección de $\varphi(\Delta)$: Si los errores en el análisis de Gauss fueran sumados, entonces el teorema del límite central tendría como consecuencia que éstos deberían estar aproximadamente distribuidos como lo que más tarde se conocería con el nombre de una distribución *normal* o *curva gaussiana* para $\varphi(\Delta)$. Una vez realizada la elección para la curva de los errores, el desarrollo de los mínimos cuadrados tomó forma, siguiendo la línea que Gauss había comenzado un año antes. De hecho, Laplace tomó ventaja de su propio resultado de 1774, concluyendo que el principio de mínimos cuadrados se encuentra en esta distribución, haciendo que el error esperado se minimice. Incluso en 1809, el mismo Gauss había ido más lejos, simplemente observando que esta curva le daba paso al método de mínimos cuadrados, explorando más allá las consecuencias e intentando encontrar formas de simplificar su uso. Gauss notó que si los errores se comportaban de acuerdo a la curva $\varphi(\Delta)$, entonces las variables $p, q, r, \text{etc.}$, seguirían una distribución proporcional a e^{-h^2W} , donde $W = \Delta^2 + \Delta'^2 + \Delta''^2 + \text{etc.}$ Posteriormente, demostró cómo este hecho podría ser usado para asentar la precisión del método: obteniendo la distribución marginal de las variables en forma separada, además de su definida "precisión" -lo que hoy se escribe como $1/(\sigma\sqrt{2})$ -. También generalizó para el caso en donde las funciones $V', V'', V''', \text{etc.}$ fueran no lineales en $p, q, r, \text{etc.}$, pero sí de forma local. Por último, explicó un algoritmo de eliminación para resolver en forma simultánea las variables $p, q, r, \text{etc.}$, famoso procedimiento conocido hoy en día como *eliminación gaussiana*.

La forma en la que el razonamiento de Gauss encaja con el principio de Legendre es abrumadora, desde un particular punto de vista. El hecho de que $W = \Delta^2 + \Delta'^2 + \Delta''^2 + \text{etc.}$ sea normalmente distribuida, es equivalente a minimizar la suma de los cuadrados de los errores, ya que la función e^{-h^2W} es mínima cuando W también lo es, todo ello considerando la independencia de los errores.

A pesar de lo especial que fue la formulación de Gauss, su desarrollo contaba con algunas limitaciones. En la discusión sobre la precisión de los estimadores para las variables, éste sólo señala precisión relativa. Un estimador para p fue catalogado como 4.6 veces más preciso que una simple observación de la variable (h era igual a la unidad en este caso). Ningún intento fue realizado para traducir esto a un planteamiento, expresando la incertidumbre del estimador para p en su propia escala de medida; y ningún esquema fue presentado para determinar la precisión desconocida h , en el caso de una sola medición. Cualquier crítica está de sobra, considerando lo importante que resultó su síntesis: La fusión del trabajo de un siglo a través de dos líneas distintas -manejo de observaciones y teoría de la probabilidad- en una.

Sólo resta tratar una cuestión: ¿El principio de Legendre sirvió como herramienta base? O ¿la síntesis de Gauss constituía en sí mismo el principio de Legendre? Gauss afrontó este hecho refiriéndose al método de mínimos cuadrados como “nuestro principio” (*principium nostrum* en latín) y reclamando que él había estado usando el método desde 1795. La consiguiente disputa prioritaria además de otra que involucraba la *ley de reciprocidad cuadrática* en *teoría de números*, hizo que se exacerbara la relación entre los dos hombres. El problema no fue tan grande como la controversia Newton-Leibniz, pero tuvo sus momentos de tensión. Legendre lanzó algunos ataques a Gauss mediante correspondencia, además de criticarlo -de forma mesurada- en la versión de 1820 de su *Nouvelles méthodes pour la détermination des orbites des comètes*. Por su parte, Gauss solicitó testimonios convincentes de amigos que constataran que él había usado el método de mínimos cuadrados antes de 1805. Un estudio reciente y demás evidencias sugieren que Gauss tendría razón en haber sido el primero en usar el método, pero no tuvo éxito en todo intento que hiciese para mostrarlo antes de 1805; además de que parece no haber visto su gran potencial sino hasta después de ver el trabajo del francés. Por otro lado, el apéndice de Legendre de 1805 fue una clara proclamación de un método general por un hombre que no tuvo duda de su importancia.

El suplemento de Laplace de 1810 fue sólo el primer signo de la gran influencia que el libro de Gauss le había dejado. En este suplemento, había confinado su razonamiento probabilístico, cuyo contenido mostraba cómo su teorema del límite central podía mejorar el argumento de Gauss al seleccionar una *normal* como distribución del error y usar esto en su propio análisis para llegar a la conclusión en un sólo salto de que el método de mínimos cuadrados encuentra a los estimadores “más probables” y los “más precisos” -en el sentido del error mínimo esperado-.

El nuevo trabajo de Laplace, en el cual usó el teorema del límite central de un modo diferente, fue presentado en un largo escrito que fue publicado en 1811. Éste fue

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

reiterado en su libro del año siguiente titulado *Théorie Analytique des Probabilités* (Laplace, 1812, pp. 312-329). Él comenzó con la más simple de las situaciones donde, en una cercana aproximación, los errores fueron relacionados con las observaciones mediante la fórmula $\varepsilon^{(i)} = p^{(i)}z - \alpha^{(i)}$. En esta notación -la cual se asemeja más a la de Legendre que a la de Gauss-, $p^{(i)}$ y $\alpha^{(i)}$ están determinados por las observaciones, z es un coeficiente a determinar y $\varepsilon^{(i)}$ son los errores. En este esfuerzo de 1811, Laplace evitó cualquier suposición inicial acerca de la distribución de los errores, contrariamente a lo que Gauss hizo -suponiendo que los errores están simétricamente distribuidos alrededor del cero-. En vez de ello, debido a que

$$\sum m^{(i)} \varepsilon^{(i)} = z \sum m^{(i)} p^{(i)} - \sum m^{(i)} \alpha^{(i)},$$

una vez que los pesos $m^{(i)}$ estuvieran especificados, Laplace resolvió esta ecuación para encontrar que

$$z = \frac{\sum m^{(i)} \alpha^{(i)}}{\sum m^{(i)} p^{(i)}}$$

suponiendo que los pesos $m^{(i)}$ eran enteros positivos o negativos. Esta aproximación le dio un conjunto de estimadores a considerar, uno por cada asignación de los pesos $m^{(i)}$. A primera vista, ésta es una manera curiosa de tratar el problema pero, por supuesto, fue esencialmente una formulación algebraica de la forma generalizada del método de Mayer: dado un sistema de más ecuaciones inconsistentes que variables, tomadas en varios conjuntos, tantos como variables hubiese, y resolver para las incógnitas. En esta ocasión, Laplace procedió con tan sólo una variable. Siempre supuso -aunque esto no es trascendental para el análisis- que los pesos $m^{(i)}$ son todos números; entonces la suma era realmente una combinación de las ecuaciones observacionales originales -como en el caso del método de Mayer-, pero con el peso de $m^{(i)}$ veces la i -ésima ecuación dentro de esta combinación. Sólo restaba especificar los pesos, y aquí el desarrollo *gaussiano* dio una nueva dirección. Por ahora, Laplace estaba pensando de una forma probabilística; los errores fueron tomados como aleatorios, simétricamente distribuidos alrededor del cero. No debió tener $\sum m^{(i)} \varepsilon^{(i)} = 0$ en forma exacta, siendo apenas diferente por una cantidad aleatoria. Entonces, el último cociente no tendría que ser exactamente z , difiriendo por una cantidad denotada por u :

$$z = \frac{\sum m^{(i)} \alpha^{(i)}}{\sum m^{(i)} p^{(i)}} + u,$$

donde

$$u = \frac{\sum m^{(i)} \varepsilon^{(i)}}{\sum m^{(i)} p^{(i)}}$$

Una vez que había expresado el problema en estos términos, Laplace estaba listo para determinar los pesos. Con una ligera extensión de su teorema del límite central, pudo mostrar que el error u era -si el número de observaciones es grande- aproximadamente *normal* y que, para esta distribución límite, no sólo el error esperado era proporcional a

$$\frac{\sum m^{(i)^2}}{\sum m^{(i)} p^{(i)}},$$

sino que también la probabilidad de cualquier intervalo simétrico alrededor del cero variaba inversamente con esta cantidad. Resultó entonces simple calcular y concluir que esta cantidad sería la más pequeña posible si los $m^{(i)}$ fueran proporcionales a los $p^{(i)}$. Pero eso justamente era el estimador de z por mínimos cuadrados:

$$\frac{\sum p^{(i)} \alpha^{(i)}}{\sum p^{(i)^2}},$$

si se deriva la ecuación $\sum [p^{(i)} z - \alpha^{(i)}]^2 = 0$ con respecto a z para después despejarla. Fue entonces que la credibilidad probabilística que el método de mínimos cuadrados reclamaba, le fue otorgada por el teorema del límite central.

La conclusión de Laplace puede ser resumida en términos modernos: Todos los estimadores de z son funciones lineales de los valores observables $\alpha^{(i)}$ (donde los coeficientes de $\alpha^{(i)}$ pueden depender de $p^{(i)}$) y están normalmente distribuidos de forma aproximada y, de entre todos ellos, el estimador por mínimos cuadrados tiene el error esperado mínimo. Continuó generalizando esta conclusión en problemas de estimación lineal con varias incógnitas (Laplace, 1811; 1812, pp. 322-329); con un argumento que derivó la función *normal multivariada* para dos o más valores a ser estimados mediante mínimos cuadrados. Esta conclusión, aunque estaba basada en argumentos sólidos, no consideraba ninguna formulación rígida para la curva de los errores. Sugirió que la función $(1/2a) \log(a/|x|)$, que había derivado en 1777, podría ser considerada como posible curva para los errores (1811; 1812, p.321).

Laplace encontró en los mínimos cuadrados el principio de precisión para un gran número de ecuaciones observacionales en dos conceptos: proporcionar el error mínimo esperado,

1.2. LA INVENCION DE LOS MÍNIMOS CUADRADOS

y el estimador con el valor más probable de la variable a estimar. Más tarde, calificó al método de mínimos cuadrados como “el más ventajoso”. La síntesis de Gauss-Laplace conjuntó el desarrollo de dos líneas: por un lado, la combinación de observaciones a través de la agregación (o suma) de ecuaciones de condición lineales y, por otro, el uso de la probabilidad matemática para realizar inferencia estadística. En muchos aspectos, significó uno de los mayores éxitos en la historia de la ciencia. Plantea toda una serie de aplicaciones, incluso en las ciencias sociales con Adolphe Quetelet como su principal precursor.

Doce años después, Gauss publicó una extensión de estos argumentos notando que el análisis involucraba sólo segundos momentos, y que si uno buscaba medir la precisión mediante errores cuadrados, entonces la conclusión llegaría sin considerar el número de ecuaciones observacionales en cuestión. Esta extensión es conocida como el *teorema de Gauss-Markov* revisado en el apéndice de la *Parte IV* de este trabajo en su forma generalizada.

PARTE II

FUNDAMENTOS

2. El modelo de regresión lineal tradicional

Este capítulo considera el modelo de regresión lineal simple, esto es, un modelo con una única variable regresora x que tiene una relación con la respuesta y en forma de línea recta, es decir

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.1)$$

donde la intercepción β_0 y la pendiente β_1 son constantes desconocidas, y ε es un componente de error aleatorio. Se supone que los errores tienen media cero y varianza constante desconocida σ^2 . De forma adicional, usualmente se considera que los errores son no correlacionados.

Es conveniente ver a la regresora x como una variable controlada dentro del análisis, medida con un error despreciable, mientras que la respuesta y es una variable aleatoria. Esto es, existe una distribución de probabilidad para y en cada posible valor de x . La media o esperanza de esta distribución es

$$E(y | x) = \beta_0 + \beta_1 x \quad (2.2a)$$

Y la varianza es

$$\text{Var}(y | x) = \text{Var}(\beta_0 + \beta_1 x + \varepsilon_i) = \sigma^2 \quad (2.2b)$$

Entonces la media de y es una función lineal de x , mientras que su varianza no depende del valor de la regresora. Además, debido a que los errores son no correlacionados, las respuestas y también son no correlacionadas.

Los parámetros β_0 y β_1 son usualmente llamados los *coeficientes de la regresión*. Estos tienen una simple y útil interpretación. La pendiente β_1 es el cambio en la media de la distribución de y por unidad de variación en x . Si el rango de los datos incluye el valor $x = 0$, entonces la intercepción β_0 es la media de la respuesta y cuando x toma este valor. Si el rango de observaciones no incluye $x = 0$, entonces β_0 no tiene una interpretación práctica.

2.1. Estimación de los parámetros por mínimos cuadrados

Los parámetros β_0 y β_1 son desconocidos y pueden ser estimados usando una muestra de observaciones, suponiendo por ejemplo que se tienen n parejas de datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Como se ha visto a lo largo de todo el capítulo anterior, estos datos pueden provenir de un estudio diseñado específicamente para recabar los datos, de un estudio observacional, o de un registro histórico de datos.

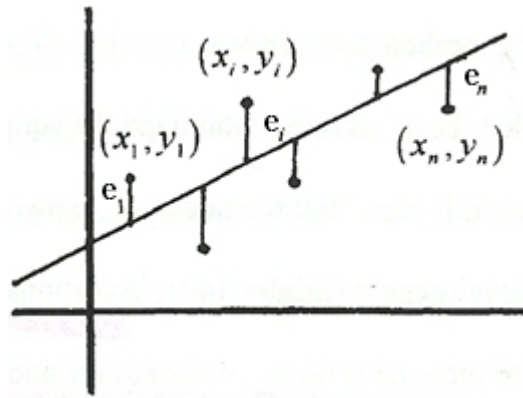


Figura 2.1. Definición tradicional de los residuales.

2.1.1. Estimación de los parámetros

El método de mínimos cuadrados es usado para estimar β_0 y β_1 . Esto es, se deben encontrar sus estimadores tal que la suma de cuadrados de las diferencias entre las observaciones y_i y la línea recta sea un mínimo.

De la ecuación (2.1) se puede escribir

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.3)$$

La ecuación (2.3) es una *muestra del modelo de regresión*, escrita en términos de los n pares de datos (x_i, y_i) ($i = 1, 2, \dots, n$). Entonces, el criterio de mínimos cuadrados establece minimizar la siguiente suma:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2 \quad (2.4)$$

2.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

Los estimadores de β_0 y β_1 , por ejemplo $\hat{\beta}_0$ y $\hat{\beta}_1$, deben satisfacer

$$\begin{aligned}\frac{\partial S}{\partial \beta_1} \Big|_{\hat{\beta}_0, \hat{\beta}_1} &= -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_1 x_i - \hat{\beta}_0) = 0 \\ \frac{\partial S}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1} &= -2 \sum_{i=1}^n (y_i - \hat{\beta}_1 x_i - \hat{\beta}_0) = 0\end{aligned}\quad (2.5)$$

A partir de este momento, se hace una aclaración con respecto a la notación. Debido a la evidente relación entre este *método tradicional* de regresión por mínimos cuadrados y el *alternativo*, revisado en la *Parte III* de este trabajo, es adecuado conservar la *notación usual* del método alternativo en su contexto geométrico de *punto-pendiente* dentro del tradicional para la estimación de sus parámetros; facilitando así la comparación entre ambos métodos. De esta forma se tiene que el estimador $\hat{\beta}_1$ será ahora m , mientras que $\hat{\beta}_0$ será representado por la tradicional ordenada al origen b , quedando entonces la línea estimada $\hat{y} = b + mx$. Se retomará la notación de $\hat{\beta}_0$ y $\hat{\beta}_1$ para los estimadores cuando se estudien sus propiedades, ya que es natural comparar ahora éstas con sus generalizaciones de la *Parte IV*.

Una vez hecha esta aclaración, se establece entonces que para el criterio de mínimos cuadrados hay que minimizar:

$$S(m, b) = \sum_{i=1}^n (y_i - mx_i - b)^2$$

por lo que es posible rescribir las ecuaciones (2.5) de la siguiente manera:

$$\begin{aligned}\frac{\partial S}{\partial m} &= -2 \sum_{i=1}^n x_i (y_i - mx_i - b) = 0 \\ \frac{\partial S}{\partial b} &= -2 \sum_{i=1}^n (y_i - mx_i - b) = 0\end{aligned}\quad (2.6)$$

Los puntos críticos, o extremos de $S(m, b)$, son encontrados resolviendo el sistema (2.6):

$$\begin{aligned}\frac{\partial S}{\partial m} &= -2 \sum_{i=1}^n x_i (y_i - mx_i - b) = 0 \\ \sum_{i=1}^n x_i (y_i - mx_i - b) &= 0\end{aligned}$$

PARTE II. FUNDAMENTOS

$$\begin{aligned} \sum_{i=1}^n x_i y_i - m \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i &= 0 \\ \sum_{i=1}^n x_i y_i &= m \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i \quad (2.7) \\ \frac{\partial S}{\partial b} &= -2 \sum_{i=1}^n (y_i - m x_i - b) = 0 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^n (y_i - m x_i - b) &= 0 \\ \sum_{i=1}^n y_i - m \sum_{i=1}^n x_i - b n &= 0 \\ \frac{1}{n} \sum_{i=1}^n y_i - \frac{m}{n} \sum_{i=1}^n x_i - b &= 0 \end{aligned}$$

Entonces

$$\bar{y} - m\bar{x} - b = 0 \quad (2.8)$$

pero también

$$\sum_{i=1}^n y_i = m \sum_{i=1}^n x_i + b n \quad (2.9)$$

La ecuación (2.7) junto con (2.9) son llamadas *ecuaciones normales*. La intercepción b puede ser encontrada fácilmente de la ecuación (2.8), quedando entonces

$$b = \bar{y} - m\bar{x} \quad (2.10)$$

Aquí se usa la regla de Cramer para resolver las ecuaciones normales para m :

$$m = \frac{\begin{vmatrix} \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i & n \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{vmatrix}} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (2.11)$$

2.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{y} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Para probar que los valores obtenidos en (2.10) y (2.11) minimizan la suma de cuadrados $S(m, b)$, se obtienen sus segundas derivadas parciales con respecto a m y a b .

$$\begin{aligned} \frac{\partial^2 S}{\partial m^2} &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n x_i (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n x_i y_i + 2m \sum_{i=1}^n x_i^2 + 2b \sum_{i=1}^n x_i \right) \\ &= 2 \sum_{i=1}^n x_i^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 S}{\partial b \partial m} &= \frac{\partial^2 S}{\partial m \partial b} = \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial m} \left(-2 \sum_{i=1}^n y_i + 2m \sum_{i=1}^n x_i + 2bn \right) \\ &= 2 \sum_{i=1}^n x_i \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 S}{\partial b^2} &= \frac{\partial}{\partial b} \left(-2 \sum_{i=1}^n (y_i - mx_i - b) \right) \\ &= \frac{\partial}{\partial b} \left(-2 \sum_{i=1}^n y_i + 2m \sum_{i=1}^n x_i + 2bn \right) \\ &= 2n \end{aligned}$$

Se colocan las segundas derivadas parciales en la siguiente matriz:

$$D^2(S(m, b)) = \begin{bmatrix} \frac{\partial^2 S}{\partial b^2} & \frac{\partial^2 S}{\partial m \partial b} \\ \frac{\partial^2 S}{\partial b \partial m} & \frac{\partial^2 S}{\partial m^2} \end{bmatrix} = \begin{bmatrix} 2n & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2 \sum_{i=1}^n x_i^2 \end{bmatrix}$$

Para el punto crítico que minimiza la suma de los cuadrados de los errores, la matriz anterior tiene que ser *definida positiva*. Se dice que una matriz es definida positiva si el determinante de todas las submatrices es positivo. La submatriz superior izquierda de $D^2(S(m, b))$ es siempre positiva, debido a que representa el doble de número de puntos.

PARTE II. FUNDAMENTOS

El determinante de la matriz de la segunda derivada parcial de $S(m,b)$ es encontrado mediante:

$$|D^2(S(m,b))| = 4n \sum_{i=1}^n x_i^2 - 4 \left(\sum_{i=1}^n x_i \right)^2 \quad (2.12)$$

Debido a que se buscan los valores que hacen que la matriz sea definida positiva, se pone a (2.12) mayor a cero llegando a la desigualdad (2.13). Resta entonces demostrarla como verdadera para todos los valores de x_i :

$$\begin{aligned} n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 &> 0 & (2.13) \\ \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 &> 0 \\ \sum_{i=1}^n x_i^2 - n \left(\frac{1}{n^2} \right) \left(\sum_{i=1}^n x_i \right)^2 &> 0 \\ \sum_{i=1}^n x_i^2 - n \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 &> 0 \\ \sum_{i=1}^n x_i^2 - n \bar{x}^2 &> 0 \\ \sum_{i=1}^n x_i^2 + (-2n\bar{x}^2 + n\bar{x}^2) &> 0 \\ \sum_{i=1}^n x_i^2 - 2n\bar{x} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + n\bar{x}^2 &> 0 \\ \sum_{i=1}^n x_i^2 - 2\bar{x} \left(\sum_{i=1}^n x_i \right) + n\bar{x}^2 &> 0 \\ \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) &> 0 \\ \sum_{i=1}^n (x_i - \bar{x})^2 &> 0 & (2.14) \end{aligned}$$

La desigualdad (2.14) es cierta si y sólo si existe algún $x_i \neq \bar{x}$. Es válido suponer que no todas las x_i son iguales debido a que la línea de regresión sería entonces una recta vertical con $x = c$, donde c es una constante.

2.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

La regresión lineal, por el método tradicional de mínimos cuadrados, no puede arrojar líneas verticales. Debido a que el determinante de la matriz para la segunda derivada es siempre positiva, los valores de la pendiente m y la intersección b dados por las ecuaciones (2.10) y (2.11) son los valores que minimizan $S(m,b)$, la suma de los cuadrados de los errores.

La diferencia entre el valor observado y_i y su correspondiente valor ajustado \hat{y}_i se define como un *residual*. Matemáticamente, el i -ésimo residual es

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n$$

Los residuales juegan un papel importante investigando la adecuación del modelo ajustado de regresión detectando violaciones a sus suposiciones, tales como su varianza constante o independencia.

De acuerdo a la definición de los residuales, definidos como la distancia vertical entre las observaciones y la línea de regresión, el método tradicional de regresión por mínimos cuadrados producirá la recta que minimiza estas distancias.

2.1.2. Propiedades de los estimadores y del modelo ajustado de regresión

Como se señaló anteriormente, la notación de $\hat{\beta}_0$ y $\hat{\beta}_1$ será retomada en esta sección en un intento por hacer que la generalización de la *Parte IV* se vea de forma más natural. Entonces, siendo esto, se pueden sustituir $\hat{\beta}_0$ y $\hat{\beta}_1$ en (2.10) y (2.11) quedando

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.15)$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (2.16)$$

y el modelo ajustado de regresión simple es

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (2.17)$$

La ecuación (2.17) da un punto estimado de la esperanza de y para un particular x .

PARTE II. FUNDAMENTOS

Ya que el denominador de la ecuación (2.16) es la suma corregida de cuadrados de las x_i , y el numerador es la suma corregida de los productos cruzados de x_i y y_i , se pueden escribir estas cantidades con una notación más compacta:

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.18)$$

y

$$S_{xy} = \sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n} = \sum_{i=1}^n y_i (x_i - \bar{x}) \quad (2.19)$$

En diversos escritos, S_{xx} también es denotada como S_x^2 . Una forma conveniente de escribir (2.16) es

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (2.20)$$

Los estimadores por mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ tienen varias propiedades importantes. Primero, nótese que de las ecuaciones (2.15) y (2.20), $\hat{\beta}_0$ y $\hat{\beta}_1$ son combinaciones lineales de las observaciones y_i . Por ejemplo,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^n c_i y_i$$

donde $c_i = (x_i - \bar{x}) / S_{xx}$ para $i = 1, 2, \dots, n$.

Los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ por mínimos cuadrados, son estimadores *insesgados* de los parámetros del modelo β_0 y β_1 . Para probar esto en $\hat{\beta}_1$, considérese:

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\sum_{i=1}^n c_i y_i\right) = \sum_{i=1}^n c_i E(y_i) \\ &= \sum_{i=1}^n c_i (\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i \end{aligned}$$

ya que $E(\varepsilon_i) = 0$ por hipótesis. Se puede demostrar directamente que $\sum_{i=1}^n c_i = 0$ y $\sum_{i=1}^n c_i x_i = 1$, por lo que

2.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

$$E(\hat{\beta}_1) = \beta_1$$

Esto es, suponiendo que el modelo es correcto, $\hat{\beta}_1$ es un estimador insesgado para β_1 . Similarmente, se demuestra que $\hat{\beta}_0$ es un estimador insesgado de β_0 , o

$$E(\hat{\beta}_0) = \beta_0$$

La varianza de $\hat{\beta}_1$ es encontrada como

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var}\left(\sum_{i=1}^n c_i y_i\right) \\ &= \sum_{i=1}^n c_i^2 \text{Var}(y_i) \end{aligned} \quad (2.21)$$

ya que las observaciones y_i son independientes, la varianza de la suma es la suma de las varianzas. Para cada término se tiene $c_i^2 \text{Var}(y_i)$, y ya que se supone que $\text{Var}(y_i) = \sigma^2$, entonces

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^n c_i^2 = \frac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(S_{xx})^2} = \frac{\sigma^2}{S_{xx}} \quad (2.22)$$

La varianza de $\hat{\beta}_0$ es

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \text{Var}(\bar{y} - \hat{\beta}_1 \bar{x}) \\ &= \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1) - 2\bar{x} \text{Cov}(\bar{y}, \hat{\beta}_1) \end{aligned}$$

La varianza de \bar{y} es justamente $\text{Var}(\bar{y}) = \sigma^2 / n$, y puede demostrarse fácilmente que la covarianza entre \bar{y} y $\hat{\beta}_1$ es cero. Por lo tanto,

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad (2.23)$$

Otro resultado importante es que el teorema de Gauss-Markov establece que para el modelo de regresión (2.1) con las suposiciones $E(\varepsilon) = 0$ y $\text{Var}(\varepsilon) = \sigma^2$, con errores independientes, los estimadores por mínimos cuadrados son insesgados y tienen la varianza mínima cuando son comparados con todos los demás que son combinaciones lineales de las y_i . Se dice con frecuencia que son los *mejores estimadores lineales e insesgados*, donde “mejores” se refiere a que tienen la varianza mínima. El teorema de

Gauss-Markov será revisado en el **Apéndice D** de la *Parte IV* para un modelo de regresión lineal múltiple, del cual la regresión simple es un caso particular.

Existen otras propiedades prácticas del ajuste de los mínimos cuadrados:

1. La suma de los residuales en cualquier modelo de regresión que contiene una intercepción β_0 es siempre cero, esto es,

$$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0$$

Esta propiedad se sigue directamente de la primera ecuación normal (2.9).

2. La suma de los valores observados y_i iguala la suma de los valores ajustados \hat{y}_i , o

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

3. La línea de regresión por mínimos cuadrados siempre pasa por el *centroide* [el punto (\bar{x}, \bar{y})] de los datos.
4. La suma de residuales, ponderados por el correspondiente valor de la variable regresora, es siempre cero, esto es

$$\sum_{i=1}^n x_i e_i = 0$$

5. La suma de residuales, cargados por el correspondiente valor ajustado, es siempre cero, esto es

$$\sum_{i=1}^n \hat{y}_i e_i = 0$$

2.1.3. Estimación de σ^2

Un estimador de σ^2 también se requiere para realizar las pruebas de hipótesis y construir los intervalos de confianza concernientes al modelo de regresión. El estimador de σ^2 es obtenido a partir de la *suma de los cuadrados de residuales o errores cuadrados*,

$$SC_{\text{Res}} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.24)$$

2.2. PRUEBAS DE HIPÓTESIS PARA LA PENDIENTE Y LA INTERCEPCIÓN

Una fórmula conveniente para SC_{Res} puede ser encontrada sustituyendo $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ en la ecuación (2.24) y simplificando, obteniendo entonces

$$SC_{\text{Res}} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy} \quad (2.25)$$

Pero se define

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

como la suma de cuadrados corregida de las observaciones de la variable de respuesta, por lo que

$$SC_{\text{Res}} = SC_T - \hat{\beta}_1 S_{xy} \quad (2.26)$$

La suma de cuadrados de residuales tiene $n-2$ grados de libertad, debido a que dos grados de libertad están asociados al número de estimadores. En el **Apéndice C** de la *Parte IV* se muestra que el valor esperado de SC_{Res} es $(n-2)\sigma^2$, por lo que un estimador insesgado de σ^2 es

$$\hat{\sigma}^2 = \frac{SC_{\text{Res}}}{n-2} = CM_{\text{Res}} \quad (2.27)$$

La cantidad CM_{Res} es llamada el *cuadrado medio de los residuales*. La raíz cuadrada de $\hat{\sigma}^2$ es llamada el *error estándar de la regresión*, y tiene las mismas unidades de medida que la variable de respuesta y .

2.2. Pruebas de hipótesis para la pendiente y la intercepción

Ahora es de interés realizar pruebas de hipótesis y construir intervalos de confianza para los parámetros del modelo. Las primeras son tratadas en esta sección, mientras que los segundos en la siguiente. Estos procedimientos requieren la suposición adicional de que los errores son normalmente distribuidos. Entonces, todas las suposiciones son que éstos son independientes y normalmente distribuidos con media 0 y varianza σ^2 , que de forma abreviada se escribe $\text{NID}(0, \sigma^2)$ por sus siglas en inglés.

2.2.1. Desarrollo de la prueba de hipótesis sobre la pendiente

Supóngase que se desea probar la hipótesis de que la pendiente es igual a una constante, por ejemplo β_{01} . Las hipótesis apropiadas serían

$$\begin{aligned} H_0 : \beta_1 &= \beta_{01} \\ H_1 : \beta_1 &\neq \beta_{01} \end{aligned} \quad (2.28)$$

donde se han especificado dos alternativas. Debido a que los errores son $NID(0, \sigma^2)$, las observaciones y_i son $NID(\beta_0 + \beta_1 x_i, \sigma^2)$. Ahora, $\hat{\beta}_1$ es una combinación lineal de las observaciones, por lo que $\hat{\beta}_1$ se distribuye normalmente con media β_1 y varianza σ^2 / S_{xx} como se encontró anteriormente. Entonces, la estadística

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{01}}{\sqrt{\sigma^2 / S_{xx}}}$$

tiene una distribución $N(0,1)$ si la hipótesis nula $H_0 : \beta_1 = \beta_{01}$ es verdadera. Si σ^2 fuera conocida, se podría usar Z_0 para probar las hipótesis (2.28), pero generalmente, es desconocida. Ya que CM_{Res} es un estimador insesgado para σ^2 y que en el **Apéndice C** de la *Parte IV* se establece que $(n-2)CM_{Res} / \sigma^2$ sigue una distribución χ^2_{n-2} para el caso específico de la regresión lineal simple, y que CM_{Res} y $\hat{\beta}_1$ son independientes; por la definición dada en el **Apéndice A** de la *Parte IV*, la estadística t_0 queda de la siguiente forma:

$$t_0 = \frac{\hat{\beta}_1 - \beta_{01}}{\sqrt{CM_{Res} / S_{xx}}} \quad (2.29)$$

la cual sigue una distribución t_{n-2} si la hipótesis nula $H_0 : \beta_1 = \beta_{01}$ es verdadera. Los grados de libertad asociados con t_0 son los mismos en número que con CM_{Res} . Entonces, el cociente t_0 es la estadística usada para probar $H_0 : \beta_1 = \beta_{01}$. La prueba calcula t_0 y lo compara con el cuantil $1 - \alpha / 2$ de una distribución t_{n-2} ($t_{\alpha/2, n-2}$). Este procedimiento rechaza H_0 si

$$|t_0| > t_{\alpha/2, n-2} \quad (2.30)$$

De forma alternativa, el cálculo del *P-value* se puede ser usar para la toma de decisiones.

2.2. PRUEBAS DE HIPÓTESIS PARA LA PENDIENTE Y LA INTERCEPCIÓN

El denominador en (2.29) es llamado el *error estándar estimado*, o más fácil, el *error estándar de la pendiente*. Esto es

$$ee(\hat{\beta}_1) = \sqrt{\frac{CM_{\text{Res}}}{S_{xx}}} \quad (2.31)$$

Entonces, es frecuente escribir t_0 como

$$t_0 = \frac{\hat{\beta}_1 - \beta_{01}}{ee(\hat{\beta}_1)} \quad (2.32)$$

Un procedimiento similar puede ser usado para probar la hipótesis de la intercepción:

$$\begin{aligned} H_0 : \beta_0 &= \beta_{00} \\ H_1 : \beta_0 &\neq \beta_{00} \end{aligned} \quad (2.33)$$

Ahora, usando la estadística de prueba

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{CM_{\text{Res}} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} = \frac{\hat{\beta}_0 - \beta_{00}}{ee(\hat{\beta}_0)} \quad (2.34)$$

donde $ee(\hat{\beta}_0) = \sqrt{CM_{\text{Res}} (1/n + \bar{x}^2 / S_{xx})}$ es el *error estándar de la intercepción*, se rechaza la hipótesis nula si $|t_0| > t_{\alpha/2, n-2}$.

2.2.2. Prueba de significancia

Un caso muy importante de las hipótesis (2.28) es

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ H_1 : \beta_1 &\neq 0 \end{aligned} \quad (2.35)$$

Estas hipótesis constituyen la *prueba de significancia de la regresión*. No rechazar $H_0 : \beta_1 = 0$ significa que no existe relación lineal entre x y y . Esta situación se ilustra en la Figura 2.2. Esto puede significar que x tiene un pequeño peso para explicar la variación en y y que el mejor estimador de y para toda x es $\hat{y} = \bar{y}$ (Figura 2.2a), o que la

verdadera relación entre x y y no es lineal (Figura 2.2b). Por lo tanto, no rechazar H_0 es equivalente a decir que no existe relación lineal entre x y y .

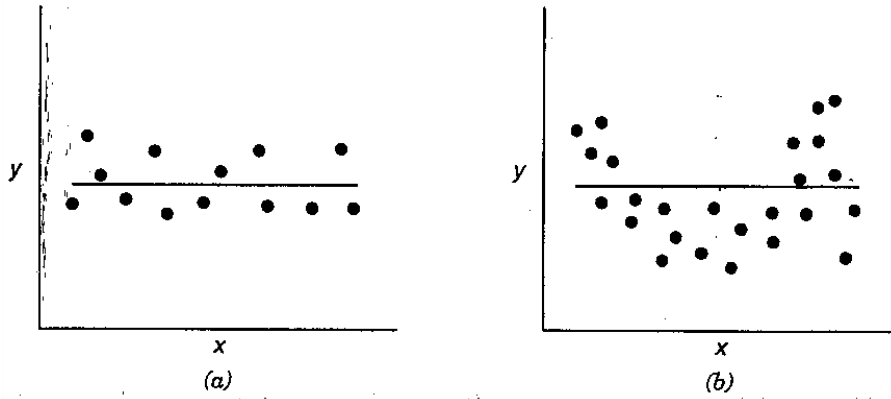


Figura 2.2. Situaciones en donde la hipótesis nula no es rechazada.

Por otro lado, si $H_0 : \beta_1 = 0$ es rechazada, significa que x tiene peso en la explicación de la variabilidad en y y se dice que el modelo tiene sentido (Figura 2.3).

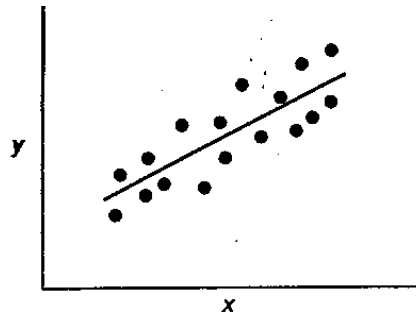


Figura 2.3. Situación en donde la hipótesis nula es rechazada.

El procedimiento para probar $H_0 : \beta_1 = 0$, hace uso de la estadística t de la ecuación (2.32):

$$t_0 = \frac{\hat{\beta}_1}{ee(\hat{\beta}_1)}$$

Se rechaza la hipótesis H_0 si $|t_0| > t_{\alpha/2, n-2}$.

2.2.3. El análisis de varianza

También se puede realizar un *análisis de varianza* para probar la significancia de la regresión. Éste consiste en particionar la variabilidad total de la variable de respuesta y . Para realizar esto se considera la identidad

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \quad (2.36)$$

Elevando al cuadrado ambos lados de (2.36) y sumando sobre todos los valores de las n observaciones se tiene

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

El tercer término de la expresión anterior puede ser rescrito como:

$$\begin{aligned} 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= 2 \sum_{i=1}^n \hat{y}_i (y_i - \hat{y}_i) - 2\bar{y} \sum_{i=1}^n (y_i - \hat{y}_i) \\ &= 2 \sum_{i=1}^n \hat{y}_i e_i - 2\bar{y} \sum_{i=1}^n e_i = 0 \end{aligned}$$

Debido a que la suma de residuales es siempre cero (propiedad 1, sección 2.1.2) y la suma de residuales ponderados por su correspondiente valor ajustado \hat{y}_i es también cero (propiedad 5, sección 2.1.2), se tiene

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.37)$$

El lado izquierdo de la ecuación (2.37) es la *suma de cuadrados corregida de las observaciones* denotada como SC_T , y mide la variabilidad total de la muestra. Sus dos componentes se refieren, respectivamente, a la cantidad de variabilidad en las observaciones y_i representada por la línea de regresión, y a la variación residual no explicada por esta línea. $SC_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ es la *suma de cuadrados de residuales* de la ecuación (2.24). Se acostumbra llamar a $SC_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ *suma de cuadrados debida a la regresión o al modelo*.

La ecuación (2.37) es la *identidad fundamental del análisis de varianza para un modelo de regresión*. De manera compacta se escribe

PARTE II. FUNDAMENTOS

$$SC_T = SC_R + SC_{Res} \quad (2.38)$$

Comparando la ecuación (2.38) con (2.26), otra forma de escribir SC_R es

$$SC_R = \hat{\beta}_1 S_{xy} \quad (2.39)$$

El análisis de los grados de libertad se realiza como sigue: la suma de cuadrados total SC_T tiene $n-1$ grados de libertad, ya que uno de ellos se pierde como resultado de la restricción $\sum_{i=1}^n (y_i - \bar{y}) = 0$ sobre las desviaciones $y_i - \bar{y}$. La suma de cuadrados de la regresión SC_R tiene $df_R = 1$ grado de libertad debido a que es completamente determinada por un parámetro llamado β_1 . Finalmente, se vio antes que SC_R tiene $df_{Res} = n-2$ grados de libertad, ya que dos restricciones son impuestas sobre las desviaciones $y_i - \hat{y}_i$ como resultado de estimar β_0 y β_1 . Los grados de libertad tienen una propiedad adicional:

$$\begin{aligned} df_T &= df_R + df_{Res} \\ n-1 &= 1 + (n-2) \end{aligned} \quad (2.40)$$

Se puede usar la *prueba F usual del análisis de varianza* para probar la hipótesis $H_0 : \beta_1 = 0$. En el **Apéndice C** de la *Parte IV* se prueba que (1) $SC_{Res} = (n-2)CM_{Res}$ sigue una distribución χ_{n-2}^2 ; (2) si la hipótesis nula $H_0 : \beta_1 = 0$ es verdadera, entonces SC_R sigue una distribución χ_{n-2}^2 ; (3) SC_{Res} y SC_R son independientes. Por definición de la estadística F dada en el **Apéndice A** de la *Parte IV* se tiene

$$F_0 = \frac{SC_R / 1}{SC_{Res} / (n-2)} = \frac{CM_R}{CM_{Res}} \quad (2.41)$$

que tiene una distribución $F_{1, n-2}$. En el **Apéndice C** de la *Parte IV* se demuestra que el valor esperado de estos cuadrados medios en el caso particular de la regresión simple son

$$\begin{aligned} E(CM_{Res}) &= \sigma^2 \\ E(CM_R) &= \sigma^2 + \beta_1^2 S_{xx} \end{aligned}$$

Estas esperanzas indican que si el valor observado de F_0 es grande, entonces es probable que la pendiente $\beta_1 \neq 0$. En el **Apéndice C** de la *Parte IV* también se prueba que si

2.3. INTERVALOS DE CONFIANZA EN LA REGRESIÓN LINEAL SIMPLE

$\beta_1 \neq 0$, entonces F_0 sigue una distribución no-central F con 1 y $n-2$ grados de libertad y parámetro de no-centralidad

$$\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$$

Este parámetro también indica que el valor observado de F_0 debería ser grande si $\beta_1 \neq 0$. Entonces, para probar la hipótesis $H_0 : \beta_1 = 0$, se rechaza H_0 si

$$F_0 > F_{\alpha,1,n-2}$$

El análisis de varianza es resumido en la siguiente tabla, también llamada ANOVA.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F_0
Regresión	SC_R	1	CM_R	CM_R / CM_{Res}
Residual	SC_{Res}	$n-2$	CM_{Res}	
Total	SC_T	$n-1$		

Tabla 2.1. Tabla de Análisis de Varianza.

2.3. Intervalos de confianza en la regresión lineal simple

En esta sección se presentan los intervalos de confianza para los estimadores de los parámetros del modelo de regresión. También se verá el intervalo de estimación de la respuesta media $E(y)$ para los valores dados de x . Las suposiciones de normalidad para los errores siguen siendo vigentes.

2.3.1. Intervalos de confianza para β_0 , β_1 y σ^2

La longitud de estos intervalos de confianza es una medida sobre la calidad total de la línea de regresión. Si los errores son independientes y normalmente distribuidos, entonces

PARTE II. FUNDAMENTOS

$$\frac{\hat{\beta}_1 - \beta_1}{ee(\hat{\beta}_1)} \text{ y } \frac{\hat{\beta}_0 - \beta_0}{ee(\hat{\beta}_0)}$$

tienen una distribución t con $n - 2$ grados de libertad. Con esto, un intervalo de confianza del $(1 - \alpha) \times 100\%$ para la pendiente β_1 está dado por

$$\hat{\beta}_1 - t_{\alpha/2, n-2} ee(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} ee(\hat{\beta}_1) \quad (2.42)$$

y un intervalo de confianza del $(1 - \alpha) \times 100\%$ para la intercepción β_0 es

$$\hat{\beta}_0 - t_{\alpha/2, n-2} ee(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2, n-2} ee(\hat{\beta}_0) \quad (2.43)$$

Por otra parte, si los errores son independientes y normalmente distribuidos, en el **Apéndice C** de la *Parte IV* se muestra que la distribución de

$$(n - 2)CM_{\text{Res}} / \sigma^2$$

es una χ^2 con $n - 2$ grados de libertad, de donde:

$$P \left[\chi_{1-\alpha/2, n-2}^2 \leq \frac{(n - 2)CM_{\text{Res}}}{\sigma^2} \leq \chi_{\alpha/2, n-2}^2 \right] = 1 - \alpha$$

Por consecuencia, un intervalo de confianza del $(1 - \alpha) \times 100\%$ para σ^2 es

$$\frac{(n - 2)CM_{\text{Res}}}{\chi_{\alpha/2, n-2}^2} \leq \sigma^2 \leq \frac{(n - 2)CM_{\text{Res}}}{\chi_{1-\alpha/2, n-2}^2} \quad (2.44)$$

2.3.2. Intervalo de confianza para la respuesta media

Uno de los usos del modelo de regresión es poder estimar la respuesta media $E(y)$ para un valor particular de x . Por ejemplo, sea x_0 el nivel de la variable regresora para la cual se desea estimar la respuesta media, esto es $E(y | x_0)$. Se supone que x_0 es cualquier valor de x dentro o fuera del rango de los datos originales usados para ajustar el modelo. Un estimador insesgado de $E(y | x_0)$ es encontrado del modelo ajustado como

$$\widehat{E(y | x_0)} = \hat{\mu}_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (2.45)$$

2.4. PREDICCIÓN DE NUEVAS OBSERVACIONES

Para obtener un intervalo de confianza del $(1-\alpha)\times 100\%$ para $E(y | x_0)$, se ve que $\hat{\mu}_{y|x_0}$ es una variable aleatoria normalmente distribuida, pues es una combinación lineal de las observaciones y_i . La varianza de $\hat{\mu}_{y|x_0}$ es

$$\begin{aligned}\text{Var}(\hat{\mu}_{y|x_0}) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \text{Var}\left[\bar{y} + \hat{\beta}_1(x_0 - \bar{x})\right] \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2(x_0 - \bar{x})^2}{S_{xx}} = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]\end{aligned}$$

ya que, como se había visto anteriormente, $\text{Cov}(\bar{y}, \hat{\beta}_1) = 0$. Entonces, la distribución muestral de

$$\frac{\hat{\mu}_{y|x_0} - E(y | x_0)}{\sqrt{CM_{\text{Res}} (1/n + (x_0 - \bar{x})^2 / S_{xx})}}$$

es una t con $n-2$ grados de libertad. De esta forma, un intervalo de confianza del $(1-\alpha)\times 100\%$ para la respuesta media en el punto x_0 es

$$\begin{aligned}\hat{\mu}_{y|x_0} - t_{\alpha/2, n-2} \sqrt{CM_{\text{Res}} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \\ \leq E(y | x_0) \leq \hat{\mu}_{y|x_0} + t_{\alpha/2, n-2} \sqrt{CM_{\text{Res}} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}\end{aligned}\quad (2.46)$$

La longitud del intervalo de confianza para $E(y | x_0)$ es una función de x_0 . Ésta es un mínimo cuando $x_0 = \bar{x}$ y crece cuando $|x_0 - \bar{x}|$ se incrementa. Intuitivamente esto es razonable, ya que se espera que las mejores estimaciones de y se realicen cerca del centro de gravedad de los datos x , y la precisión se deteriore cuando se mueve hacia los límites del rango de valores de x .

2.4. Predicción de nuevas observaciones

Una aplicación importante del modelo de regresión es la predicción de nuevas observaciones y correspondientes a un nivel específico de la variable regresora x . Si x_0 es el valor de interés, entonces

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (2.47)$$

PARTE II. FUNDAMENTOS

es el estimador puntual del nuevo valor de la respuesta y_0 .

Ahora, se pretende estimar el intervalo para esta observación futura y_0 . El intervalo de confianza para la respuesta media en $x = x_0$ de (2.46) no es apropiado para este problema, ya que es una estimación del intervalo para la media de y -un parámetro-, y no es un planteamiento probabilístico de la distribución sobre futuras observaciones. Se desarrollará un intervalo de predicción para las y_0 .

Se observa que la variable aleatoria

$$\psi = y_0 - \hat{y}_0$$

es normalmente distribuida con media cero y varianza

$$\text{Var}(\psi) = \text{Var}(y_0 - \hat{y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

ya que la observación futura y_0 es independiente de \hat{y}_0 . Si se usa \hat{y}_0 para predecir y_0 , entonces el error estándar de $\psi = y_0 - \hat{y}_0$ es la estadística apropiada para usarla en un intervalo de predicción. Por esto, el intervalo de predicción del $(1-\alpha) \times 100\%$ de confianza para una observación futura en x_0 es

$$\begin{aligned} \hat{y}_0 - t_{\alpha/2, n-2} \sqrt{CM_{\text{Res}} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} &\leq y_0 \\ &\leq \hat{y}_0 + t_{\alpha/2, n-2} \sqrt{CM_{\text{Res}} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \end{aligned} \quad (2.48)$$

Éste tiene una longitud mínima en $x_0 = \bar{x}$ y crece cuando $|x_0 - \bar{x}|$ se incrementa. Comparando (2.48) con (2.46), se observa que el intervalo de predicción en x_0 siempre es mayor que el de confianza en el mismo punto, ya que el primero depende tanto del error del modelo ajustado como del relacionado con futuras observaciones.

2.5. Coeficiente de determinación

El cociente

$$R^2 = \frac{SC_R}{SC_T} = 1 - \frac{SC_{Res}}{SC_T} \quad (2.49)$$

es llamado *coeficiente de determinación*. Debido a que SC_T es una medida de la variabilidad en y sin considerar el efecto de la variable regresora x , y SC_{Res} es una medida de la variabilidad restante en y después de que x ha sido considerada, R^2 es frecuentemente llamada la proporción de la variación explicada por la variable regresora x . Ya que $0 \leq SC_{Res} \leq SC_T$, se sigue que $0 \leq R^2 \leq 1$. Un valor R^2 cercano a 1, significa que la mayoría de la variabilidad en y es explicada por el modelo de regresión.

El coeficiente R^2 debe ser usado con precaución, debido a que siempre es posible hacerlo crecer agregando los suficientes términos al modelo. Por ejemplo, si no hay puntos repetidos (más de una y para la misma x), entonces un polinomio de grado $n-1$ dará un ajuste perfecto $R^2 = 1$ para los n puntos. Cuando existen puntos repetidos, R^2 nunca puede ser exactamente igual a 1, ya que el modelo no puede explicar por completo la variabilidad relacionada al *error puro*. Éste es un concepto que se utiliza en la prueba de carencia de ajuste -en este trabajo no se contempla el tema-.

2.6. Estimación por máxima verosimilitud

El método de mínimos cuadrados puede ser usado para estimar los parámetros de un modelo de regresión lineal sin considerar la forma de la distribución de los errores ε , produciendo también los mejores estimadores lineales insesgados de β_0 y β_1 . Otros procedimientos estadísticos, tales como las pruebas de hipótesis y la construcción de intervalos de confianza, requieren que los errores sean normalmente distribuidos. Si la distribución de los errores es conocida, existe un método alternativo para la estimación de los parámetros: *el método de máxima verosimilitud*.

Considérese los datos (x_i, y_i) , $i = 1, 2, \dots, n$. Si se supone que los errores en el modelo de regresión son $NID(0, \sigma^2)$, entonces las observaciones y_i en la muestra son variables aleatorias independientes y normalmente distribuidas con media $\beta_0 + \beta_1 x_i$ y

PARTE II. FUNDAMENTOS

varianza σ^2 . La función de verosimilitud es encontrada a partir de la distribución conjunta de las observaciones. Si se considera esta distribución junto con las observaciones dadas y los parámetros β_0 , β_1 y σ^2 constantes desconocidas, entonces se tiene la *función de verosimilitud*. Para la regresión simple con errores normales, la función de verosimilitud es

$$\begin{aligned} L(y_i, x_i, \beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right] \end{aligned} \quad (2.50)$$

Los estimadores por máxima verosimilitud, por ejemplo: $\tilde{\beta}_0$, $\tilde{\beta}_1$ y $\tilde{\sigma}^2$, maximizan L , o de forma equivalente, $\ln L$. Esto es

$$\begin{aligned} \ln L(y_i, x_i, \beta_0, \beta_1, \sigma^2) &= -\left(\frac{n}{2}\right) \ln 2\pi - \left(\frac{n}{2}\right) \ln \sigma^2 \\ &\quad - \left(\frac{1}{2\sigma^2}\right) \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned} \quad (2.51)$$

y los *estimadores máximo-verosímiles* $\tilde{\beta}_0$, $\tilde{\beta}_1$ y $\tilde{\sigma}^2$ deben satisfacer

$$\left. \frac{\partial \ln L}{\partial \beta_0} \right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2} \sum_{i=1}^n (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i) = 0 \quad (2.52a)$$

$$\left. \frac{\partial \ln L}{\partial \beta_1} \right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2} \sum_{i=1}^n (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i) x_i = 0 \quad (2.52b)$$

y

$$\left. \frac{\partial \ln L}{\partial \sigma^2} \right|_{\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2} = -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \sum_{i=1}^n (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2 = 0 \quad (2.52c)$$

Las soluciones al sistema de ecuaciones (2.52) dan los estimadores máximo-verosímiles:

2.7. CASO CUANDO LA REGRESORA x ES ALEATORIA

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x} \quad (2.53a)$$

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.53b)$$

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2}{n} \quad (2.53c)$$

Los estimadores por máxima verosimilitud de la intercepción y de la pendiente son idénticos a los obtenidos por mínimos cuadrados. Por su parte, $\tilde{\sigma}^2$ es un estimador sesgado de σ^2 (ya que $E(\tilde{\sigma}^2) = [(n-2)/n]\sigma^2$). Este sesgo es pequeño si n es moderadamente grande. Normalmente, es usado el estimador insesgado $\hat{\sigma}^2$.

En general, los estimadores máximo-verosímiles tienen mejores *propiedades estadísticas* que los mismos por mínimos cuadrados. Los primeros, son insesgados y tienen *varianza mínima* cuando se les compara con todos los demás estimadores. Son consistentes, y son también un conjunto de *estadísticas suficientes* -esto significa que los estimadores contienen toda la información relativa a los parámetros de la muestra original de tamaño n -. Por otra parte, requieren suposiciones estadísticas más fuertes que los estimadores por mínimos cuadrados. Éstos últimos requieren sólo suposiciones de segundos momentos -sobre el valor esperado, las varianzas, y las covarianzas entre los errores aleatorios-. Los estimadores máximo-verosímiles requieren una suposición completa de su distribución, en este caso, se supone que los errores tienen una distribución normal con los mismos segundos momentos requeridos por los estimadores por mínimos cuadrados.

2.7. Caso cuando la regresora x es aleatoria

En el modelo lineal que se ha presentado en este capítulo, se supone que la variable regresora x es una constante conocida. Hay muchas situaciones en las que fijar los valores de x es inapropiado. En la vida real, es más razonable que tanto y como x sean variables aleatorias. Afortunadamente, bajo ciertas circunstancias, todos los resultados mostrados sobre la estimación de parámetros, pruebas de hipótesis, construcción de intervalos de confianza y predicción, son válidos. Suponiendo que x y y son variables

aleatorias conjuntamente distribuidas -pero la forma de esta distribución es desconocida-, se puede demostrar que todos los resultados mostrados de la regresión son válidos si las siguientes condiciones se satisfacen:

1. La distribución condicional de y dado x es normal con media $\beta_0 + \beta_1 x$ y varianza condicional σ^2 .
2. Las x son variables aleatorias independientes cuya distribución no incluye a β_0 , β_1 , ni a σ^2 .

Un ejemplo numérico de esta situación es construido en el capítulo siguiente, cuando se compare el modelo tradicional de regresión por mínimos cuadrados con un método alternativo, también por mínimos cuadrados, y se encontrará que éste último es “ligeramente” mejor que el primero en el caso específico cuando las x son aleatorias.

2.8. Un ejemplo numérico en R

En este ejemplo numérico se muestra una aplicación del cálculo de los estimadores por mínimos cuadrados en el modelo de regresión lineal simple. Antes de usar **R**, se consideran algunos cálculos que reforzarán las ideas de la teoría aquí desarrollada. Después, estos cálculos serán verificados en **R**, mostrando lo simple que se vuelven con la ayuda de este tipo de software. **R** es un lenguaje y entorno de programación para análisis estadístico y gráfico. **R** es quizá el lenguaje más utilizado por la comunidad estadística, siendo además muy popular en el campo de la investigación biomédica, la bioinformática y las matemáticas financieras.

2.8.1. Los cohetes de la NASA

Los propelentes, usados para motores de cohete, son sustancias muy diversas pudiendo estar en estado sólido, líquido o mixto. Éstos reaccionan en la cámara de empuje o combustión, generando gases a alta presión y gran temperatura. Cuando estos gases salen por la tobera, generan el empuje necesario para elevar y acelerar el cohete. En este ejemplo, la NASA fabrica un motor de cohete mezclando hidrógeno líquido (combustible) y oxígeno líquido (oxidante) como propelentes. La resistencia de la cámara de empuje a la reacción de los dos tipos de agentes propulsores es una característica importante de calidad en los motores de cohetes, medida en libras por pulgada cuadrada (psi). Se

2.8. UN EJEMPLO NUMÉRICO EN R

sospecha que esta resistencia está relacionada con la duración en semanas de la reacción. Veinte observaciones sobre la resistencia de la cámara de combustión y duración de los propelentes han sido recabadas y se muestran en la siguiente tabla.

Observación	Resistencia de la cámara (psi)	Duración de la reacción (semanas)
i	y_i	x_i
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

Tabla 2.2. Observaciones para el ejemplo de los cohetes.

El siguiente *diagrama de dispersión* sugiere que existe una relación lineal entre la resistencia de la cámara de combustión y la duración de la reacción. Una supuesta relación mediante un modelo de línea recta $y = \beta_0 + \beta_1 x + \varepsilon$ parece ser razonable.

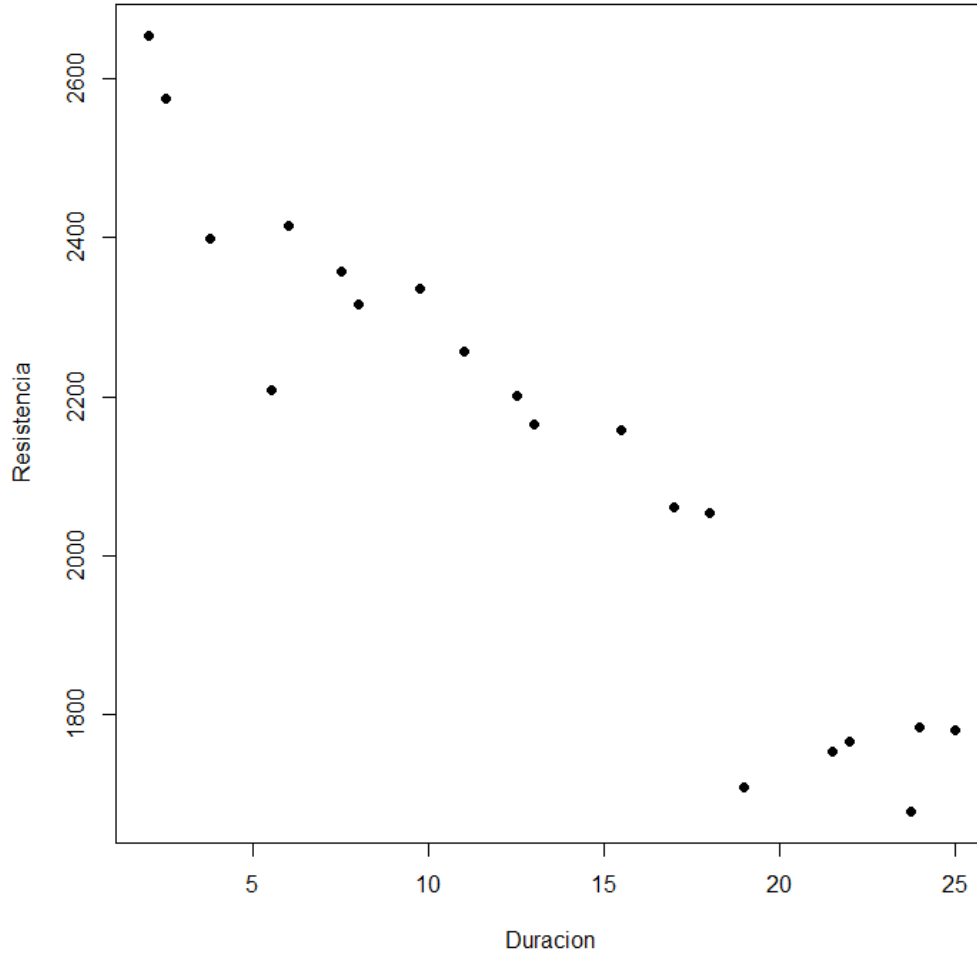


Figura 2.4. Diagrama de dispersión.

Estimación de los parámetros

Primero se calcula:

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = 4677.69 - \frac{71,422.56}{20} = 1106.56$$

y

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} = 528,492.64 - \frac{(267.25)(42,627.15)}{20} = -41,112.65$$

2.8. UN EJEMPLO NUMÉRICO EN R

Entonces, de las ecuaciones (2.20) y (2.15) se encuentra que

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-41,112.65}{1106.56} = -37.15$$

y

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2131.3575 - (-37.15)(13.3625) = 2627.82$$

La línea ajustada por mínimos cuadrados es entonces

$$\hat{y} = 2627.82 - 37.15x$$

Se interpreta la pendiente como el decremento semanal promedio de la resistencia de la cámara de combustión debido a la reacción de los propelentes. La intercepción representa la resistencia del motor justo después de haber sido fabricado.

Para estimar σ^2 , primero se encuentra

$$\begin{aligned} SC_T &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \\ &= 92,547,433.45 - \frac{(42,627.15)^2}{20} = 1,693,737.60 \end{aligned}$$

La suma de cuadrados de residuales es

$$SC_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 166,254.90$$

Entonces, el estimador de σ^2 es encontrado de la ecuación (2.27) como sigue

$$\hat{\sigma}^2 = CM_{Res} = \frac{SC_{Res}}{n-2} = \frac{166,254.90}{18} = 9236.38$$

Prueba de significancia para la regresión

Primero se calculan los errores estándar de los estimadores:

$$ee(\hat{\beta}_1) = \sqrt{\frac{CM_{Res}}{S_{xx}}} = \sqrt{\frac{9236.38}{1106.56}} = 2.89$$

$$ee(\hat{\beta}_0) = \sqrt{CM_{\text{Res}} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} = \sqrt{9236.38 \left(\frac{1}{20} + \frac{(13.3625)^2}{1106.56} \right)} = 44.18$$

Por lo que las estadísticas de prueba respectivas están dadas por

$$t_0 = \frac{\hat{\beta}_1}{ee(\hat{\beta}_1)} = \frac{-37.15}{2.89} = -12.85$$

$$t_0 = \frac{\hat{\beta}_0}{ee(\hat{\beta}_0)} = \frac{2627.82}{44.18} = 59.48$$

Elijiendo $\alpha = 0.05$, el valor crítico de t es $t_{0.025,18} = 2.101 < |-12.85|$. Entonces se rechaza $H_0 : \beta_1 = 0$ y se concluye que tiene sentido el modelo.

De la misma forma con $t_{0.025,18} = 2.101 < |59.48|$, se rechaza $H_0 : \beta_0 = 0$, por lo que se concluye que la intercepción no es cero dentro del modelo.

Análisis de varianza

La otra forma para probar la significancia de la regresión es mediante el *análisis de varianza*, calculando primero la suma de cuadrados debida a la regresión

$$SC_R = SC_T - SC_{\text{Res}} = 1,693,737.80 - 166,254.90 = 1,527,482.90$$

De esta forma, la estadística F esta dada por

$$F_0 = \frac{SC_R / 1}{SC_{\text{Res}} / (n - 2)} = \frac{1,527,842.90 / 1}{166,254.90 / 18} = 165.41$$

Tomando $\alpha = 0.05$, se tiene que $F_{0.05,1,18} = 4.41 < 165.41$. El *p-value* (obtenido directamente de **R**) para esta prueba es de 1.64E-10. Por consecuencia, se rechaza $H_0 : \beta_1 = 0$ ó se dice que la prueba es significativa.

Se construye la tabla de análisis de varianza (ANOVA) con los datos obtenidos:

2.8. UN EJEMPLO NUMÉRICO EN R

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F_0	P value
Regresión	1,527,482.90	1	1,527,482.90	165.41	1.64×10^{-10}
Residual	166,254.90	18	9,236.38		
Total	1,693,737.60	19			

Tabla 2.3. Tabla de ANOVA para el ejemplo de los cohetes.

Intervalos de confianza para β_1 y β_0

Se van a construir intervalos de confianza del 95% para β_1 y β_0 . Ya se habían obtenido los errores estándar de los estimadores $ee(\hat{\beta}_1) = 2.89$ y $ee(\hat{\beta}_0) = 44.18$. También se tenía $t_{0.025,18} = 2.101$. Entonces, de la ecuación (2.42), el intervalo de confianza del 95% para la pendiente es

$$\begin{aligned} \hat{\beta}_1 - t_{0.025,18} ee(\hat{\beta}_1) &\leq \beta_1 \leq \hat{\beta}_1 + t_{0.025,18} ee(\hat{\beta}_1) \\ -37.15 - (2.101)(2.89) &\leq \beta_1 \leq -37.15 + (2.101)(2.89) \\ -43.22 &\leq \beta_1 \leq -31.08 \end{aligned}$$

En otras palabras, 95% de tales intervalos incluirán el valor verdadero de la pendiente.

De la misma manera, de la ecuación (2.43), el intervalo de confianza del 95% para la intercepción es

$$\begin{aligned} \hat{\beta}_0 - t_{0.025,18} ee(\hat{\beta}_0) &\leq \beta_0 \leq \hat{\beta}_0 + t_{0.025,18} ee(\hat{\beta}_0) \\ 2627.82 - (2.101)(44.18) &\leq \beta_0 \leq 2627.82 + (2.101)(44.18) \\ 2534.98 &\leq \beta_0 \leq 2720.64 \end{aligned}$$

El 95% de tales intervalos incluirán el valor verdadero de la intercepción.

Coefficiente de determinación R^2

De la ecuación (2.49) se tiene

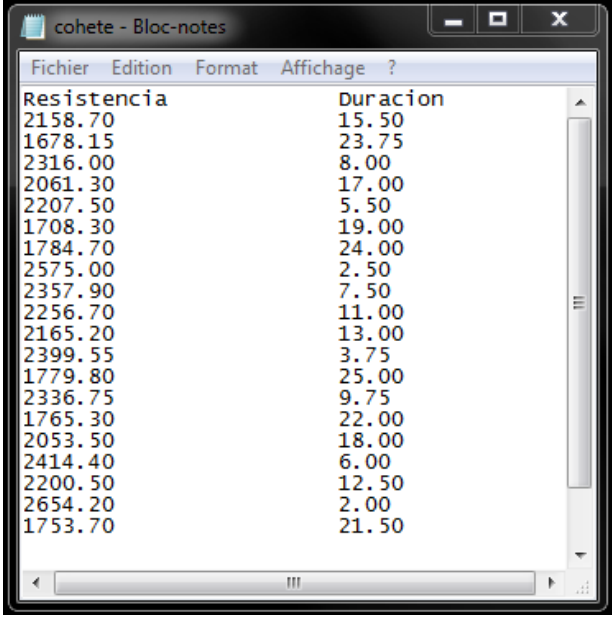
$$R^2 = \frac{SC_R}{SC_T} = \frac{1,527,482.90}{1,693,737.80} = 0.9018$$

Esto es, 90.18% de la variabilidad en la resistencia del motor está representada por el modelo de regresión planteado.

Para la estimación por intervalos para la respuesta media, así como para el de predicción, se obtienen los resultados únicamente sobre el lenguaje **R**.

2.8.2. Resultados en R para el ejemplo de los cohetes

Se comienza guardando los datos en un archivo .txt como se muestra en la siguiente imagen.



Resistencia	Duracion
2158.70	15.50
1678.15	23.75
2316.00	8.00
2061.30	17.00
2207.50	5.50
1708.30	19.00
1784.70	24.00
2575.00	2.50
2357.90	7.50
2256.70	11.00
2165.20	13.00
2399.55	3.75
1779.80	25.00
2336.75	9.75
1765.30	22.00
2053.50	18.00
2414.40	6.00
2200.50	12.50
2654.20	2.00
1753.70	21.50

Figura 2.5. Datos para el ejemplo de los cohetes.

Para acceder a los datos, se usa la función `setwd()` especificando la ruta en donde se encuentra almacenado el archivo .txt. Se cargan los datos leyendo el archivo con la función `read.table()` incluyendo los nombres de las etiquetas de la siguiente forma:

```
> setwd("c:/Users/Gérard/Documents/R")
> datos.cohete=read.table(file="cohete.txt",header=TRUE)
```

2.8. UN EJEMPLO NUMÉRICO EN R

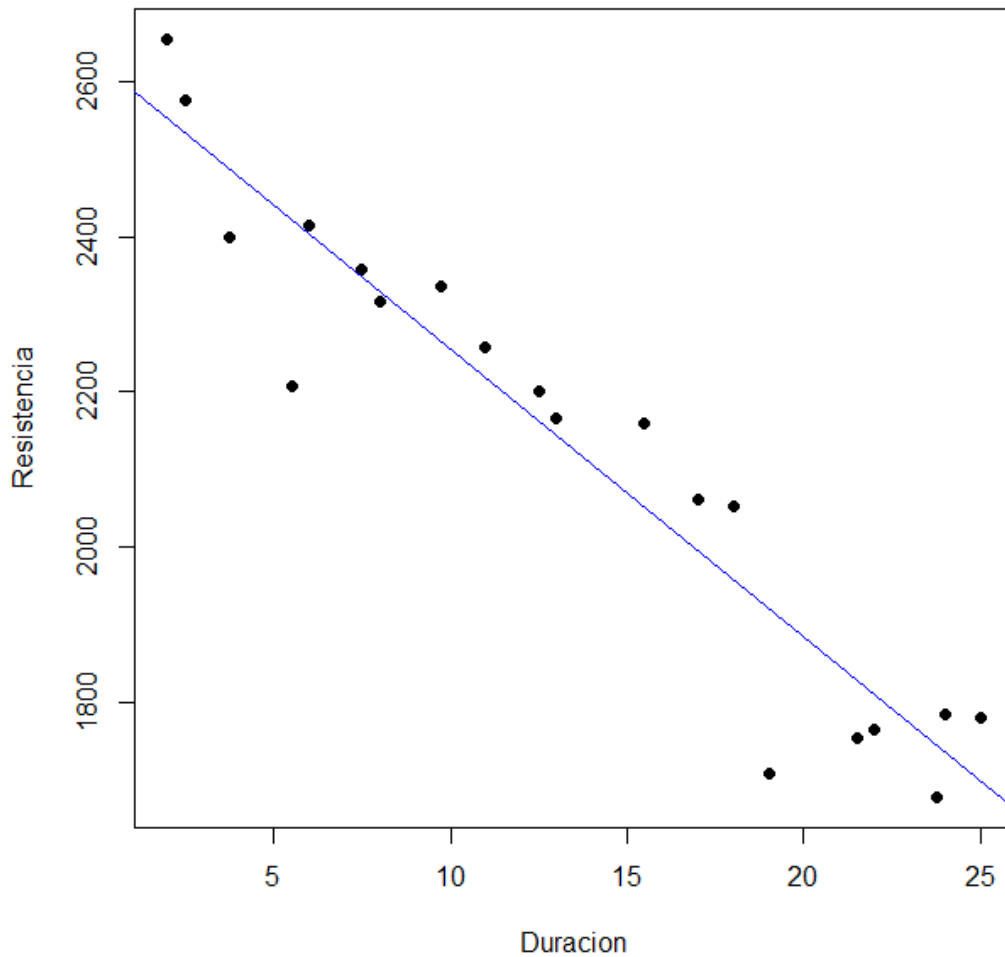
Se asignan las columnas a cada variable y se grafica el diagrama de dispersión.

```
> Resistencia<-datos.cohete[,1]
> Duracion<-datos.cohete[,2]
> plot(Duracion,Resistencia,col="black",pch=19)
```

Estimación de los parámetros

Se usa la fórmula de R para el modelo de regresión Resistencia~Duracion dentro de la función lm(), y se grafica la línea ajustada. Esto es:

```
> regresion=lm(Resistencia~Duracion)
> abline(regresion,col="blue")
```



PARTE II. FUNDAMENTOS

La función `summary()` despliega la información del ajuste:

```
> summary(regresion)

Call:
lm(formula = Resistencia ~ Duracion)

Residuals:
    Min       1Q   Median       3Q      Max
-215.98  -50.68   28.74   66.61  106.76

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2627.822    44.184   59.48 < 2e-16 ***
Duracion    -37.154     2.889  -12.86 1.64e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.11 on 18 degrees of freedom
Multiple R-squared:  0.9018,    Adjusted R-squared:  0.8964
F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10
```

Los resultados obtenidos anteriormente de forma manual pueden ser verificados.

Los valores ajustados pueden ser obtenidos usando la función `predict()`, y los residuales mediante la función `resid()`.

```
> data.frame(datos.cohete, valores.ajustados=predict(regresion),
+ residuales=resid(regresion))
  Resistencia Duracion valores.ajustados residuales
1    2158.70    15.50    2051.942  106.758301
2    1678.15    23.75    1745.425  -67.274574
3    2316.00     8.00    2330.594  -14.593631
4    2061.30    17.00    1996.211   65.088687
5    2207.50     5.50    2423.478 -215.977609
6    1708.30    19.00    1921.904 -213.604131
7    1784.70    24.00    1736.136   48.563824
8    2575.00     2.50    2534.938   40.061618
9    2357.90     7.50    2349.170    8.729573
10   2256.70    11.00    2219.133   37.567141
11   2165.20    13.00    2144.826   20.374323
12   2399.55     3.75    2488.496  -88.946393
13   1779.80    25.00    1698.983   80.817415
14   2336.75     9.75    2265.575   71.175153
15   1765.30    22.00    1810.443  -45.143358
16   2053.50    18.00    1959.058   94.442278
17   2414.40     6.00    2404.901    9.499187
18   2200.50    12.50    2163.402   37.097528
19   2654.20     2.00    2553.515  100.684823
20   1753.70    21.50    1829.020  -75.320154
```

2.8. UN EJEMPLO NUMÉRICO EN R

Análisis de varianza

Se puede desplegar la tabla de análisis de varianza con la función `anova()`.

```
> anova(regresion)
Analysis of Variance Table

Response: Resistencia
      Df Sum Sq Mean Sq F value    Pr(>F)
Duracion  1 1527483 1527483  165.38 1.643e-10 ***
Residuals 18  166255    9236
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los resultados obtenidos en la ANOVA construida manualmente pueden ser verificados.

Intervalos de confianza para β_1 y β_0

Primero, se verá que es posible obtener la matriz de varianzas y covarianzas de los parámetros estimados gracias a la función `vcov()`:

```
> vcov(regresion)
      (Intercept) Duracion
(Intercept) 1952.2181 -111.535941
Duracion    -111.5359   8.346937
```

Con esto, es posible extraer fácilmente los errores estándar de $\hat{\beta}_1$ y $\hat{\beta}_0$ que, junto con el valor de $t_{0,025,18}$, se encuentran los intervalos de confianza del 95% para β_1 y β_0 .

```
> n=length(Resistencia)
> SE0=sqrt(vcov(regresion)[1,1]) #error estándar de B0 gorro
> SE1=sqrt(vcov(regresion)[2,2]) #error estándar de B1 gorro
> tstar = qt(1 - 0.05/2,df= n - 2) #valor de t
> B0.gorro=coef(regresion)[1] #primer coeficiente
> B1.gorro=coef(regresion)[2] #segundo coeficiente
> c(B0.gorro - tstar*SE0, B0.gorro + tstar*SE0)
(Intercept) (Intercept)
 2534.995    2720.649
> c(B1.gorro - tstar*SE1, B1.gorro + tstar*SE1)
Duracion Duracion
-43.22338 -31.08380
```

Otra forma de obtener los intervalos de confianza para los parámetros, es haciéndolo de manera directa con la función `confint()` especificando el nivel de confianza.

PARTE II. FUNDAMENTOS

```
> confint (regresion, level=0.95)
              2.5 %      97.5 %
(Intercept) 2534.99540 2720.6493
Duracion     -43.22338  -31.0838
```

Estimación de nuevas observaciones y de la respuesta media

La función `predict()` es usada para hacer diferentes tipos de predicción. Su estructura general es la siguiente: `predict (name, newdata=..., interval=..., level =...)`

Como se ha visto, la función `predict()` regresa los valores de \hat{y} para cada x cuando su único argumento es el nombre dado para la regresión (`name=regresión` en este ejemplo).

También se puede usar `predict()` para encontrar los estimados de nuevas observaciones o de respuesta media. Para esto se usa la estructura:

```
predict(name, newdata=data.frame(x=x0))
```

Por ejemplo, a continuación se calcula la respuesta media $\hat{\mu}_{y|x_0}$ para $x_0 = 0, 5, 10$:

```
> predict (regresion, newdata=data.frame(Duracion=0))
      1
2627.822
> predict (regresion, newdata=data.frame(Duracion=5))
      1
2442.054
> predict (regresion, newdata=data.frame(Duracion=10))
      1
2256.286
```

De igual forma, se obtiene la predicción de nuevas observaciones \hat{y}_0 para $x_0 = 30, 50, 70$:

```
> predict (regresion, newdata=data.frame(Duracion=30))
      1
1513.215
> predict (regresion, newdata=data.frame(Duracion=50))
      1
 770.1428
> predict (regresion, newdata=data.frame(Duracion=70))
      1
 27.07099
```

2.8. UN EJEMPLO NUMÉRICO EN R

Predicción de intervalos para nuevas observaciones

La predicción de intervalos se realiza para cada valor de x en forma simultanea. A menudo se trazan dos curvas en el diagrama de dispersión para mostrar los límites superior e inferior. Se especifica ahora `interval="prediction"` y un nivel de confianza, que por defecto es del 95% . Se tiene entonces:

```
> pred.nuevas = predict(regresion, interval = "prediction")
Mensajes de aviso perdidos
In predict.lm(regresion, interval = "prediction") :
  Predictions on current data refer to _future_ responses

> pred.nuevas
      fit      lwr      upr
1  2051.942 1844.638 2259.246
2  1745.425 1529.133 1961.716
3  2330.594 2121.151 2540.036
4  1996.211 1788.139 2204.284
5  2423.478 2211.147 2635.808
6  1921.904 1712.196 2131.612
7  1736.136 1519.398 1952.875
8  2534.938 2317.789 2752.088
9  2349.170 2139.235 2559.106
10 2219.133 2011.739 2426.527
11 2144.826 1937.916 2351.735
12 2488.496 2273.529 2703.464
13 1698.983 1480.359 1917.606
14 2265.575 2057.519 2473.631
15 1810.443 1597.006 2023.880
16 1959.058 1750.254 2167.861
17 2404.901 2193.232 2616.570
18 2163.402 1956.439 2370.366
19 2553.515 2335.425 2771.605
20 1829.020 1616.308 2041.732
```

El valor de retorno es una matriz. Se desea ahora graficar los límites superiores e inferiores de estos intervalos. Debido a que la variable *Duración* no esta ordenada, será un error graficar simplemente los datos de salida. Hay que ordenarlos primero. Se usa entonces la función `sort(unique())` para ordenar los valores de x en forma ascendente:

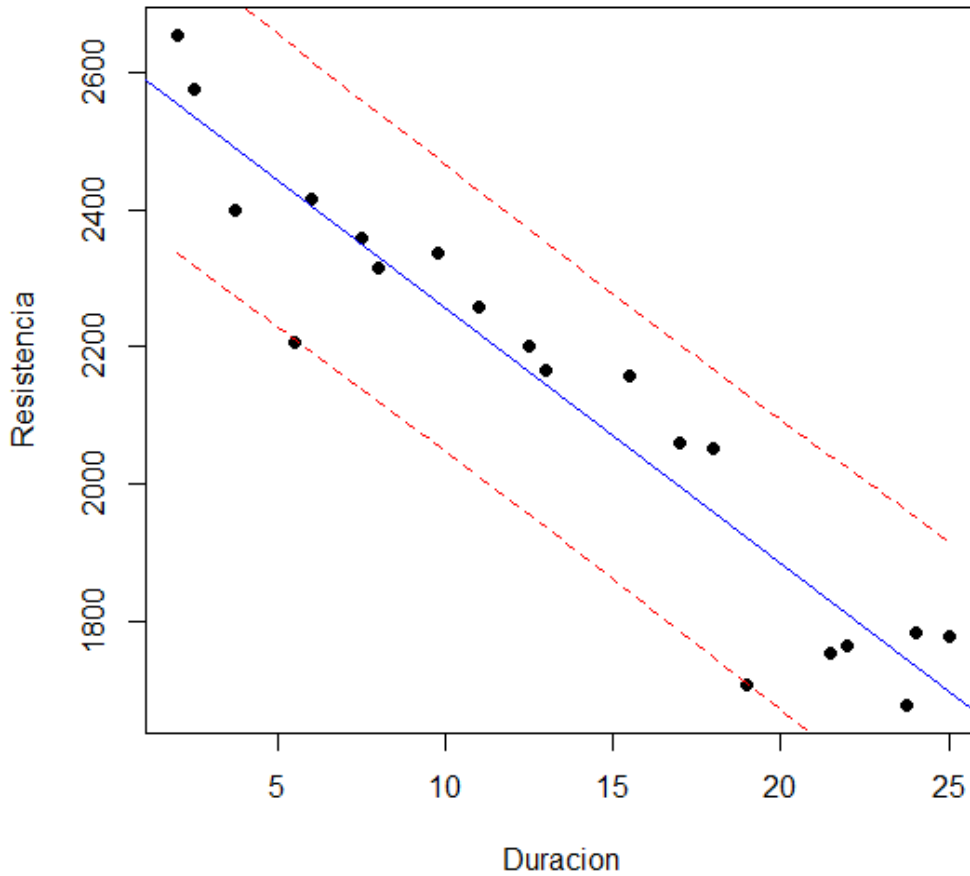
PARTE II. FUNDAMENTOS

```
> duracion.ord = sort(unique(Duracion))
> inter.nuevas = predict(regresion, newdata =
+ data.frame(Duracion=duracion.ord), interval="prediction")
> inter.nuevas
      fit      lwr      upr
1  2553.515 2335.425 2771.605
2  2534.938 2317.789 2752.088
3  2488.496 2273.529 2703.464
4  2423.478 2211.147 2635.808
5  2404.901 2193.232 2616.570
6  2349.170 2139.235 2559.106
7  2330.594 2121.151 2540.036
8  2265.575 2057.519 2473.631
9  2219.133 2011.739 2426.527
10 2163.402 1956.439 2370.366
11 2144.826 1937.916 2351.735
12 2051.942 1844.638 2259.246
13 1996.211 1788.139 2204.284
14 1959.058 1750.254 2167.861
15 1921.904 1712.196 2131.612
16 1829.020 1616.308 2041.732
17 1810.443 1597.006 2023.880
18 1745.425 1529.133 1961.716
19 1736.136 1519.398 1952.875
20 1698.983 1480.359 1917.606
```

Se aprecia que los valores están ahora ordenados. Se procede a agregar las curvas a la gráfica que representan los intervalos del 95% de confianza para nuevas observaciones. Esto se hace con la función `lines()`.

```
> lines(duracion.ord,inter.nuevas[,2] , lty=2,col="red")
> lines(duracion.ord,inter.nuevas[,3] , lty=2,col="red")
```

2.8. UN EJEMPLO NUMÉRICO EN R



Se puede decir con certeza que la anchura de esta franja alcanza su mínimo justo en el punto $(\bar{x} = 13.3625, \bar{y} = 2131.40)$.

Estimación de intervalos para la respuesta media

El procedimiento para estimar los intervalos para la respuesta media $\hat{\mu}_{y|x_0}$ es el mismo que en la predicción de intervalos para nuevas observaciones. La única diferencia, es que ahora se cambia el argumento a `interval="confidence"`. Se tiene entonces:

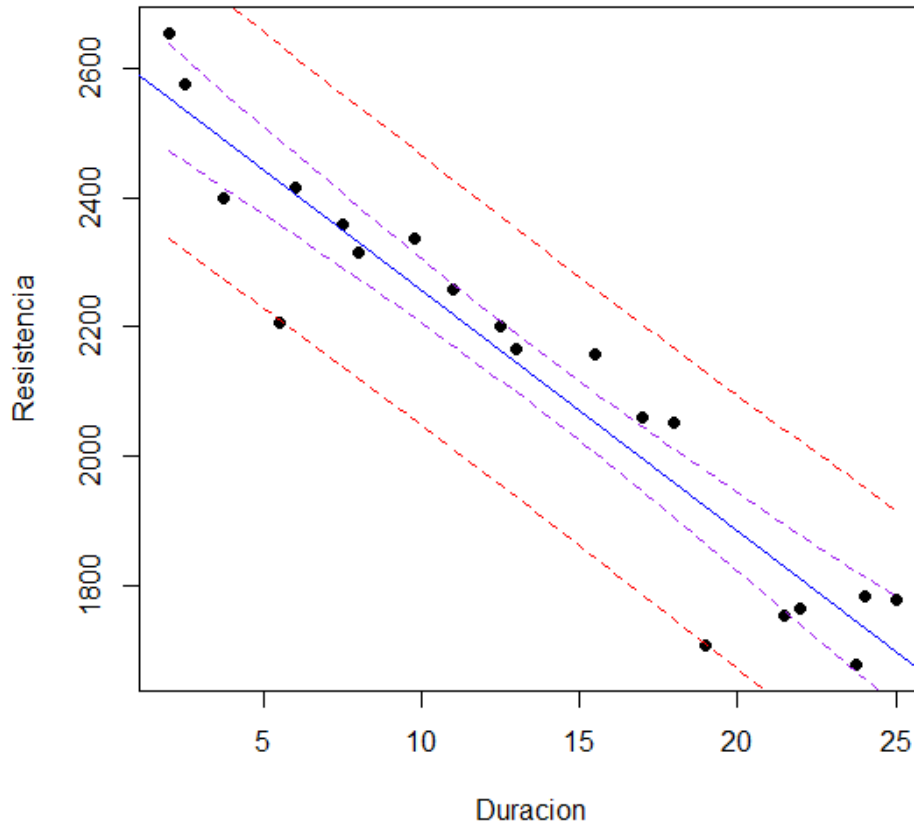
PARTE II. FUNDAMENTOS

```
> duracion.ord = sort(unique(Duracion))
> inter.respmed = predict(regresion, newdata =
+ data.frame(Duracion=duracion.ord), interval="confidence")
> inter.respmed
      fit      lwr      upr
1  2553.515 2471.083 2635.947
2  2534.938 2455.029 2614.848
3  2488.496 2414.722 2562.271
4  2423.478 2357.782 2489.174
5  2404.901 2341.375 2468.426
6  2349.170 2291.684 2406.656
7  2330.594 2274.935 2386.252
8  2265.575 2215.383 2315.767
9  2219.133 2171.762 2266.504
10 2163.402 2117.951 2208.854
11 2144.826 2099.623 2190.028
12 2051.942 2004.966 2098.918
13 1996.211 1945.953 2046.470
14 1959.058 1905.853 2012.263
15 1921.904 1865.253 1978.555
16 1829.020 1762.102 1895.939
17 1810.443 1741.255 1879.632
18 1745.425 1667.877 1822.973
19 1736.136 1657.349 1814.923
20 1698.983 1615.149 1782.816
```

Agregando las curvas a la gráfica de la misma forma que en el caso anterior:

```
> lines(duracion.ord,inter.respmed[,2] , lty=2,col="purple")
> lines(duracion.ord,inter.respmed[,3] , lty=2,col="purple")
```

2.8. UN EJEMPLO NUMÉRICO EN R



Nuevamente, es posible asegurar que la anchura de esta franja alcanza su mínimo en el punto $(\bar{x} = 13.3625, \bar{y} = 2131.40)$.

PARTE III ALTERNANCIA

3. Método alternativo de regresión lineal por mínimos cuadrados

El método alternativo de regresión lineal por mínimos cuadrados se deriva de una forma similar al tradicional presentado anteriormente. Se siguen los mismos pasos, pero la definición de los residuales, o distancia de un punto a la línea de regresión, es ahora diferente. Las fórmulas derivadas encontrarán la “mejor” línea de acuerdo a la definición alternativa para los residuales.

3.1. Estimación de los parámetros por mínimos cuadrados

Una recta con pendiente m y ordenada al origen b es paralela al vector $[1, m]$ y perpendicular a $[m, -1]$. La distancia perpendicular, d_i , entre un punto (x_i, y_i) y la recta $y = mx + b$ puede ser encontrada de la siguiente manera. Considérese que (x_0, y_0) es un punto sobre la recta y \vec{a} es el vector que va de (x_0, y_0) a (x_i, y_i) .

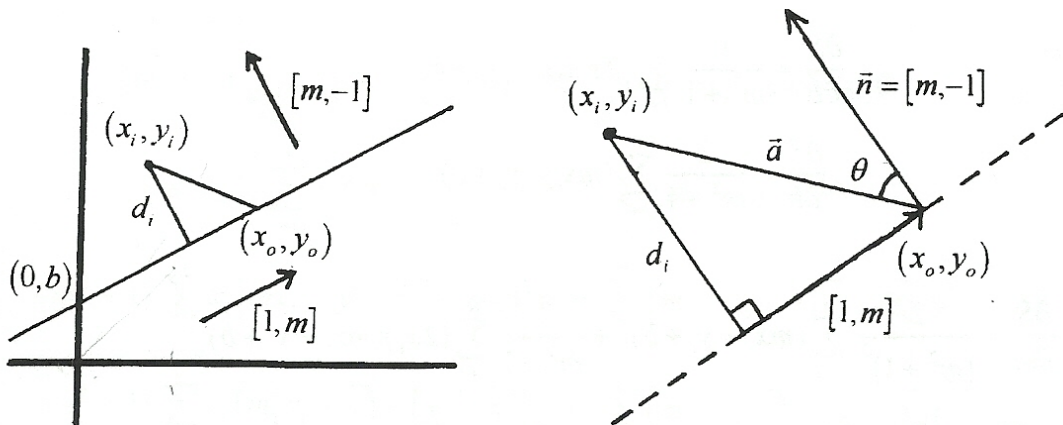


Figura 3.1. Definición alternativa de los residuales.

$$d_i = \|\vec{a}\| (\cos \theta) \left(\frac{\|\vec{n}\|}{\|\vec{n}\|} \right)$$

PARTE III. ALTERNANCIA

$$d_i = \frac{\vec{a} \cdot \vec{n}}{\|\vec{n}\|}$$

$$d_i = \frac{[x_i - x_0, y_i - y_0] \cdot [m, -1]}{\sqrt{m^2 + 1}}$$

$$d_i = \frac{m(x_i - x_0) - (y_i - y_0)}{\sqrt{m^2 + 1}}$$

Como en el método tradicional, se usa el cuadrado de la distancia, llamada ahora d_i^2 , para encontrar la suma de los cuadrados de los residuales $S(m, b)$, o el error total de la línea de regresión.

$$d_i^2 = \frac{(m(x_i - x_0) - (y_i - y_0))^2}{m^2 + 1}$$

$$= \frac{(mx_i - y_i + (y_0 - mx_0))^2}{m^2 + 1}$$

$$\therefore d_i^2 = \frac{(mx_i - y_i + b)^2}{m^2 + 1}$$

Entonces:

$$S(m, b) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \left(\frac{(mx_i - y_i + b)^2}{m^2 + 1} \right)$$

o alternativamente:

$$S(m, b) = \frac{1}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b)^2$$

Para minimizar $S(m, b)$, el primer paso consiste en tomar sus derivadas parciales con respecto a m y b .

$$\frac{\partial S}{\partial b} = \frac{1}{m^2 + 1} \sum_{i=1}^n (2)(mx_i - y_i + b)$$

$$= \frac{2}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b)$$

3.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

$$\frac{\partial S}{\partial m} = \frac{-2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 + \frac{1}{m^2 + 1} \sum_{i=1}^n (2x_i)(mx_i - y_i + b)$$

O de forma equivalente:

$$\frac{\partial S}{\partial m} = \frac{2}{m^2 + 1} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2$$

Para encontrar los valores de m y b que minimizan a los errores, las primeras derivadas parciales son igualadas a cero, resolviendo entonces el sistema de ecuaciones resultante. Estos puntos de la forma (m, b) , son los puntos críticos de $S(m, b)$.

$$\frac{\partial S}{\partial b} = \frac{2}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + b) = 0$$

$$\sum_{i=1}^n (mx_i - y_i + b) = 0 \quad (3.1)$$

$$m \sum_{i=1}^n x_i - \sum_{i=1}^n y_i + nb = 0$$

$$b = \frac{1}{n} \sum_{i=1}^n y_i - \frac{m}{n} \sum_{i=1}^n x_i$$

$$\therefore b = \bar{y} - m\bar{x} \quad (3.2)$$

$$\frac{\partial S}{\partial m} = \frac{2}{m^2 + 1} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{2m}{(m^2 + 1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 = 0$$

$$(m^2 + 1) \sum_{i=1}^n x_i (mx_i - y_i + b) - m \sum_{i=1}^n (mx_i - y_i + b)^2 = 0$$

$$(m^2 + 1) \sum_{i=1}^n x_i (mx_i - y_i + b) = m \sum_{i=1}^n (mx_i - y_i + b)^2 \quad (3.3)$$

$$(m^2 + 1) \sum_{i=1}^n x_i (mx_i - y_i + (\bar{y} - m\bar{x})) = m \sum_{i=1}^n (mx_i - y_i + (\bar{y} - m\bar{x}))^2$$

$$(m^2 + 1) \sum_{i=1}^n x_i (m(x_i - \bar{x}) - (y_i - \bar{y})) = m \sum_{i=1}^n (m(x_i - \bar{x}) - (y_i - \bar{y}))^2$$

PARTE III. ALTERNANCIA

Definiendo $X_i = x_i - \bar{x}$ y $Y_i = y_i - \bar{y}$, se tiene:

$$\begin{aligned}
 (m^2 + 1) \sum_{i=1}^n (X_i + \bar{x})(mX_i - Y_i) &= m \sum_{i=1}^n (mX_i - Y_i)^2 \\
 (m^2 + 1) \sum_{i=1}^n (mX_i^2 + m\bar{x}X_i - X_iY_i - \bar{x}Y_i) &= m \sum_{i=1}^n (m^2 X_i^2 - 2mX_iY_i + Y_i^2) \\
 m^3 \sum_{i=1}^n X_i^2 + m \sum_{i=1}^n X_i^2 + m^3 \bar{x} \sum_{i=1}^n X_i + m\bar{x} \sum_{i=1}^n X_i - m^2 \sum_{i=1}^n X_iY_i - \sum_{i=1}^n X_iY_i - m^2 \bar{x} \sum_{i=1}^n Y_i - \bar{x} \sum_{i=1}^n Y_i \\
 &= m^3 \sum_{i=1}^n X_i^2 - 2m^2 \sum_{i=1}^n X_iY_i + m \sum_{i=1}^n Y_i^2 \tag{3.4}
 \end{aligned}$$

La ecuación (3.4) puede ser simplificada mediante cancelación de términos y usando los siguientes resultados:

$$\begin{aligned}
 \sum_{i=1}^n X_i &= \sum_{i=1}^n (x_i - \bar{x}) & \sum_{i=1}^n Y_i &= \sum_{i=1}^n (y_i - \bar{y}) \\
 &= \sum_{i=1}^n x_i - n\bar{x} & &= \sum_{i=1}^n y_i - n\bar{y} \\
 &= \sum_{i=1}^n x_i - n \left(\frac{1}{n} \sum_{i=1}^n x_i \right) & &= \sum_{i=1}^n y_i - n \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \\
 &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i & &= \sum_{i=1}^n y_i - \sum_{i=1}^n y_i \\
 &= 0 & &= 0
 \end{aligned}$$

Con esta sustitución, la ecuación (3.4) puede ser resuelta para m usando la fórmula de la ecuación cuadrática, obteniendo entonces dos resultados para esta línea de regresión alterna:

$$\begin{aligned}
 m \sum_{i=1}^n X_i^2 - m^2 \sum_{i=1}^n X_iY_i - \sum_{i=1}^n X_iY_i &= -2m^2 \sum_{i=1}^n X_iY_i + m \sum_{i=1}^n Y_i^2 \\
 m^2 \left(\sum_{i=1}^n X_iY_i \right) + m \left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n Y_i^2 \right) + \left(-\sum_{i=1}^n X_iY_i \right) &= 0 \\
 m = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 \pm \sqrt{\left(\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 \right)^2 + 4 \left(\sum_{i=1}^n X_iY_i \right)^2}}{2 \sum_{i=1}^n X_iY_i} & \tag{3.5}
 \end{aligned}$$

3.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

Sea m_+ la solución con el radical positivo y m_- la solución con el negativo:

$$m_+ = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 + \sqrt{\left(\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2\right)^2 + 4\left(\sum_{i=1}^n X_i Y_i\right)^2}}{2\sum_{i=1}^n X_i Y_i}$$

$$m_- = \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2 - \sqrt{\left(\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n X_i^2\right)^2 + 4\left(\sum_{i=1}^n X_i Y_i\right)^2}}{2\sum_{i=1}^n X_i Y_i}$$

Las pendientes m_+ y m_- producen dos intersecciones con respecto al eje y , b_+ y b_- respectivamente.

Sobre la gráfica de $S(m,b)$ en \mathbb{R}^3 (Figura 3.2), los dos puntos mínimos son visibles con coordenadas $(m_+, b_+, S(m_+, b_+))$ y $(m_-, b_-, S(m_-, b_-))$. Estos puntos críticos son los valores que minimizan la suma de los cuadrados de los errores vista como función. Es posible construir gráficas de $S(m,b)$ una vez que se conocen los valores de x_i y y_i . La siguiente gráfica es generada con el software matemático *Maple* con base en los valores de x_i y y_i del **Ejemplo 1** desarrollado en el **Apéndice A** de la *Parte III*.

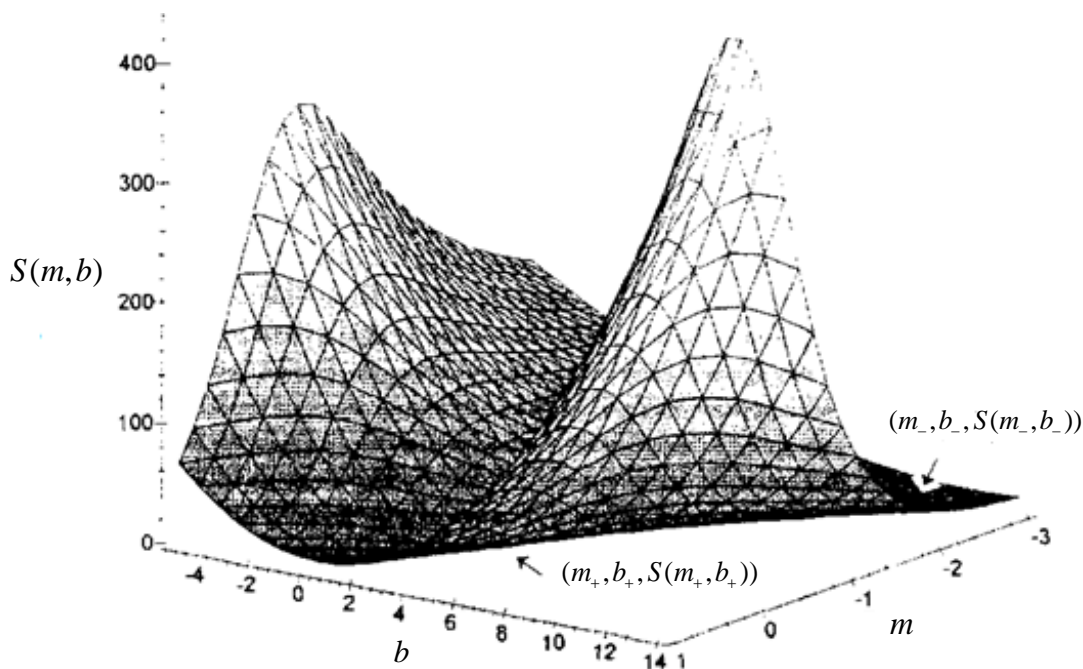


Figura 3.2. Gráfica de $S(m,b)$

Para encontrar teóricamente los únicos valores de m y b que minimizan la suma de los cuadrados de los errores, $S(m,b)$, la matriz de las segundas derivadas parciales -o matriz Hessiana- debe ser determinada.

$$D^2(S(m,b)) = \begin{bmatrix} \frac{\partial^2 S}{\partial b^2} & \frac{\partial^2 S}{\partial m \partial b} \\ \frac{\partial^2 S}{\partial b \partial m} & \frac{\partial^2 S}{\partial m^2} \end{bmatrix} = \begin{bmatrix} S_{bb} & S_{mb} \\ S_{bm} & S_{mm} \end{bmatrix}$$

Se deben encontrar los valores de m y b que hacen que esta matriz de 2×2 sea definida positiva, pero antes, las segundas derivadas parciales deben ser determinadas. Esto se puede hacer a partir de las primeras derivadas parciales de $S(m,b)$ que ya se han obtenido con anterioridad. Debido a que los valores de m y b son posibles extremos, se pueden usar las ecuaciones (3.1) y (3.3) que sirvieron para derivar los puntos críticos, y de esta manera, hacer simplificaciones posteriores en los cálculos:

$$\begin{aligned} S_{bb} &= \frac{\partial^2 S}{\partial b^2} = \frac{\partial}{\partial b} \left(\frac{2}{m^2+1} \sum_{i=1}^n (mx_i - y_i + b) \right) \\ &= \left(\frac{2}{m^2+1} \right) \left(\sum_{i=1}^n \frac{\partial}{\partial b} (mx_i - y_i + b) \right) \\ &= \frac{2}{m^2+1} \sum_{i=1}^n 1 \\ &= \frac{2n}{m^2+1} \end{aligned} \tag{3.6}$$

$$\begin{aligned} S_{mm} &= \frac{\partial^2 S}{\partial m^2} = \frac{\partial}{\partial m} \left(\frac{2}{m^2+1} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{2m}{(m^2+1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 \right) \\ &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{4m}{(m^2+1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) - \frac{4m}{(m^2+1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) \\ &\quad + \left(\frac{8m^2}{(m^2+1)^3} - \frac{2}{(m^2+1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \\ &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{8m}{(m^2+1)^2} \sum_{i=1}^n x_i (mx_i - y_i + b) + \left(\frac{8m^2}{(m^2+1)^3} - \frac{2}{(m^2+1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \\ &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{8m^2}{(m^2+1)^3} \sum_{i=1}^n (mx_i - y_i + b)^2 + \left(\frac{8m^2}{(m^2+1)^3} - \frac{2}{(m^2+1)^2} \right) \sum_{i=1}^n (mx_i - y_i + b)^2 \end{aligned}$$

este resultado por (3.3)

3.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

$$\begin{aligned}
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{(m^2+1)^2} \sum_{i=1}^n (mx_i - y_i + b)^2 \\
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \left(\frac{2}{(m^2+1)^2} \right) \left(\frac{(m^2+1)}{m} \right) \sum_{i=1}^n x_i (mx_i - y_i + b) \\
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m(m^2+1)} \left(m \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i \right) \\
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m^2+1} \sum_{i=1}^n x_i^2 - \frac{2}{m(m^2+1)} \left(- \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i \right) \\
 &= \frac{2}{m(m^2+1)} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) \tag{3.7}
 \end{aligned}$$

Las segundas derivadas parciales con respecto a m y a b son iguales, $S_{mb} = S_{bm}$, entonces sólo una de las dos debe ser encontrada. La primera derivada parcial de $S(m,b)$ con respecto a b es más simple, por lo que se procede a encontrar S_{mb} .

$$\begin{aligned}
 S_{mb} &= \frac{\partial^2 S}{\partial m \partial b} = \frac{\partial}{\partial m} \left(\frac{2}{m^2+1} \sum_{i=1}^n (mx_i - y_i + b) \right) \\
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i - \frac{4m}{(m^2+1)^2} \sum_{i=1}^n (mx_i - y_i + b) \\
 &= \frac{2}{m^2+1} \sum_{i=1}^n x_i
 \end{aligned}$$

este resultado por (3.1) (3.8)

Estos valores conforman la matriz de las segundas derivadas parciales de $S(m,b)$:

$$D^2(S(m,b)) = \frac{2}{m^2+1} \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \frac{1}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) \end{bmatrix} \tag{3.9}$$

Para los valores de m y b que minimizan $S(m,b)$, la matriz debe ser definida positiva en estos valores. El determinante de la submatriz superior izquierda es siempre positivo, porque n , el número de puntos, es siempre positivo. La segunda condición requiere que el determinante de la matriz de las segundas derivadas parciales (3.9) sea positivo, es decir $S_{bb}S_{mm} - S_{mb}^2 > 0$.

Graficando la pendiente o la intercepción contra el determinante de la matriz de las segundas derivadas parciales, es fácil observar que sólo un intervalo acotado para m o

para b producirá un determinante positivo. Como en la gráfica anterior, es posible graficar sólo cuando los valores de x y y son sustituidos en las soluciones. Ésta en particular es del **Ejemplo 1** presentado en el **Apéndice A** de la *Parte III*.

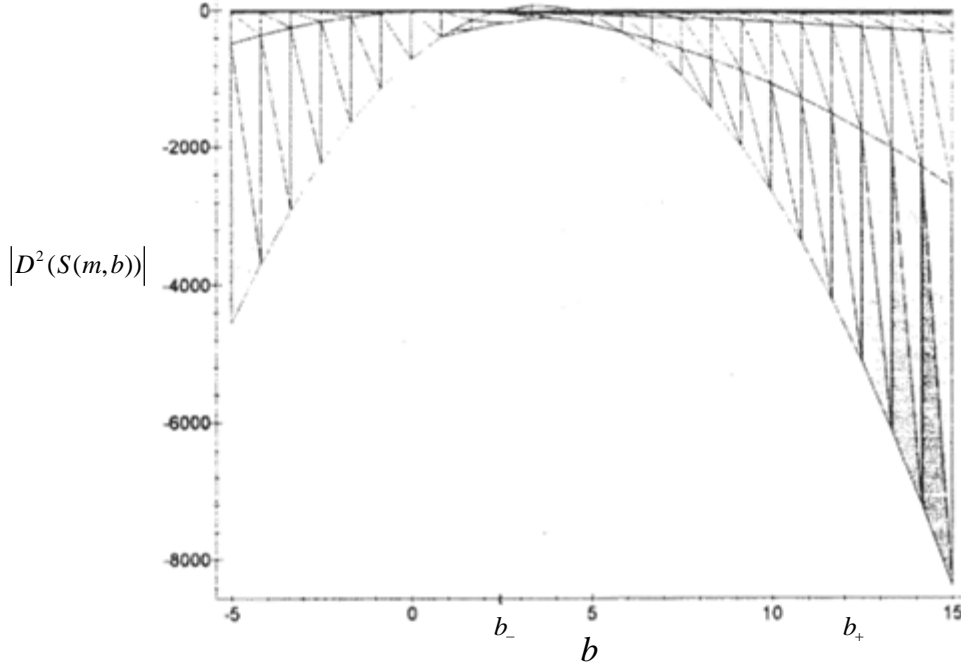


Figura 3.3. Gráfica de $|D^2(S(m,b))|$

En el intento por determinar analíticamente cuál de los dos valores para m y para b debería ser usado, minimizando de esta forma la suma de los cuadrados de los errores, la desigualdad del determinante de la matriz de las segundas derivadas parciales debe ser simplificada.

$$\frac{2}{m^2 + 1} \left(\frac{n}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) - \left(\sum_{i=1}^n x_i \right)^2 \right) > 0$$

$$\Leftrightarrow \frac{n}{m} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) > \left(\sum_{i=1}^n x_i \right)^2$$

$$\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i - \frac{bn}{m} \sum_{i=1}^n x_i > \left(\sum_{i=1}^n x_i \right)^2$$

$$\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{bn}{m} \sum_{i=1}^n x_i + \left(\sum_{i=1}^n x_i \right)^2$$

3.1. ESTIMACIÓN DE LOS PARÁMETROS POR MÍNIMOS CUADRADOS

$$\begin{aligned}
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > \left(\sum_{i=1}^n x_i \right) \left(\frac{bn}{m} + \sum_{i=1}^n x_i \right) \\
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > n \left(\sum_{i=1}^n x_i \right) \left(\frac{b}{m} + \frac{1}{n} \sum_{i=1}^n x_i \right) \\
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > n \left(\sum_{i=1}^n x_i \right) \left(\frac{b}{m} + \bar{x} \right) \\
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{n}{m} \left(\sum_{i=1}^n x_i \right) (b + m\bar{x}) \\
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{n}{m} \left(\sum_{i=1}^n x_i \right) (\bar{y}) \\
 &\Leftrightarrow \frac{n}{m} \sum_{i=1}^n x_i y_i > \frac{1}{m} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \tag{3.10}
 \end{aligned}$$

No pueden hacerse simplificaciones posteriores de la ecuación (3.10) sin antes conocer la naturaleza de m . Se demostrará que las pendientes m_- y m_+ son inversas negativas una de la otra, por lo que las rectas producidas serán siempre perpendiculares. Se usarán a las variables a , b , y c para simplificar los cálculos.

$$\begin{aligned}
 a &= \sum_{i=1}^n X_i^2, \quad b = \sum_{i=1}^n Y_i^2, \quad y \quad c = \sum_{i=1}^n X_i Y_i \\
 m_+ m_- &= \left(\frac{b-a+\sqrt{(a-b)^2+4c^2}}{2c} \right) \left(\frac{b-a-\sqrt{(a-b)^2+4c^2}}{2c} \right) \\
 &= \frac{(b-a)^2 - ((a-b)^2 + 4c^2)}{4c^2} \\
 &= \frac{(b-a)^2 - (a-b)^2 - 4c^2}{4c^2} \\
 \therefore m_+ m_- &= -1
 \end{aligned}$$

Debido a que el producto de las dos pendientes con valores reales es siempre negativo, una debe ser positiva y la otra negativa. Esta información es utilizada a continuación para simplificar la ecuación (3.10).

$$\text{Si la pendiente es positiva: } n \sum_{i=1}^n x_i y_i > \sum_{i=1}^n x_i \sum_{i=1}^n y_i \quad (3.11)$$

$$\text{Si la pendiente es negativa: } n \sum_{i=1}^n x_i y_i < \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

Todos los pasos en la simplificación (3.10) son reversibles, por lo que la siguiente declaración es verdadera:

$$\text{Si } n \sum_{i=1}^n x_i y_i > \sum_{i=1}^n x_i \sum_{i=1}^n y_i, \text{ entonces la pendiente positiva minimiza } S(m,b). \quad (3.12)$$

$$\text{Si } n \sum_{i=1}^n x_i y_i < \sum_{i=1}^n x_i \sum_{i=1}^n y_i, \text{ entonces la pendiente negativa minimiza } S(m,b).$$

Ya que la submatriz superior izquierda de la matriz para la segunda derivada de $S(m,b)$ es siempre positiva, la matriz de las segundas derivadas parciales nunca será definida negativa en los puntos críticos, y por lo tanto, no hay valores críticos de m y b que puedan maximizar $S(m,b)$. Lógicamente esto es cierto debido a que una línea puede ser extendida infinitamente lejos de los puntos, creando una suma de cuadrados de errores tan grande como se quiera sin ninguna cota superior. El segundo punto crítico (m,b) de la función $S(m,b)$, el cual hace que el determinante de $D^2(S(m,b))$ sea negativo, será un punto silla en la gráfica de $S(m,b)$, y da una suma de cuadrados de errores más grande que el mínimo.

Este método alternativo encontrará la pendiente m y la intercepción b que minimizan los errores de acuerdo a la definición del error visto como una distancia perpendicular. Usando las ecuaciones (3.2) y (3.5), y la condición (3.12), la línea de regresión que minimiza la suma de los cuadrados de los errores puede ser encontrada con este método alternativo.

Trabajando “a mano” con el método alternativo de regresión por mínimos cuadrados puede resultar tedioso. Sin embargo, con el uso del software adecuado, este método no es más difícil que el método tradicional. Dos ejemplos numéricos en *Maple* son elaborados en el **Apéndice A** de este capítulo, así como también el código que sirve de modelo.

3.2. Ventajas y desventajas del método alternativo

El método alternativo de regresión lineal presentado en la sección anterior tiene ventajas como desventajas cuando se le compara con el tradicional. Como se aprecia en el **Apéndice A** de la *Parte III*, los dos métodos producen resultados similares, aunque ambas rectas son visiblemente diferentes. Diversas cuestiones deberían ser consideradas primero antes de comparar la precisión entre los dos métodos de regresión.

A diferencia del método tradicional, el alternativo no puede producir líneas horizontales -rectas de la forma $y = b$ que tienen pendiente cero-. Ninguno de los dos puede arrojar líneas verticales -de la forma $x = c$ -. Cuando los puntos se encuentran exactamente sobre una recta, ambos métodos de regresión producen la misma línea -la recta sobre la cual se encuentran todos los puntos- por lo que las dos sumas de cuadrados de los errores son iguales a cero.

El método tradicional por mínimos cuadrados es más sencillo; la expresión para m es mucho más corta que en el alternativo permitiendo su fácil memorización. Es mejor usar el método tradicional si se calcula la línea de regresión a mano. Sin embargo, con el uso de software, el método alternativo no es más complicado.

En el método tradicional, el error está presente sólo en la variable y -la variable dependiente-. Esta suposición es importante en muchas aplicaciones prácticas en donde es útil considerar a una variable sin defectos o errores, y tomar al error como una variable dependiente. En el método alternativo de estimación de los parámetros del modelo de regresión lineal por mínimos cuadrados, ya no se supone que la variable x se encuentra libre de error. El método alternativo permite la presencia del error en las dos variables midiendo la distancia de un punto a la línea de regresión de manera perpendicular; esta distancia puede variar de ser casi vertical, como en el método tradicional, a ser casi horizontal, dependiendo del arreglo del conjunto de puntos.

La suma de los cuadrados de los errores en la línea alternativa es siempre menor o igual a la misma suma por el tradicional. Este hecho hace que la regresión por el método alternativo parezca superior al tradicional. De manera frecuente, la línea de regresión por mínimos cuadrados es conocida como “el mejor ajuste”, ya que ésta tiene la suma de cuadrados de errores mínima. Comparando la suma de los cuadrados de errores por cada método, sólo resta determinar cuál de los dos es más preciso, siendo fácil suponer que el método alternativo es “mejor” debido a que tiene la suma de cuadrados de errores más pequeña. Dependiendo de cuáles cantidades se consideren ser más importantes, el método alternativo de regresión debería ser la mejor elección. Sin embargo, aún no se ha

demostrado que el método alternativo de regresión por mínimos cuadrados es necesariamente más preciso.

3.3. Comparación de precisión

A pesar de que el método alternativo de estimación de los parámetros del modelo de regresión lineal por mínimos cuadrados da una suma de cuadrados de errores, menor o igual a la del tradicional, estos dos valores no son realmente comparables. Debido a que la distancia de un punto a la línea de regresión -o error- es definido de forma diferente en cada método, cada valor de la suma de cuadrados de los errores describe algo diferente. Comparando éstas para la línea alternativa y la tradicional, es como si se compararan peras con manzanas. Para encontrar cuál de las dos líneas de regresión es más precisa, debe ser utilizado un método diferente.

Un método para encontrar el área entre dos rectas es estudiado en el *cálculo integral*. Para encontrar el área entre dos funciones, $f(x)$ y $g(x)$, en los valores de x entre a y b , la integral de a a b es tomada de la diferencia de $f(x)$ y $g(x)$:

$$\int_b^a (f(x) - g(x)) dx.$$

Se deben tomar precauciones con esta fórmula, debido a que se podría obtener un área negativa si $f(x)$ no es siempre mayor a $g(x)$ en el intervalo de a a b . Una alternativa para garantizar un área positiva es utilizar el valor absoluto de la función que se está integrando:

$$\int_b^a |f(x) - g(x)| dx. \quad (3.13)$$

Otra alternativa es tomar el cuadrado de la diferencia:

$$\int_b^a (f(x) - g(x))^2 dx. \quad (3.14)$$

Evidentemente, esta alternativa no proporciona el área real entre dos funciones, pero puede ser usada para realizar comparaciones.

El método utilizado para comparar la precisión entre el método de regresión tradicional y el alternativo, equivale a elegir de manera aleatoria una recta, a la cual se referirá como la *línea real*, graficando puntos aleatorios alrededor de ésta. La pendiente de la

3.3. COMPARACIÓN DE PRECISIÓN

línea real permite el error en ambas variables. Estos puntos son usados para calcular ambas líneas de regresión: la tradicional y la alternativa. Utilizando el método de la integral, el área entre la línea tradicional y la línea real, así como el área entre la línea alternativa y la línea real, son encontradas y comparadas entre sí. La línea de regresión con la menor área entre las diferencias será considerada como la más precisa.

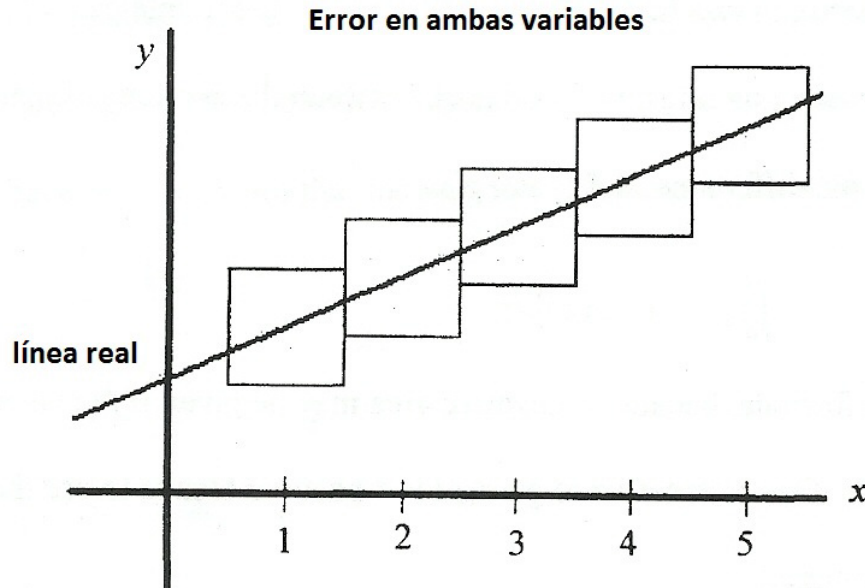


Figura 3.4. Error en ambas variables

Un programa elaborado en *Maple* utiliza distribuciones uniformes para producir la pendiente y la ordenada al origen de la línea real en forma aleatoria, así como los errores en las dos variables. Esta serie de comandos es presentada en el **Apéndice B** de la *Parte III*. También en este apéndice se incluye una versión de iteración múltiple del mismo programa, la cual permite su ejecución 100 veces mientras registra cuál de los dos métodos produce una línea más precisa en cada iteración. Después de realizar 20 iteraciones de éste último, el método alternativo de regresión lineal por mínimos cuadrados muestra ser más preciso en un 51.95% de acuerdo al criterio (3.13). El programa puede ser fácilmente modificado para usar el criterio (3.14).

Un experimento similar fue llevado a cabo para puntos con errores aleatorios presentes sólo en la variable y , siendo en ese caso más preciso el método tradicional en un 52.35%. El método tradicional de regresión por mínimos cuadrados supone que se tienen datos de este tipo.

PARTE III. ALTERNANCIA

El programa realizado en *Maple* para la comparación de precisión, con datos que cuentan con errores sólo en la variable y , es mostrado también en el **Apéndice B**. Una versión con múltiples iteraciones es incluida también. De nuevo, la pendiente para la línea real puede variar de $\frac{1}{10}$ a 3; y la intercepción b en un intervalo de 0 a $3\frac{1}{3}$. Los puntos son graficados de forma aleatoria alrededor de la línea, pero uniformemente distribuidos a lo largo de un intervalo de una unidad y centrados alrededor del valor de y en la línea real. De forma sorprendente, después de cientos de iteraciones, ambos métodos no muestran tener alguna diferencia significativa.

PARTE IV EXTENSIÓN

4. Regresión lineal múltiple

Un modelo de regresión que involucra más de una variable regresora es llamado un *modelo de regresión múltiple*. Su ajuste y discusión son hechos en este capítulo. Como su nombre lo indica, esta sección es una extensión de lo estudiado en la *Parte II* de este trabajo.

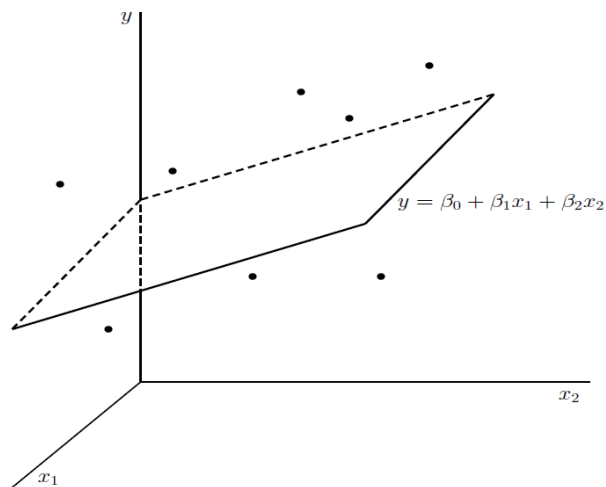


Figura 4.1. Modelo de regresión lineal múltiple.

4.1. Modelos de regresión múltiple

Suponiendo que el rendimiento en libras de un proceso químico depende de dos variables: la temperatura y la concentración de catalizador, el modelo de regresión múltiple que debería describir esta relación es

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \varepsilon \quad (4.1)$$

donde y denota el rendimiento, x_1 denota la temperatura, y x_2 denota la concentración de catalizador. Este es un *modelo de regresión lineal múltiple* con dos variables regresoras.

PARTE IV. EXTENSIÓN

El término *lineal* es utilizado debido a que la ecuación (4.1) es una función lineal de los parámetros desconocidos β_0 , β_1 , y β_2 .

El modelo de regresión (4.1) describe un plano en el espacio tridimensional de y , x_1 y x_2 . La Figura 4.2a muestra este plano para el modelo

$$E(y) = 50 + 10x_1 + 7x_2$$

donde se ha supuesto que el valor esperado del error x en (4.1) es cero. El parámetro β_0 es la intercepción del plano de la regresión. Si el rango de datos incluye $x_1 = x_2 = 0$, entonces β_0 es la media de y en este caso. De otra forma, no tiene ninguna interpretación física. El parámetro β_1 es el cambio esperado en la respuesta (y) por unidad de cambio en x_1 cuando x_2 se mantiene constante. La Figura 4.2b muestra la *gráfica de contorno* del modelo de regresión, esto es, líneas constantes de respuesta esperada $E(y)$ como función de x_1 y x_2 . Nótese que las líneas de contorno en esta gráfica son líneas rectas paralelas.

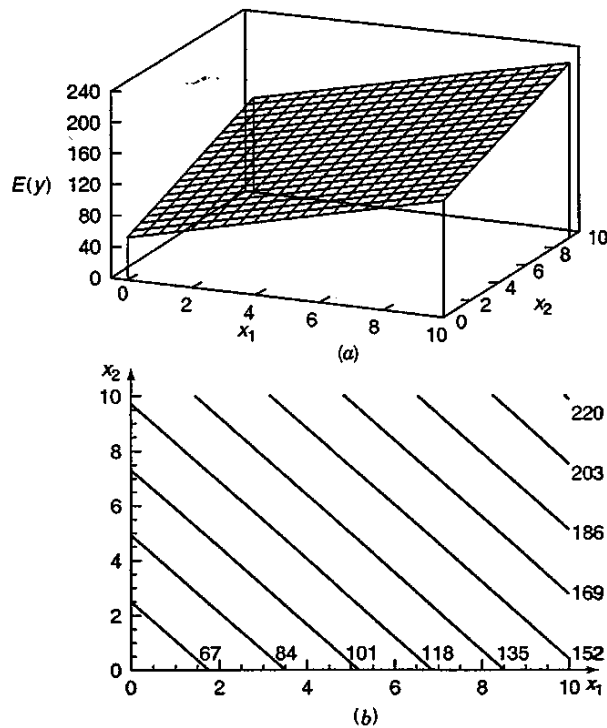


Figura 4.2. (a) El plano de regresión para el modelo $E(y) = 50 + 10x_1 + 7x_2$. (b) La gráfica de contorno.

4.1. MODELOS DE REGRESIÓN MÚLTIPLE

En general, la respuesta y puede estar relacionada a las k variables *regresoras* o *predictoras*. El modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (4.2)$$

es llamado *modelo de regresión lineal múltiple* con k regresoras. Los parámetros β_j , $j = 0, 1, \dots, k$, son llamados *coeficientes de regresión*. Este modelo describe un hiperplano en el espacio k -dimensional de las variables x_j . El parámetro β_j representa el cambio esperado en la respuesta y por unidad de cambio en x_j cuando todas las demás regresoras $x_i (i \neq j)$ permanecen constantes. Por esta razón, los parámetros β_j , $j = 0, 1, \dots, k$, son llamados de forma frecuente *coeficientes de regresión parcial*.

Los modelos de regresión lineal múltiple son frecuentemente utilizados como modelos empíricos o funciones de aproximación. Esto es, la verdadera relación funcional entre y y x_1, x_2, \dots, x_k es desconocida, pero bajo ciertos rangos de las variables regresoras, el modelo de regresión lineal es una adecuada aproximación a la función real desconocida.

Modelos que son más complejos en estructura que la ecuación (4.2), pueden ser analizados de la misma forma mediante técnicas de regresión lineal múltiple. Por ejemplo, considérese el modelo de un polinomio cúbico

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \quad (4.3)$$

Haciendo $x_1 = x$, $x_2 = x^2$ y $x_3 = x^3$, entonces la ecuación (4.3) puede ser escrita como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (4.4)$$

el cual es un modelo de regresión múltiple con tres variables regresoras. Un modelo polinomial como éste, es implementado en el siguiente capítulo.

Los modelos que incluyen efectos de interacción también pueden ser analizados por un método de regresión lineal múltiple. Por ejemplo, supóngase que el modelo es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \quad (4.5)$$

Haciendo $x_3 = x_1 x_2$ y $\beta_3 = \beta_{12}$, la ecuación (4.5) puede ser escrita como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (4.6)$$

el cual es un modelo de regresión lineal.

PARTE IV. EXTENSIÓN

La Figura 4.3a muestra la gráfica tridimensional del modelo

$$y = 50 + 10x_1 + 7x_2 + 5x_1x_2,$$

y la Figura 4.3b su correspondiente gráfica de contorno en dos dimensiones. A pesar de que este modelo corresponde a uno de regresión lineal, la forma de la superficie generada es no lineal. En general, cualquier modelo de regresión lineal en los parámetros -los β_j -, es un modelo de regresión lineal, sin tener en cuenta la forma de la superficie generada.

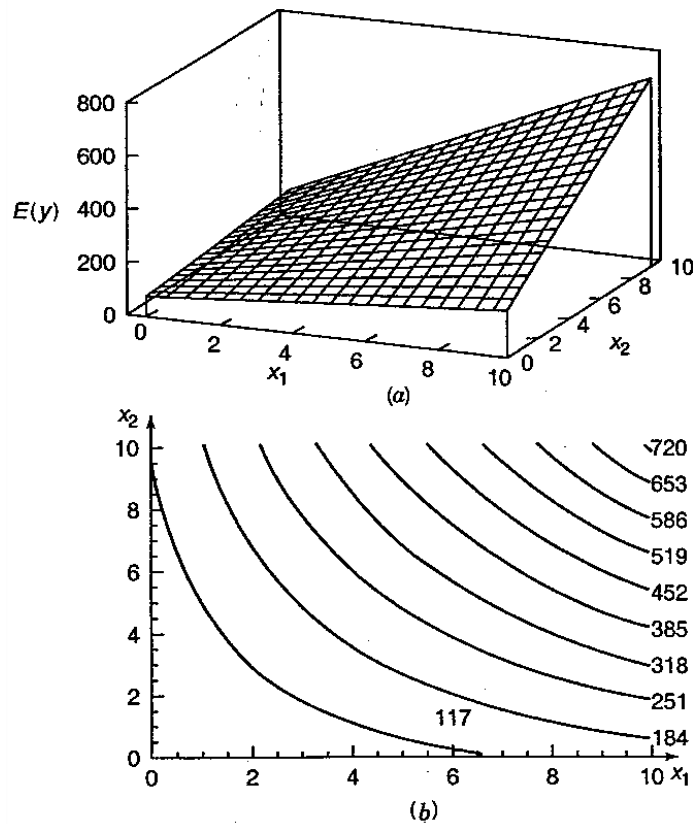


Figura 4.3. (a) Gráfica tridimensional para el modelo de regresión $E(y) = 50 + 10x_1 + 7x_2 + 5x_1x_2$. (b) La gráfica de contorno (o curvas de nivel).

La figura anterior otorga una buena interpretación gráfica de una interacción. Generalmente, la interacción establece que el efecto producido al cambiar una variable - x_1 por ejemplo-, depende del nivel de la otra variable (x_2). La Figura 4.3 muestra que cambiando x_1 de 2 a 8 produce un cambio mucho más pequeño en $E(y)$ cuando $x_2 = 2$ que cuando $x_2 = 10$. Los efectos de interacción ocurren frecuentemente en el estudio y

4.1. MODELOS DE REGRESIÓN MÚLTIPLE

análisis de sistemas del mundo real, siendo los métodos de regresión una de las técnicas que pueden ser usadas para describir dichos sistemas.

Como un ejemplo final, considérese el modelo de segundo orden con interacción

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad (4.7)$$

Haciendo $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$ y $\beta_5 = \beta_{12}$, entonces la ecuación anterior puede ser escrita como un modelo de regresión lineal múltiple como sigue:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

La Figura 4.4 muestra la gráfica tridimensional y de contorno (o curvas de nivel) para

$$E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$$

Estos puntos indican que el cambio esperado en y cuando x_1 es cambiado en una unidad -por ejemplo-, es una función de ambas variables x_1 y x_2 . Los términos cuadráticos y cruzados en este modelo, producen una función en forma de montículo. Dependiendo de los valores de los coeficientes de la regresión, el modelo de segundo orden con interacción es capaz de asumir una amplia variedad de formas; debido a esto, es un modelo de regresión muy flexible.

PARTE IV. EXTENSIÓN

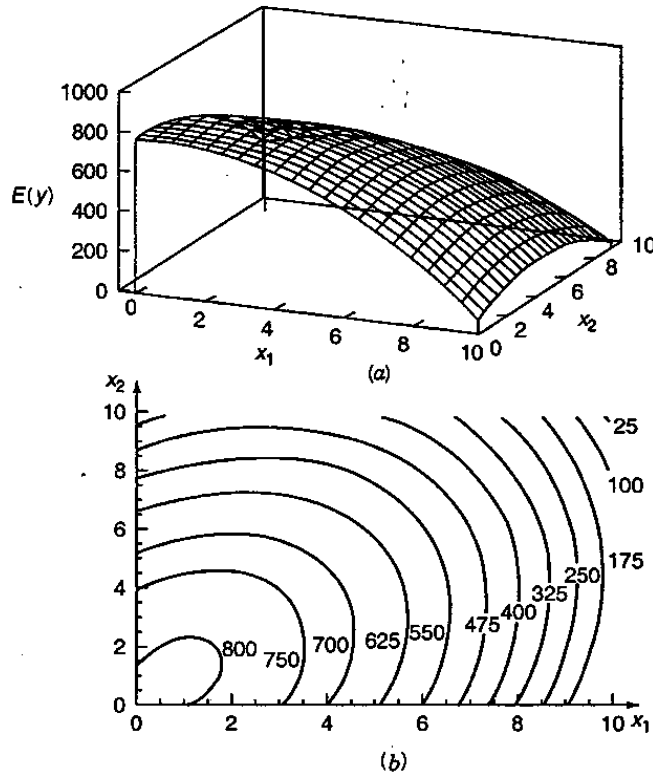


Figura 4.4. (a) Gráfica tridimensional para el modelo de regresión $E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$. (b) La gráfica de contorno (o curvas de nivel).

En la mayoría de problemas del mundo real, los valores de los parámetros -los coeficientes de la regresión- y la varianza del error σ^2 no son conocidos, debiendo ser estimados con los datos de la muestra. El modelo ajustado de la regresión es típicamente utilizado en la predicción de futuras observaciones de la variable de respuesta y , o para estimar la respuesta media en niveles particulares de las x .

4.2. Estimación de los parámetros del modelo

4.2.1. Estimación de los coeficientes de regresión mediante mínimos cuadrados

El método de mínimos cuadrados también puede ser usado para estimar los coeficientes de la ecuación (4.2). Suponiendo que $n > k$ observaciones están disponibles, sea y_i la

4.2. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

i -ésima respuesta observada y x_{ij} la i -ésima observación o nivel de la regresora x_j . Se supone que el término ε en el modelo tiene $E(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2$, además de que estos errores no están correlacionados.

A lo largo de este capítulo, se considera que las regresoras x_1, x_2, \dots, x_k son variables fijas -es decir, no aleatorias- medidas sin error. Sin embargo, todos los resultados son válidos para el caso en donde también son aleatorias. Esto es importante, debido a que cuando los datos de la regresión surgen de un estudio observacional, algunos o la mayoría de las regresoras serán variables aleatorias. Cuando éstos provienen de un experimento diseñado, es más conveniente que las variables sean fijas. Cuando las x son variables aleatorias, únicamente es necesario que las observaciones de cada regresora sean independientes, y que su distribución no dependa de los coeficientes de la regresión -los β_j - o de σ^2 . Cuando se realicen las pruebas de hipótesis o se estudien los intervalos de confianza, se supondrá que la distribución condicional de y dadas x_1, x_2, \dots, x_k es normal con media $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ y varianza σ^2 .

Se puede escribir el modelo de regresión correspondiente a la ecuación (4.2) como

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \end{aligned} \quad (4.8)$$

Esta vez, la función de mínimos cuadrados está definida como

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_k) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \end{aligned} \quad (4.9)$$

La función $S(\beta_0, \beta_1, \dots, \beta_k)$ debe ser minimizada con respecto a $\beta_0, \beta_1, \dots, \beta_k$. Los estimadores por mínimos cuadrados de $\beta_0, \beta_1, \dots, \beta_k$ deben satisfacer

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) = 0 \quad (4.10a)$$

$$\left. \frac{\partial S}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} = 0, \quad j = 1, 2, \dots, k \quad (4.10b)$$

4.2. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

La suma anterior puede ser expresada como

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

debido a que $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ es una matriz de 1×1 -o un escalar- y su transpuesta $(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ es el mismo valor. Los estimadores por mínimos cuadrados tienen que satisfacer

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

Lo cual se simplifica como

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (4.12)$$

Las ecuaciones (4.12) son las ecuaciones normales por mínimos cuadrados. Este modelo es la representación análoga de la forma escalar en (4.11).

Para resolver las ecuaciones normales, se multiplican los dos lados de (4.12) por la inversa de $\mathbf{X}'\mathbf{X}$. Entonces, el estimador por mínimos cuadrados de $\boldsymbol{\beta}$ es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (4.13)$$

siempre que la matriz inversa $(\mathbf{X}'\mathbf{X})^{-1}$ exista. Ésta existirá siempre que las regresoras sean *linealmente independientes*, esto es, si ninguna columna de la matriz \mathbf{X} es una combinación lineal de las otras columnas.

Es fácil ver que la forma matricial de las ecuaciones normales (4.12) es idéntica a la forma escalar (4.11). Escribiendo (4.12) en detalle, se obtiene:

PARTE IV. EXTENSIÓN

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix}$$

Si se realiza la multiplicación, se obtiene la forma escalar (4.11). En este arreglo, se aprecia que $\mathbf{X}'\mathbf{X}$ es una matriz simétrica de $p \times p$, y $\mathbf{X}'\mathbf{y}$ es un vector columna de $p \times 1$. La matriz $\mathbf{X}'\mathbf{X}$ tiene una estructura especial. Sus elementos diagonales son las sumas de los cuadrados de las columnas de \mathbf{X} , y los elementos fuera de la diagonal son las sumas de los productos cruzados de los elementos de las columnas de \mathbf{X} . Además, los elementos de $\mathbf{X}'\mathbf{y}$ son las sumas de los productos cruzados de las columnas de \mathbf{X} y las observaciones y_i .

El modelo de regresión ajustado correspondiente a los niveles -o renglones- de las variables regresoras $\mathbf{x}' = [1, x_1, x_2, \dots, x_k]$ es

$$\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_j$$

El vector de valores ajustados \hat{y}_i correspondientes a los valores observados y_i es

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \quad (4.14)$$

La matriz $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ de $n \times n$, es frecuentemente llamada la *matriz sombrero*. Ésta relaciona el vector de los valores observados en un vector de valores ajustados. La matriz sombrero y sus propiedades, tienen un papel central en el análisis de regresión.

La diferencia entre el valor observado y_i y su correspondiente valor ajustado \hat{y}_i , es el residual $e_i = y_i - \hat{y}_i$. Los n residuales pueden ser escritos de forma conveniente en notación matricial como

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (4.15a)$$

Existen diversas formas de expresar el vector de residuales \mathbf{e} que demuestran ser útiles, tales como

4.2. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (4.15b)$$

4.2.2. Una interpretación geométrica de los estimadores por mínimos cuadrados en el modelo de regresión lineal múltiple

Una interpretación geométrica intuitiva de los estimadores por mínimos cuadrados es a veces útil. Considerando el vector de observaciones $\mathbf{y}' = [y_1, y_2, \dots, y_n]$ definido desde el origen al punto A , y_1, y_2, \dots, y_n forman las coordenadas de un espacio muestral n -dimensional.

El siguiente espacio muestral es tridimensional.

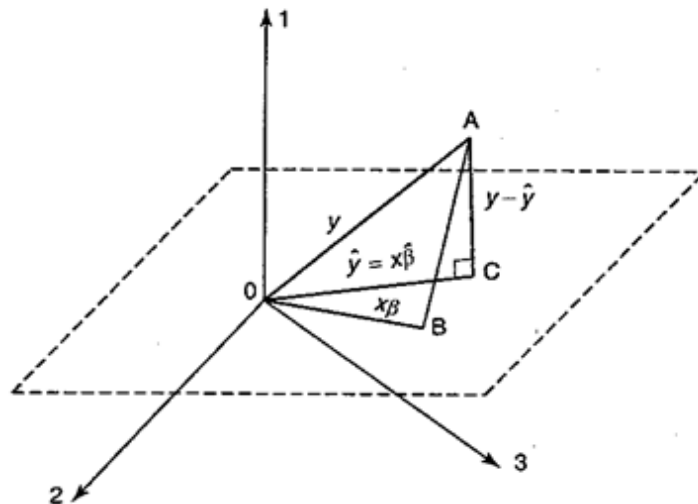


Figura 4.5. Una interpretación geométrica de los estimadores por mínimos cuadrados.

La matriz \mathbf{X} consiste de p vectores columna de $n \times 1$, por ejemplo: $\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ (el primero es un vector de unos). Cada una de estas columnas define un vector desde el origen en el espacio muestral. Se puede representar cada punto en este subespacio mediante una combinación lineal de $\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. Estos p vectores forman un subespacio p -dimensional llamado el *espacio de estimación* mostrado en la Figura 4.5 para $p = 2$. Entonces, cualquier punto en el espacio de estimación es de la forma $\mathbf{X}\boldsymbol{\beta}$. El vector $\mathbf{X}\boldsymbol{\beta}$ determina el punto B en la Figura 4.5. La distancia al cuadrado de A a B es justo

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Por lo tanto, minimizar la distancia cuadrada del punto A definido por el vector de observaciones \mathbf{y} al espacio de estimación, requiere encontrar el punto en el espacio de estimación más cercano a A . La distancia cuadrada será un mínimo cuando éste se encuentre en la línea normal -o perpendicular- de A . Este es el punto C en la Figura 4.5 definido por el vector $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Debido a que $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ es perpendicular al espacio de estimación, se puede escribir

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \quad \text{ó} \quad \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

que se reconocen como las ecuaciones normales por mínimos cuadrados.

4.2.3. Propiedades de los estimadores por mínimos cuadrados

Las propiedades estadísticas de los estimadores $\hat{\boldsymbol{\beta}}$ por mínimos cuadrados, pueden ser fácilmente demostrables. Se tiene primero:

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] = \boldsymbol{\beta} \end{aligned}$$

ya que $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ y $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$. Por lo tanto, $\hat{\boldsymbol{\beta}}$ es un *estimador insesgado* de $\boldsymbol{\beta}$.

El cálculo de la varianza de $\hat{\boldsymbol{\beta}}$ es expresado mediante la *matriz de covarianzas*:

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

la cual es una matriz simétrica de $p \times p$, cuyo j -ésimo elemento de la diagonal es la varianza de $\hat{\boldsymbol{\beta}}_j$, y cuyo ij -ésimo elemento fuera de la diagonal es la covarianza entre $\hat{\boldsymbol{\beta}}_i$ y $\hat{\boldsymbol{\beta}}_j$. La matriz de covarianzas de $\hat{\boldsymbol{\beta}}$ es

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

Por lo tanto, escribiendo $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$; la varianza de $\hat{\boldsymbol{\beta}}_j$ es $\sigma^2 C_{jj}$, y la covarianza entre $\hat{\boldsymbol{\beta}}_i$ y $\hat{\boldsymbol{\beta}}_j$ es $\sigma^2 C_{ij}$. El teorema de Gauss-Markov, revisado en el **Apéndice D** de este capítulo, establece que $\hat{\boldsymbol{\beta}}$ es el mejor estimador insesgado por mínimos cuadrados de $\boldsymbol{\beta}$. Si después se supone que los errores ε_i son normalmente distribuidos, entonces se verá (en la

4.2. ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

sección 4.2.5) que $\hat{\beta}$ es también el estimador de β por máxima verosimilitud. Éste es un estimador insesgado de varianza mínima de β .

4.2.4. Estimación de σ^2

Como en la regresión lineal simple, es posible desarrollar un estimador de σ^2 a partir de la suma de cuadrados de residuales denotada por SC_{Res} :

$$\begin{aligned} SC_{\text{Res}} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n e_i^2 \\ &= \mathbf{e}'\mathbf{e} \end{aligned}$$

Sustituyendo $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$, se tiene

$$\begin{aligned} SC_{\text{Res}} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned}$$

Ya que $\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$, la última ecuación queda

$$SC_{\text{Res}} = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \quad (4.16)$$

La suma de residuales tiene asociados $n - p$ grados de libertad, debido a que se estiman p parámetros en el modelo de regresión. El *cuadrado medio de los residuales* es

$$CM_{\text{Res}} = \frac{SC_{\text{Res}}}{n - p} \quad (4.17)$$

Por el **Apéndice C** de este capítulo, el valor esperado de CM_{Res} es σ^2 , por lo que un estimador insesgado de σ^2 está dado por

$$\hat{\sigma}^2 = CM_{\text{Res}} \quad (4.18)$$

4.2.5. Estimación por máxima verosimilitud

Como en el caso de la regresión lineal simple, se puede demostrar que los estimadores por máxima verosimilitud para los parámetros del modelo en la regresión lineal múltiple, cuando los errores son normalmente distribuidos e independientes, son los mismos por mínimos cuadrados. El modelo es

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

y los errores son normalmente distribuidos e independientes, con una varianza constante σ^2 , o equivalentemente, $\boldsymbol{\varepsilon}$ se distribuye como $N(\mathbf{0}, \sigma^2 \mathbf{I})$. La función de densidad normal para los errores es

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2}$$

La función de verosimilitud es la densidad conjunta de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, ó $\prod_{i=1}^n f(\varepsilon_i)$. Por lo tanto se tiene

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}$$

Ya que se puede escribir $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, esta función se convierte en

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}$$

Como en el caso de la regresión lineal simple, es más conveniente tratar con el logaritmo de la función anterior:

$$\ln L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

No es difícil percatarse que para un valor fijo de σ , la función anterior es maximizada cuando el término

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

es minimizado. Por lo tanto, el estimador por máxima verosimilitud de $\boldsymbol{\beta}$ cuando los errores son normalmente distribuidos, es igual al encontrado por mínimos cuadrados $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$. El estimador por máxima verosimilitud de σ^2 es

4.3. PRUEBAS DE HIPÓTESIS EN LA REGRESIÓN LINEAL MÚLTIPLE

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$

4.3. Pruebas de hipótesis en la regresión lineal múltiple

Una vez que se han estimado los parámetros en el modelo, se establecen dos preguntas inmediatas:

1. ¿Qué tan adecuado es el modelo?
2. ¿Qué regresoras específicas parecen importantes?

Diversas pruebas de hipótesis demuestran ser buenas para contestar estas preguntas. Las pruebas formales requieren que los errores sean independientes e idénticamente distribuidos, siguiendo una distribución normal con media $E(\varepsilon_i) = 0$ y $\text{Var}(\varepsilon_i) = \sigma^2$.

4.3.1. Prueba de significancia

La *prueba de significancia de la regresión* es utilizada para determinar si existe una *relación lineal* entre la respuesta y y las variables regresoras x_1, x_2, \dots, x_k . Este procedimiento es frecuentemente pensado como una prueba global del modelo. Las hipótesis adecuadas son:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ para al menos una } j$$

Rechazar la hipótesis nula implica que al menos una de las regresoras x_1, x_2, \dots, x_k contribuye de forma significativa al modelo.

Esta prueba de significancia, es una generalización del *análisis de varianza* usado en la regresión simple. La *suma de cuadrados total* SC_T es particionada en una *suma de cuadrados debida a la regresión*, SC_R , y en una *suma de cuadrados de los residuales*, SC_{Res} . Entonces,

$$SC_T = SC_R + SC_{Res}$$

PARTE IV. EXTENSIÓN

En el **Apéndice C** de la *Parte IV* se muestra que si la hipótesis nula es verdadera, entonces SC_R / σ^2 sigue una distribución χ_k^2 , la cual tiene el mismo número de grados de libertad que variables regresoras en el modelo. También se muestra que $SC_{Res} / \sigma^2 \sim \chi_{n-k-1}^2$, y que SC_{Res} y SC_R son independientes. Por definición, la estadística F está dada por

$$F_0 = \frac{SC_R / k}{SC_{Res} / (n - k - 1)} = \frac{CM_R}{CM_{Res}}$$

con una distribución $F_{k, n-k-1}$. El **Apéndice C** muestra también que

$$E(CM_{Res}) = \sigma^2$$

$$E(CM_R) = \sigma^2 + \frac{\boldsymbol{\beta}^* \mathbf{X}'_C \mathbf{X}_C \boldsymbol{\beta}^*}{k}$$

donde $\boldsymbol{\beta}^* = (\beta_1, \beta_2, \dots, \beta_k)'$ y \mathbf{X}_C es la matriz del modelo “centrado” dada por

$$\mathbf{X}_C = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{i1} - \bar{x}_1 & x_{i2} - \bar{x}_2 & \cdots & x_{ik} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{bmatrix}$$

Estas esperanzas de cuadrados medios indican que si el valor observado de F_0 es grande, entonces es probable que exista al menos un $\beta_j \neq 0$. Por lo tanto, para probar la hipótesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$, se debe calcular la estadística F_0 y rechazar H_0 si

$$F_0 > F_{\alpha, k, n-k-1}$$

La prueba de significancia global es normalmente resumida en una *tabla de análisis de varianza* (ANOVA), cuyo modelo se muestra a continuación:

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F_0
Regresión	SC_R	k	CM_R	CM_R / CM_{Res}
Residual	SC_{Res}	$n - k - 1$	CM_{Res}	
Total	SC_T	$n - 1$		

Tabla 4.1. Tabla de ANOVA.

4.3. PRUEBAS DE HIPÓTESIS EN LA REGRESIÓN LINEAL MÚLTIPLE

Una expresión para SC_R puede ser encontrada comenzando por

$$SC_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \quad (4.19)$$

y ya que

$$SC_T = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$$

Se puede escribir entonces

$$SC_{Res} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} - \left[\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \right] \quad (4.20)$$

o equivalentemente

$$SC_{Res} = SC_T - SC_R \quad (4.21)$$

Por lo tanto, la *suma de cuadrados debida a la regresión* es

$$SC_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (4.22)$$

la *suma de cuadrados de residuales* es

$$SC_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \quad (4.23)$$

y la *suma de cuadrados total* es

$$SC_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (4.24)$$

4.3.2. Coeficientes de determinación

Otras dos formas de evaluar la adecuación del modelo son R^2 y R^2 ajustada, denotada esta última como R_{Ajust}^2 . En general, R^2 siempre se incrementa cuando una variable regresora es agregada al modelo sin importar la significancia de su contribución. Por lo tanto, es difícil juzgar si un incremento en R^2 está diciendo algo importante.

Algunos estadísticos prefieren usar una R^2 ajustada, definida como

$$R_{Ajust}^2 = 1 - \frac{SC_{Res} / (n - p)}{SC_T / (n - 1)} \quad (4.25)$$

Debido a que $SC_T / (n - 1)$ es constante sin importar cuantas variables se encuentren en el modelo, el valor de R_{Ajust}^2 se incrementará solamente agregando variables que reduzcan el cuadrado medio de los residuales, disminuyendo de esta forma la variabilidad total. La R_{Ajust}^2 no varía por agregar términos que no son útiles, por lo que es muy eficiente evaluando y comparando distintos modelos de regresión.

4.3.3. Prueba de los coeficientes individuales de la regresión

Una vez que se ha determinado que al menos una de las regresoras es importante, es lógico pensar cuál o cuáles de ellas lo son. Agregar una variable al modelo, siempre hace que la suma de cuadrados debida a la regresión se incremente y la suma de cuadrados de los residuales disminuya. Se debe decidir si el incremento en SC_R es importante para considerar el uso de la regresora adicional en el modelo. La adición de una variable también incrementa la varianza del valor estimado \hat{y} , por lo que se debe tener mucho cuidado incluyendo sólo regresoras con valores reales para explicar a la variable de respuesta. Además, agregando una regresora sin importancia, se puede incrementar el cuadrado medio de los residuales, el cual podría disminuir la utilidad del modelo.

La prueba de hipótesis para determinar la significancia de cualquier coeficiente de manera individual, β_j por ejemplo, es

4.3. PRUEBAS DE HIPÓTESIS EN LA REGRESIÓN LINEAL MÚLTIPLE

$$\begin{aligned}H_0 : \beta_j &= 0 \\H_1 : \beta_j &\neq 0\end{aligned}\tag{4.26}$$

Si H_0 no es rechazada, entonces esto indica que la variable regresora x_j puede ser eliminada del modelo. La *estadística de prueba* para esta hipótesis es

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{\text{ee}(\hat{\beta}_j)}\tag{4.27}$$

donde C_{jj} es el elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$ correspondiente a $\hat{\beta}_j$. La hipótesis nula $H_0 : \beta_j = 0$ es rechazada si $|t_0| > t_{\alpha/2, n-k-1}$. Nótese que ésta es sólo una *prueba parcial* o *marginal*, debido a que el coeficiente $\hat{\beta}_j$ depende de todas las demás variables regresoras $x_i (i \neq j)$ que se encuentran en el modelo. Entonces, ésta es una prueba de la *contribución* de x_j *dadas las otras variables regresoras en el modelo*.

4.4. Intervalos de confianza en la regresión lineal múltiple

Los intervalos de confianza de los coeficientes individuales y de la respuesta media, juegan un papel importante en la regresión lineal múltiple, tal y como lo hacen en la regresión lineal simple.

4.4.1. Intervalos de confianza para los coeficientes de la regresión

Para construir los *intervalos de confianza* para los coeficientes β_j , se seguirá suponiendo que los errores ε_i son independientes y normalmente distribuidos con media cero y varianza σ^2 . Lo mismo para las observaciones y_i , con media $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$ y varianza σ^2 . Debido a que el estimador por mínimos cuadrados $\hat{\beta}$ es una combinación lineal de las observaciones, se sigue que $\hat{\beta}$ es normalmente distribuido con media β y matriz de covarianzas $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Esto lleva a que la distribución marginal de cada coeficiente $\hat{\beta}_j$ es normal con media β_j y varianza $\sigma^2 C_{jj}$, donde C_{jj} es el j -ésimo elemento de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Consecuentemente, cada una de las estadísticas

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \quad j = 0, 1, \dots, k \quad (4.28)$$

tiene distribución t con $n - p$ grados de libertad, donde $\hat{\sigma}^2$ es el estimador de la varianza del error obtenido en (4.18). Basándose en la ecuación anterior, se puede definir un intervalo de confianza del $(1 - \alpha) \times 100\%$ para el coeficiente β_j , $j = 0, 1, \dots, k$, como

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \quad (4.29)$$

siendo

$$ee(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \quad (4.30)$$

el error estándar del coeficiente $\hat{\beta}_j$.

4.4.2. Intervalo de confianza para la respuesta media

Se puede construir un intervalo de confianza para la respuesta media en un punto particular, por ejemplo: $x_{01}, x_{02}, \dots, x_{0k}$. Se define el vector

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

El estimador puntual es

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} \quad (4.31)$$

Este es un estimador insesgado de $E(y | \mathbf{x}_0)$, ya que $E(\hat{y}_0) = \mathbf{x}'_0 \boldsymbol{\beta} = E(y | \mathbf{x}_0)$, y la varianza de \hat{y}_0 es

$$\text{Var}(\hat{y}_0) = \sigma^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 \quad (4.32)$$

Por lo tanto, un intervalo de confianza del $(1-\alpha) \times 100\%$ para la respuesta media en el punto $x_{01}, x_{02}, \dots, x_{0k}$ es

$$\begin{aligned} \hat{y}_0 - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0} \\ \leq E(y | \mathbf{x}_0) \leq \hat{y}_0 + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0} \end{aligned} \quad (4.33)$$

Ésta es la generalización para la regresión múltiple de la ecuación (2.46).

4.5. Predicción de nuevas observaciones

El modelo de regresión puede ser usado para predecir futuras observaciones de la variable y , correspondientes a valores particulares de las regresoras, por ejemplo: $x_{01}, x_{02}, \dots, x_{0k}$.

Si $\mathbf{x}'_0 = [1, x_{01}, x_{02}, \dots, x_{0k}]$, entonces un estimador puntual de la futura observación y_0 es

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} \quad (4.34)$$

Un intervalo de predicción del $(1-\alpha) \times 100\%$ para esta futura observación es

$$\begin{aligned} \hat{y}_0 - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0)} \\ \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0)} \end{aligned} \quad (4.35)$$

Ésta es la generalización para la regresión múltiple de la ecuación (2.48).

4.6. Un ejemplo numérico en R

Aquí se aplica la teoría de la regresión lineal múltiple en el caso especial cuando se tienen dos variables regresoras. Al igual que el ejemplo numérico de la *Parte II*, se harán primero algunos cálculos, verificados después con R.

4.6.1. Los discos de vinilo de la BBC

La BBC de Londres posee una basta colección de discos de vinilo de todo tipo de música, época, y en todos los formatos que utiliza en sus transmisiones. Uno de sus almacenes, de discos sencillos (*singles*) de 7 pulgadas, debe ser cotizado durante una auditoria solicitada por la misma BBC, no sólo para conocer el valor aproximado de dicho almacén, sino también, para predecir el precio que alcanzarán todos o algunos discos específicos en años futuros, ya que, por irónico que parezca, existe una tendencia a la alza en el precio de estos objetos en un mundo cada vez más digitalizado. El auditor encargado del estudio, experto en el ámbito musical, sugiere que las dos variables más importantes que afectan el precio (y) de estos discos son la antigüedad en años (x_1) que tienen estos objetos como tales, y los niveles de audiencia o popularidad (x_2) medida en puntos *rating* que llegan a alcanzar históricamente en las señales de la BBC alrededor del mundo cuando estos *singles* son retransmitidos. El auditor toma al azar 25 discos y registra sus características en la siguiente tabla, siendo y el precio fijado para el año en curso del *single* por una prestigiosa tienda de música internacional.

4.6. UN EJEMPLO NUMÉRICO EN R

Número de la observación	Precio (Libras)	Antigüedad (años)	Popularidad (puntos de rating)
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Tabla 4.2. Datos para el ejemplo de los discos de vinilo.

Las gráficas pueden ser de gran utilidad para identificar la aparente relación lineal entre las variables involucradas. La siguiente figura es una *matriz de dispersión* compuesta por seis diagramas de dispersión bidimensionales que muestra la relación entre cada par de variables. Esta matriz da un sentido de linealidad o no-linealidad entre la relación y algunos indicios de como los puntos de datos individuales forman un arreglo sobre toda la región.

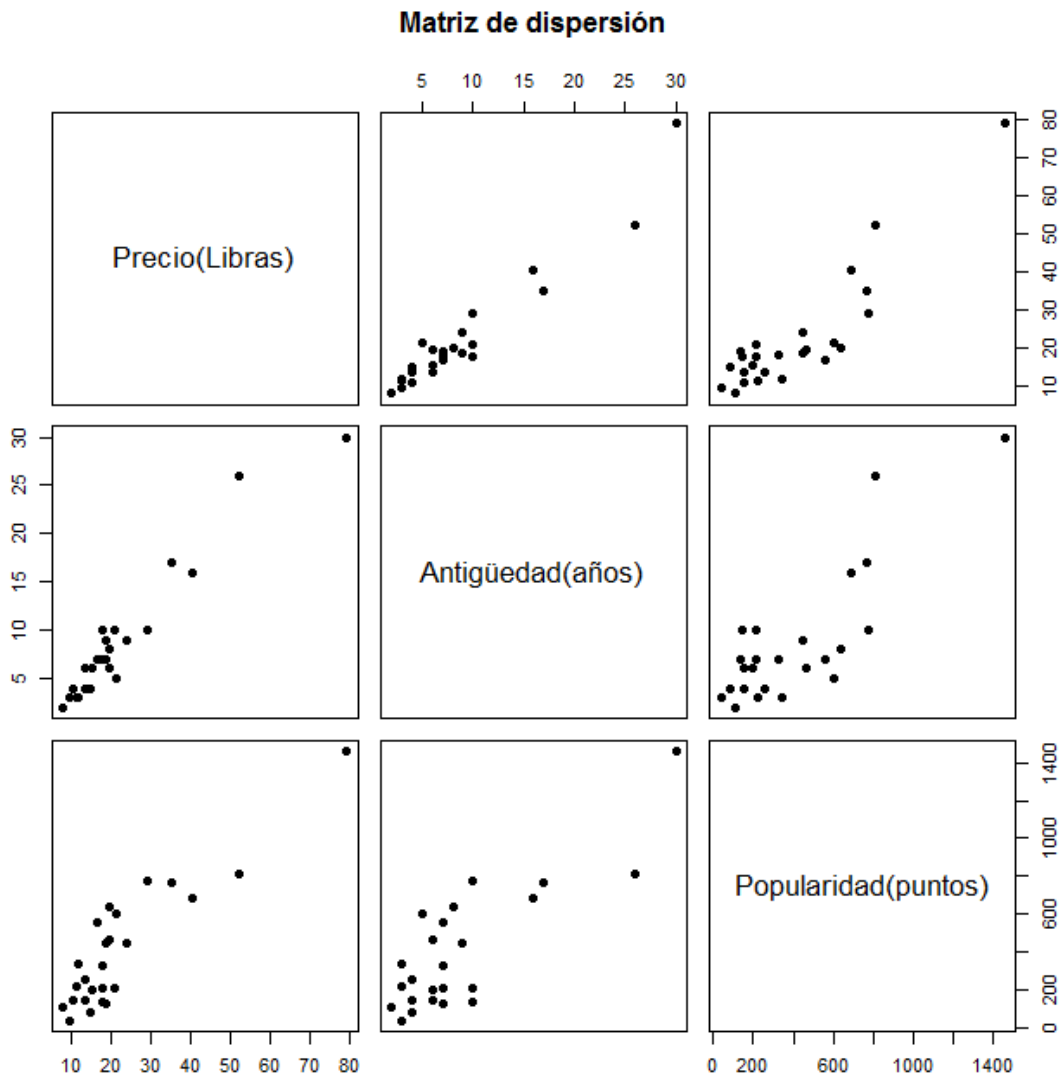


Figura 4.6. Matriz de dispersión.

4.6. UN EJEMPLO NUMÉRICO EN R

Para ajustar el modelo de regresión múltiple, primero se forma la matriz \mathbf{X} y el vector \mathbf{y} .

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ 1 & 3 & 340 \\ 1 & 4 & 80 \\ 1 & 6 & 150 \\ 1 & 7 & 330 \\ 1 & 2 & 110 \\ 1 & 7 & 210 \\ 1 & 30 & 1460 \\ 1 & 5 & 605 \\ 1 & 16 & 688 \\ 1 & 10 & 215 \\ 1 & 4 & 255 \\ 1 & 6 & 462 \\ 1 & 9 & 448 \\ 1 & 10 & 776 \\ 1 & 6 & 200 \\ 1 & 7 & 132 \\ 1 & 3 & 36 \\ 1 & 17 & 770 \\ 1 & 10 & 140 \\ 1 & 26 & 810 \\ 1 & 9 & 450 \\ 1 & 8 & 635 \\ 1 & 4 & 150 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 16.68 \\ 11.50 \\ 12.03 \\ 14.88 \\ 13.75 \\ 18.11 \\ 8.00 \\ 17.83 \\ 79.24 \\ 21.50 \\ 40.33 \\ 21.00 \\ 13.50 \\ 19.75 \\ 24.00 \\ 29.00 \\ 15.35 \\ 19.00 \\ 9.50 \\ 35.10 \\ 17.90 \\ 52.32 \\ 18.75 \\ 19.83 \\ 10.75 \end{bmatrix}$$

La matriz $\mathbf{X}'\mathbf{X}$ es

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix} \\ &= \begin{bmatrix} 25 & 219 & 10,232 \\ 219 & 3,055 & 133,899 \\ 10,232 & 133,899 & 6,725,688 \end{bmatrix} \end{aligned}$$

y el vector $\mathbf{X}'\mathbf{y}$ es

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7,375.44 \\ 337,072.00 \end{bmatrix}$$

El estimador de $\boldsymbol{\beta}$ es

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

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 219 & 10,232 \\ 219 & 3,055 & 133,899 \\ 10,232 & 133,899 & 6,725,688 \end{bmatrix}^{-1} \begin{bmatrix} 559.60 \\ 7,375.44 \\ 337,072.00 \end{bmatrix} \\ &= \begin{bmatrix} .113215118 & -.00444859 & -.00008367 \\ -.00444859 & .00274378 & -.00004786 \\ -.00008367 & -.00004786 & .00000123 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7,375.44 \\ 337,072.00 \end{bmatrix} \\ &= \begin{bmatrix} 2.34123115 \\ 1.61590712 \\ 0.01438483 \end{bmatrix} \end{aligned}$$

Entonces, el ajuste por mínimos cuadrados es

$$\hat{y} = 2.34123 + 1.61591x_1 + 0.01438x_2$$

Se encuentra ahora el estimador de la varianza del error σ^2 para el ajuste de regresión múltiple. Se tiene

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{25} y_i^2 = 18,310.6290$$

y

$$\begin{aligned} \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} &= [2.34123115 \quad 1.61590721 \quad 0.01438483] \begin{bmatrix} 559.60 \\ 7,375.44 \\ 337,072.00 \end{bmatrix} \\ &= 18,076.90304 \end{aligned}$$

La suma de cuadrados de residuales es

4.6. UN EJEMPLO NUMÉRICO EN R

$$\begin{aligned}SC_{\text{Res}} &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= 18,310.6290 - 18,076.9030 = 233.7260\end{aligned}$$

Entonces, el estimador para σ^2 es el cuadrado medio de los residuales

$$\hat{\sigma}^2 = \frac{SC_{\text{Res}}}{n-p} = \frac{233.7260}{25-3} = 10.6239$$

Prueba de significancia para la regresión

Se tiene primero que

$$\begin{aligned}SC_{\text{T}} &= \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \\ &= 18,310.6290 - \frac{(559.60)^2}{25} = 5784.5426 \\ SC_{\text{R}} &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \\ &= 18,076.9030 - \frac{(559.60)^2}{25} = 5550.8166 \\ SC_{\text{Res}} &= SC_{\text{T}} - SC_{\text{R}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{y} = 233.7260\end{aligned}$$

Para probar $H_0 : \beta_1 = \beta_2 = 0$, se calcula la estadística

$$F_0 = \frac{CM_{\text{R}}}{CM_{\text{Res}}} = \frac{5550.8166 / 2}{233.7260 / (25 - 3)} = \frac{2775.4083}{10.6239} = 261.24$$

Tomando $\alpha = 0.01$ se encuentra que $F_{0.01,2,22} = 5.72 < 261.24$. Debido a que el p -value (obtenido directamente de **R**) es muy pequeño, se rechaza $H_0 : \beta_1 = \beta_2 = 0$. Esto indica que la variable *precio* está relacionada a las variables *antigüedad* y/o *popularidad*. Se puede construir la tabla ANOVA con los datos obtenidos.

PARTE IV. EXTENSIÓN

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	F_0	$p - value$
Regresión	5550.8166	2	2775.4083	261.24	4.7×10^{-10}
Residual	233.7260	22	10.6239		
Total	5784.5426	24			

Tabla 4.3. Tabla de ANOVA para el ejemplo de los discos de vinilo.

Pruebas de hipótesis para coeficientes individuales

Para ilustrar el procedimiento, se desea probar la significancia de la variable regresora *popularidad* (x_2) dado que la regresora *antigüedad* (x_1) está en el modelo. Las hipótesis son

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

El elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$ correspondiente a β_2 es $C_{22} = 0.00000123$, por lo que la estadística t está dada por

$$t_0 = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{0.01438}{\sqrt{(10.6239)(0.00000123)}} = 3.98$$

Ya que $t_{0.025,22} = 2.074 < |3.98|$, se rechaza $H_0 : \beta_2 = 0$ y se concluye que la regresora *popularidad*, ó x_2 , contribuye significativamente al modelo dado que *antigüedad*, ó x_1 , está también en el modelo.

Intervalos de confianza para los coeficientes de la regresión

Se obtiene un intervalo de confianza para el parámetro β_1 del 95%. El estimador de β_1 es $\hat{\beta}_1 = 1.61591$, el elemento diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$ correspondiente a β_1 es $C_{11} = 0.00274378$ y $\hat{\sigma}^2 = 10.6239$. Usando la ecuación (4.29), se encuentra que

$$\hat{\beta}_1 - t_{0.025,22} \sqrt{\hat{\sigma}^2 C_{11}} \leq \beta_1 \leq \hat{\beta}_1 + t_{0.025,22} \sqrt{\hat{\sigma}^2 C_{11}}$$

4.6. UN EJEMPLO NUMÉRICO EN R

$$1.61591 - (2.074)\sqrt{(10.6239)(0.00274378)} \\ \leq \beta_1 \leq 1.61591 + (2.074)\sqrt{(10.6239)(0.00274378)} \\ 1.61591 - (2.074)(0.17073) \leq \beta_1 \leq 1.61591 + (2.074)(0.17073)$$

por lo que el intervalo de confianza de β_1 del 95% es

$$1.26181 \leq \beta_1 \leq 1.97001$$

Intervalos de confianza para la respuesta media

Se desea construir un intervalo de confianza del 95% para la respuesta media del precio para un disco con $x_1 = 8$ años, y $x_2 = 275$ puntos de popularidad. Entonces,

$$x_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$

El valor estimado en este punto es encontrado con la ecuación (4.31).

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} = [1 \quad 8 \quad 275] \begin{bmatrix} 2.34123 \\ 1.61591 \\ 0.01438 \end{bmatrix} = 19.22 \text{ £}$$

La varianza de \hat{y}_0 es estimada por

$$\hat{\sigma}^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 = 10.6239 [1 \quad 8 \quad 275] \\ \times \begin{bmatrix} .113215118 & -.00444859 & -.00008367 \\ -.00444859 & .00274378 & -.00004786 \\ -.00008367 & -.00004786 & .00000123 \end{bmatrix} \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix} \\ = 10.6239(0.05346) = 0.56794$$

Entonces, un intervalo de confianza para la respuesta media del precio en este punto es encontrado de la ecuación (4.33) como

$$19.22 - 2.074\sqrt{0.56794} \leq E(y | x_0) \leq 19.22 + 2.074\sqrt{0.56794}$$

el cual se reduce a

$$17.66 \leq E(y | x_0) \leq 20.78$$

95% de tales intervalos contendrán el verdadero valor del precio.

Intervalos de predicción para nuevas observaciones

Se desea construir un intervalo de predicción del 95% para el precio, en donde nuevamente, $x_1 = 8$ y $x_2 = 275$. Se tiene que $\mathbf{x}'_0 = [1 \ 8 \ 275]$, y el precio estimado en este punto es $\hat{y}_0 = \mathbf{x}'_0 \mathbf{b} = 19.22$ libras. También se encontró anteriormente que $\mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 = 0.05346$. Entonces, de la ecuación (4.35) se tiene

$$19.22 - 2.074 \sqrt{10.6239(1 + 0.05346)} \leq \hat{y}_0 \leq 19.22 + 2.074 \sqrt{10.6239(1 + 0.05346)}$$

y el intervalo de predicción del 95% es

$$12.28 \leq y_0 \leq 26.16$$

Este intervalo de predicción es más angosto que el intervalo de confianza para la respuesta media en el mismo punto como era de esperarse.

4.6.2. Resultados en R para el ejemplo de los vinilos

Se inicia guardando los datos en un archivo .txt como se muestra en la siguiente imagen.

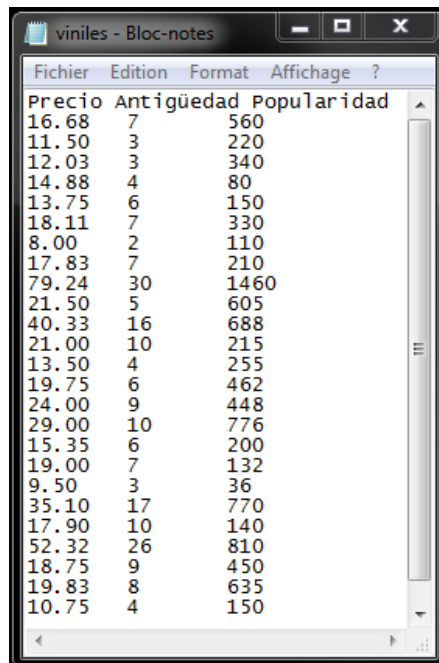


Figura 4.7. Datos para el ejemplo de los discos de vinilo.

4.6. UN EJEMPLO NUMÉRICO EN R

El procedimiento para tener acceso a los datos es el mismo que en el ejemplo numérico de la *Parte II*.

```
> setwd("c:/Users/Gérard/Documents/R")
> datos.viniles=read.table(file="viniles.txt",header=TRUE)
```

Se asignan las columnas a cada variable y se grafica la matriz de dispersión.

```
> Precio<-datos.viniles[,1]
> Antig<-datos.viniles[,2]
> Popul<-datos.viniles[,3]
> pairs(datos.viniles,labels=c("Precio (Libras)", "Antigüedad(años)",
+ "Popularidad(puntos)"),pch=19,main='Matriz de dispersión', cex.main=1.0)
```

Estimación de los parámetros

Se utiliza la fórmula de R para el modelo de regresión Precio~Antig+Popul dentro de la función lm() y se despliega la información del ajuste con summary(). Esto es:

```
> regresion=lm(Precio~Antig+Popul)
> summary(regresion)

Call:
lm(formula = Precio ~ Antig + Popul)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7880 -0.6629  0.4364  1.1566  7.4197

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.341231   1.096730   2.135 0.044170 *
Antig        1.615907   0.170735   9.464 3.25e-09 ***
Popul        0.014385   0.003613   3.981 0.000631 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared:  0.9596,    Adjusted R-squared:  0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

Los resultados obtenidos anteriormente pueden ser verificados.

Se despliegan los errores estándar de los coeficientes, así como los valores de la estadística t . Esto hace mucho más sencilla la construcción de intervalos de confianza y pruebas de hipótesis.

Los valores ajustados pueden ser obtenidos usando la función predict(), y los residuales mediante la función resid().

PARTE IV. EXTENSIÓN

```
> data.frame(datos.viniles, valores.ajustados=predict(regresion),
+ residuales=resid(regresion))
  Precio Antigüedad Popularidad valores.ajustados residuales
1  16.68           7           560         21.708084 -5.0280843
2  11.50           3           220         10.353615  1.1463854
3  12.03           3           340         12.079794 -0.0497937
4  14.88           4            80          9.955646  4.9243539
5  13.75           6           150         14.194398 -0.4443983
6  18.11           7           330         18.399574 -0.2895743
7   8.00           2           110          7.155376  0.8446235
8  17.83           7           210         16.673395  1.1566049
9  79.24          30          1460         71.820294  7.4197062
10 21.50           5           605         19.123587  2.3764129
11 40.33          16           688         38.092507  2.2374930
12 21.00          10           215         21.593041 -0.5930409
13 13.50           4           255         12.472991  1.0270093
14 19.75           6           462         18.682464  1.0675359
15 24.00           9           448         23.328798  0.6712018
16 29.00          10           776         29.662928 -0.6629284
17 15.35           6           200         14.913640  0.4363603
18 19.00           7           132         15.551379  3.4486213
19  9.50           3            36          7.706807  1.7931935
20 35.10          17           770         40.887970 -5.7879699
21 17.90          10           140         20.514179 -2.6141789
22 52.32          26           810         56.006528 -3.6865279
23 18.75           9           450         23.357568 -4.6075679
24 19.83           8           635         24.402854 -4.5728535
25 10.75           4           150         10.962584 -0.2125839
```

Análisis de varianza

Se despliega la tabla del análisis de varianza con la función `anova()`.

```
> anova(regresion)
Analysis of Variance Table

Response: Precio
      Df Sum Sq Mean Sq F value    Pr(>F)
Antig  1  5382.4   5382.4  506.619 < 2.2e-16 ***
Popul  1   168.4    168.4   15.851 0.0006312 ***
Residuals 22   233.7     10.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los resultados obtenidos anteriormente pueden ser verificados.

Intervalos de confianza para β_0 , β_1 y β_2

Se obtienen los intervalos de confianza del 95% para los parámetros. Esta vez de forma directa con la función `confint()`.

4.6. UN EJEMPLO NUMÉRICO EN R

```
> confint(regresion, level=0.95)
              2.5 %      97.5 %
(Intercept) 0.066751987 4.61571030
Antig       1.261824662 1.96998976
Popul      0.006891745 0.02187791
```

Intervalos de predicción para nuevas observaciones

Se utiliza la función predict(), especificando interval="prediction" y un nivel de confianza con un porcentaje predefinido del 95% .

```
> pred.nuevas=predict(regresion, interval="prediction")
Mensajes de aviso perdidos
In predict.lm(regresion, interval = "prediction") :
  Predictions on current data refer to _future_ responses

> pred.nuevas
      fit      lwr      upr
1  21.708084 14.6126113 28.80356
2  10.353615  3.3589989 17.34823
3  12.079794  4.9942032 19.16538
4   9.955646  2.9132656 16.99803
5  14.194398  7.1857225 21.20307
6  18.399574 11.4964758 25.30267
7   7.155376  0.1246073 14.18615
8  16.673395  9.7016031 23.64519
9  71.820294 63.5460584 80.09453
10 19.123587 11.7301065 26.51707
11 38.092507 31.0476684 45.13735
12 21.593041 14.4595011 28.72658
13 12.472991  5.5097274 19.43625
14 18.682464 11.6632578 25.70167
15 23.328798 16.4315145 30.22608
16 29.662928 22.3638539 36.96200
17 14.913640  7.9559322 21.87135
18 15.551379  8.4737709 22.62899
19  7.706807  0.6285915 14.78502
20 40.887970 33.7928734 47.98307
21 20.514179 13.2171816 27.81118
22 56.006528 48.0324043 63.98065
23 23.357568 16.4597897 30.25535
24 24.402854 17.2470809 31.55863
25 10.962584  3.9812364 17.94393
```

El valor proporcionado es nuevamente una matriz.

Intervalos de confianza para la respuesta media

El procedimiento para obtener los intervalos de confianza para la respuesta media $\hat{\mu}_{y|x_0}$ es el mismo que en el de intervalos de predicción para nuevas observaciones. La única diferencia, es que ahora se cambia el argumento a `interval="confidence"`. Se tiene entonces:

```
> inter.respmed=predict(regresion, interval="confidence")
> inter.respmed
      fit      lwr      upr
1 21.708084 19.551297 23.864872
2 10.353615  8.556216 12.151013
3 12.079794  9.955744 14.203843
4  9.955646  7.980522 11.930770
5 14.194398 12.343039 16.045757
6 18.399574 17.000017 19.799132
7  7.155376  5.222061  9.088692
8 16.673395 14.966973 18.379818
9 71.820294 67.048610 76.591977
10 19.123587 16.128667 22.118507
11 38.092507 36.108636 40.076378
12 21.593041 19.314141 23.871941
13 12.472991 10.801755 14.144227
14 18.682464 16.791631 20.573297
15 23.328798 21.958209 24.699388
16 29.662928 26.909298 32.416559
17 14.913640 13.265705 16.561574
18 15.551379 13.454112 17.648645
19  7.706807  5.607492  9.806121
20 40.887970 38.732422 43.043518
21 20.514179 17.766059 23.262299
22 56.006528 51.776559 60.236497
23 23.357568 21.984492 24.730644
24 24.402854 22.055286 26.750421
25 10.962584  9.217532 12.707636
```

PARTE V

APLICACIÓN

5. Optimización en ingeniería financiera: El método Monte Carlo de mínimos cuadrados de Longstaff-Schwartz

5.1. Contexto

Estudiando diversas aplicaciones de carácter actuarial, en donde los mínimos cuadrados son el centro de atención, tuve la fortuna de encontrar un tema que, además de cumplir con el objetivo de este capítulo, ratificó al mismo tiempo mi suposición de que había algo más en la toma de decisiones y estrategias financieras. *El método Monte Carlo de mínimos cuadrados de Longstaff-Schwartz* combina de forma elegante la minimización de errores con los derivados financieros, todo con el imprescindible sustento del álgebra matricial. Resultados importantes se presentan en este capítulo.

En este trabajo, se identifica la teoría del algoritmo de Monte Carlo de mínimos cuadrados (LSM por sus siglas en inglés) implementado por Longstaff y Schwartz para la valuación de opciones americanas. A partir de este momento se utilizarán las siglas LSM para referirse al método. Éste puede ser dividido en tres fases: simulación de precios, calibración y valuación. Se enfocará en cada una de ellas, finalizando con un ejemplo verificado con el más potente software. Se iniciará con una pequeña explicación de conceptos para situarse en contexto, sin ir demasiado profundo, ya que el lector seguramente ha estudiado estos conceptos básicos -pero necesarios- en cursos especializados de derivados.

Las opciones son derivados financieros que le dan el derecho al portador, pero no la obligación, de comprar y vender bienes subyacentes en un precio *strike*, definido de forma contractual sobre un rango de fechas futuras. Las opciones *Long Call* otorgan el derecho de comprar el bien subyacente, mientras que una *Long Put* da el derecho de vender. Hay muchas variedades de opciones. Las *opciones europeas* son ejercibles sólo a la fecha de maduración, denotada como T , mientras que las *opciones bermuda* pueden ser ejercidas en fechas específicas durante, o antes de la fecha de maduración.

La valuación y óptimo ejercicio de *opciones americanas* es uno de los más importantes problemas prácticos en el mercado accionario. Su valuación representa un alto grado de complejidad, debido a que estos instrumentos son ejercibles en cualquier momento, desde la fecha de compra hasta la de maduración. Encontrar un resultado analítico es básicamente imposible, por lo intratables que suelen ser las integrales estocásticas

resultantes después del análisis, por lo que aproximaciones numéricas deben ser implementadas. Este tipo de derivados son encontrados en la mayoría de los mercados financieros, tales como *commodity*, divisas, seguros, energía, crédito, estado, *swap*, mercados emergentes, y otros.

La intención es mostrar las oportunidades para un importante algoritmo llamado método *Monte Carlo de mínimos cuadrados* para la valuación de opciones americanas, atribuido a Longstaff y Schwartz, y que ha sido adoptado en el mercado de derivados. Este trabajo presenta una innovadora aproximación para la valuación de la fórmula de riesgo-neutral para el cálculo de precios de opciones americanas. A continuación se dará una breve explicación sobre este tipo de opciones, suficiente para entender después el marco teórico del método LSM de Longstaff-Schwartz, su conexión con el principio de mínimos cuadrados y su implementación.

5.2. Opciones americanas

Las opciones americanas son típicamente mucho más comunes que las europeas. Una opción americana es similar a una opción europea, excepto que el portador puede ejercer en cualquier momento entre la fecha inicial y la fecha de expiración. Una *Call* o *Put* americana es un contrato tal, que el propietario puede comprar o vender sin obligación, respectivamente, algún bien prescrito S -llamado el bien subyacente- al emisor en cualquier tiempo t entre la fecha inicial y la de maduración, a un precio preestablecido de ejercicio K -precio *strike*-. El tiempo de ejercicio τ puede ser representado como un tiempo de paro, por lo que las opciones americanas son un ejemplo de problemas de paro óptimo.

Sin un tiempo de ejercicio preestablecido, es mucho más difícil evaluar este tipo de opciones. El portador de una opción americana se ve ante el dilema de decidir cuando, si es que lo hace, ejercer. Si al tiempo t la opción está *out the money* -cuando el portador no obtiene ganancia al ejercer-, es claro que lo mejor es no hacer valer el contrato. Sin embargo, si la opción está *in the money* -cuando sí se obtiene ganancia-, podría ser benéfico esperar el tiempo hasta cuando el *payoff* -resultado al momento del ejercicio- sea el más grande en todo el período de vida de la opción.

Los componentes principales de la opción son el precio *strike* y la fecha de expiración. La dinámica del bien subyacente es generalmente descrita como una ecuación diferencial estocástica que, usualmente, contiene procesos de difusión y de salto. Debido a la complejidad de esta dinámica, la valuación y óptimo ejercicio de opciones americanas

5.2. OPCIONES AMERICANAS

sigue siendo uno de los más grandes retos en derivados financieros, particularmente cuando más de un factor afecta el valor de la opción. Técnicas binomiales y de diferencia finita se vuelven poco prácticas en situaciones donde hay factores múltiples, además de que, en cualquier momento, el portador de una opción americana compara de forma óptima el *payoff* del ejercicio inmediato con el *payoff* esperado por continuación y entonces, ejerce si el primero es más grande. De esta forma, la estrategia de óptimo ejercicio es fundamentalmente determinada por la esperanza condicional del *payoff* de continuidad de mantener viva la opción.

5.2.1. Valuación

Considérese el proceso de precio $S(t)$ que sigue un movimiento browniano geométrico de acuerdo a la siguiente ecuación diferencial estocástica:

$$dS = \mu S dt + \sigma S dX \quad (5.1)$$

en el cual μ es la medida de la tasa promedio de crecimiento del bien subyacente, también conocida como deriva, y σ es la volatilidad, la cual mide la desviación estándar del regreso -supuestas ambas como constantes-, y $X = X(t)$ es el movimiento browniano estándar.

La función de pago es $u(S, t)$. Una opción dependiente de la trayectoria es una para la cual $u(S, t)$ depende de $\{S(t) : 0 < t < T\}$; mientras una opción simple -no dependiente de la trayectoria- tiene $u(S, t) = u(S(t), t)$. Para una opción europea simple, el pago debe ser recibido únicamente al final de la fecha de maduración, esto es $f(T) = u(S(T), T)$. Para una opción americana simple, el ejercicio se realiza en cualquier tiempo antes de T , por lo que el pago es $f(\tau) = u(S(\tau), \tau)$ en el cual τ es un tiempo óptimo de paro.

En la mayoría de los trabajos clásicos, Black y Scholes (1973) y Merton (1973) describen dos métodos para la valuación de valores derivados. El primero es la famosa *ecuación diferencial parcial de Black-Scholes* (PDE por sus siglas en inglés), y el segundo, el cual es aplicable a opciones dependientes de sus trayectorias y a otros derivados para los cuales Black-Scholes se vuelve inválido o intratable, es la *fórmula de valuación neutral al riesgo*:

$$F(S, t) = \max_{\tau} E' [e^{-r(\tau-t)} u(S(\tau), \tau) | S(t) = S] \quad (5.2)$$

en donde E' es la esperanza neutral al riesgo -y condicionada además-, para la cual la tasa de crecimiento μ en (5.1) es remplazada por r . El máximo es tomado sobre todos los

tiempos de paro τ con $t < \tau < T$. Ésta es la fórmula en la que Longstaff-Schwartz se basan para aplicar su principio.

En este trabajo no se pretende abordar Black-Scholes, debido a que no tiene relación alguna con el objetivo de este capítulo (aplicación de mínimos cuadrados) pero, a manera de complemento, se presenta una breve explicación sobre este principio en el **Apéndice A** de la *Parte V*. De aquí en adelante, sólo se enfocará al método LSM de Longstaff-Schwartz.

5.3. El método LSM de Longstaff-Schwartz

Longstaff y Schwartz (2001) publicaron un algoritmo para la valuación de opciones americanas llamado método *Monte Carlo de mínimos cuadrados*. El LSM procede simulando trayectorias de precios usando *Monte Carlo*, y después realiza iteraciones hacia atrás en donde, en cada paso, efectúa una aproximación por mínimos cuadrados de las esperanzas condicionales -o funciones de continuación- sobre una combinación lineal de funciones base. Longstaff y Schwartz también analizaron las propiedades de convergencia del algoritmo. Ellos usaron sólo trayectorias *in the money* en la regresión para incrementar la eficiencia. Por su parte, Clément, Lamberton y Protter (2002) probaron que el estimador de la esperanza condicional se aproxima -con probabilidad uno- a su verdadero valor cuando el número de funciones base tiende a infinito.

Este algoritmo se implementa fácilmente sin usar la parte de la simulación *Monte Carlo* cuando se tiene algún indicio de las trayectorias de precios y cuenta con las ventajas adicionales de que las funciones de continuación son construidas explícitamente, además de que es fácil de calibrar de acuerdo a los precios de mercado existentes.

La ventaja principal para este capítulo, es que esta aproximación para la valuación de opciones americanas no necesita más que regresión simple y conceptos básicos del álgebra matricial. A continuación se presenta el marco teórico, así como importantes resultados de convergencia para el método de Longstaff-Schwartz.

5.3.1. Marco teórico

Considérese que se tiene un espacio de probabilidad (Ω, \mathcal{F}, P) y un horizonte de tiempo finito $[0, T]$, donde el espacio de estados Ω es el conjunto de todas las posibles realizaciones de la economía estocástica entre el tiempo 0 y T , y tiene como elemento típico a ω que representa una trayectoria. \mathcal{F} es la sigma álgebra de eventos distinguibles

5.3. EL MÉTODO LSM DE LONGSTAFF-SCHWARTZ

al tiempo T , y P una medida de probabilidad definida sobre los elementos de \mathcal{F} . Se define $F = \{\mathcal{F}_t; t \in [0, T]\}$, siendo la filtración aumentada, generada por los procesos de precios relevantes para los valores de la economía, en donde se supone que $\mathcal{F}_T = \mathcal{F}$. Consistentes con el paradigma del no arbitraje, se considera también la existencia de una medida de martingala equivalente Q para esta economía.

Es de interés evaluar derivados de estilo americano con flujos de efectivo aleatorios, los cuales pueden ocurrir en el intervalo $[0, T]$. Se restringe la atención a derivados con *payoffs* que son elementos del espacio de funciones que son cuadrado-integrables o de varianza finita $L^2(\Omega, \mathcal{F}, Q)$. Diversos resultados, tales como los de Bensoussan (1984) y Karatzas (1988), sugieren que el valor de una opción americana es igual al valor maximizado de los flujos de efectivo descontados, donde el máximo es tomado sobre todos los tiempos de paro con respecto a la filtración F . Se introduce la función $C(\omega, s; t, T)$, que denota la trayectoria de flujos de efectivo generados por la opción, condicional sobre el no-ejercicio en, o antes del tiempo t , y sobre el portador de la opción siguiendo la estrategia óptima de paro para toda s , $t < s \leq T$. Esta función es análoga a las matrices usadas en la siguiente sección durante la aplicación del algoritmo en los ejemplos numéricos.

El objetivo del algoritmo LSM es proveer una aproximación a la *regla de paro óptima* que maximiza el valor de la opción americana. Para transmitir la intuición detrás del algoritmo LSM, la discusión se enfoca cuando la opción americana sólo puede ser ejercida en los tiempos discretos $0 \leq t_1 \leq t_2 \leq t_3 \leq \dots \leq t_K = T$, y se considera la política de *paro óptimo* en cada fecha de ejercicio. En la práctica, por supuesto, se supone que las opciones americanas son ejercibles de manera continua; por lo que el algoritmo LSM puede ser usado para aproximar el valor de estas opciones tomando K lo suficientemente grande.

En la fecha de expiración, el inversionista ejerce si la opción está *in the money*, o continua sin ejercerla si está *out the money*. Al tiempo de ejercicio t_k antes de la fecha de expiración, sin embargo, el portador de la opción debe escoger entre ejercer inmediatamente o continuar la vida de la opción y visitar la *regla de paro* al siguiente tiempo de ejercicio. El valor de la opción es el *payoff* maximizado, y debido a la incondicionalidad entre las fechas de ejercicio, el inversionista ejerce tan pronto como el valor del ejercicio inmediato es más grande o igual al valor de continuación. Al tiempo t_k , el flujo de efectivo del ejercicio inmediato es conocido por el inversionista. Los flujos de efectivo de continuación, por supuesto, no son conocidos al tiempo t_k . Sin embargo, la teoría del no-arbitraje implica que el valor de continuación o, equivalentemente, el valor de la opción -asumiendo que ésta no puede ser ejercida hasta después de t_k^- , está dado

tomando la esperanza condicional de los restantes flujos de efectivo $C(\omega, s; t_k, T)$ con respecto a la medida de valuación neutral al riesgo Q . Específicamente, al tiempo t_k el valor de continuación $F(\omega; t_k)$ puede ser expresado como:

$$F(\omega; t_k) = E_Q \left[\sum_{j=k+1}^{\tau} \exp(-r(\omega, t_j; t_k)) C(\omega, t_j; t_k, T) \mid \mathcal{F}_{t_k} \right] \quad (5.3)$$

donde $r(\omega, t)$ es la tasa libre de riesgo y la esperanza es condicionada sobre el conjunto de información \mathcal{F}_{t_k} al tiempo t_k . Con esta representación, el problema de óptimo ejercicio se reduce a comparar el valor del ejercicio inmediato con la esperanza condicional y entonces, ejercer tan pronto como el valor del ejercicio inmediato es positivo y más grande o igual a la esperanza condicional. A continuación se presenta un diagrama que resume de forma general el algoritmo LSM.

Algoritmo 1: Algoritmo LSM para la valuación de una opción americana

entrada: Precio *strike* K , tasa libre de riesgo r , k tiempos de paro τ , número de trayectorias n .

salida: Precio de la opción americana.

genera n trayectorias, cada una con k tiempos de paro usando simulaciones Monte-Carlo.

para $j, j-1 \rightarrow 1$ **haz**

```

|   para cada trayectoria  $\omega$  haz
|   |   Encuentra el estimador  $\hat{F}(\omega, t_j)$  del valor de continuación en  $t_j$  para  $\omega$ .
|   |   El valor de ejercicio en  $t_j$  para la trayectoria  $\omega$  es  $E(\omega, t_j) = \max(K - S(t_j, \omega), 0)$  para la
|   |   opción Put, donde  $S(t_j, \omega)$  es el precio stock simulado al tiempo  $t_j$  para la trayectoria  $\omega$ 
|   |   si  $\hat{F}(\omega, t_j) \leq E(\omega, t_j)$  entonces
|   |   |   La estrategia óptima en  $t_j$  es ejercer en ese momento.
|   |   |   si no
|   |   |   |   La estrategia óptima en  $t_j$  es la de  $t_{j+1}$ .
|   |   fin
|   fin
fin
    
```

descuenta los flujos de efectivo resultantes, desde el tiempo de paro hasta al tiempo cero.

regresa el promedio del precio inicial de la opción sobre todas las trayectorias generadas.

Éste será el valor aproximado de V , el valor real de la opción al tiempo cero.

5.3.2. Calibración en LSM

La clave para la aproximación LSM es el uso de mínimos cuadrados para estimar el *payoff* esperado y condicionado, usando la información de la sección transversal en las simulaciones. En este paso de calibración, la función de continuación $F(\omega, t_k)$ de la ecuación (5.3) es aproximada como una combinación lineal de un conjunto numerable de funciones base para \mathcal{F}_{t_k} . En particular, si las primeras m funciones base son denotadas por $\{L_k(\cdot), k = 0, \dots, m-1\}$, entonces la aproximación para $F(\omega, t_j)$ con $j = 1, \dots, K-1$ puede ser escrita como:

$$\hat{F}(\omega; t_j) = \sum_{k=0}^{m-1} a_k L_k(S(t_j, \omega)) \quad (5.4)$$

donde $S(t_j, \omega)$ es el precio del bien subyacente en el j -ésimo tiempo para la trayectoria ω , mientras que a_k es el coeficiente correspondiente a la k -ésima función base L_k . Esta suposición puede ser formalmente justificada, por ejemplo, cuando la esperanza condicional es un elemento del espacio L^2 de funciones cuadrado-integrables para alguna medida. Debido a que L^2 es un espacio de Hilbert, existe una base ortonormal numerable, con la que la esperanza condicional puede ser representada como una función lineal de los elementos de la base.

Los estimados de las funciones de continuación pueden ser interpretados como el precio aproximado de la opción al tiempo t_j sobre la trayectoria ω . Las tuplas de los precios de la opción, así como del bien subyacente, forman los datos utilizados en la regresión por mínimos cuadrados que determina los coeficientes a_k .

En el planteamiento original del algoritmo, Longstaff y Schwartz (1999) utilizaron regresión por mínimos cuadrados sobre monomios para la ecuación (5.4). Suponiendo por ejemplo que X es el valor del bien subyacente, entonces Longstaff y Schwartz usaron las siguientes funciones base:

$$\begin{aligned} L_0(X) &= 1 \\ L_1(X) &= X \\ L_2(X) &= X^2 \end{aligned} \quad (5.5)$$

Otra posible elección de funciones base es el conjunto de polinomios de Laguerre:

PARTE V. APLICACIÓN

$$\begin{aligned}
 L_0(X) &= 1 \\
 L_1(X) &= 1 - X \\
 L_2(X) &= 1 - 2X + X^2 / 2 \\
 &\vdots \\
 L_n(X) &= \frac{e^X}{n!} \frac{d^n}{dX^n} (X^n e^{-X}),
 \end{aligned} \tag{5.6}$$

estas funciones son ortonormales entre sí con respecto al producto interior dado por

$$\langle f, g \rangle = \int_0^\infty f(x)g(x)e^{-x} dx$$

Diversos tipos de funciones base incluyen los polinomios de Hermite, Legendre, Chebyshev, Gegenbauer y Jacobi.

Se define entonces a X_j como la matriz de dimensión $n \times m$ para el j -ésimo tiempo de paro con $X_j(i, k) = L_k(S(t_j, \omega_i))$, donde ω_i es la i -ésima trayectoria, $i = 1, \dots, n$. Sea $B_j = [a_k]_j$ el vector desconocido de coeficientes para el j -ésimo tiempo, a ser determinado mediante regresión. El objetivo es encontrar el vector B_j de coeficientes para las funciones base que minimiza $\|X_j B_j - Y_j\|_2$. Los valores del vector columna $[Y_i]_j$ son determinados descontando el precio de la opción sobre la trayectoria i , desde el tiempo t_{j+1} hasta t_j . Una vez que los coeficientes a_k son determinados, entonces los precios de la trayectoria i son actualizados como sigue: se compara el valor estimado de continuación $\hat{F}(\omega; t_k)$ con el valor de ejercicio en $S(t_j, \omega_i)$. Si el valor de ejercicio es más grande, entonces se usa éste como la actualización del precio de la opción para la trayectoria i , de otra manera, se utiliza el precio de la trayectoria al tiempo t_{j+1} descontado mediante $r(\omega_i, t_j; t_{j+1})$.

Sea $U \Sigma V^T$ la descomposición del valor singular (SVD por siglas en inglés) de X_j , donde U es una matriz ortogonal de $n \times n$, V es una matriz ortogonal de $m \times m$, y Σ es una matriz diagonal de $n \times m$, entonces $\|X_j B_j - Y_j\|_2$ es minimizado para

$$B_j = (X_j^T X_j)^{-1} X_j^T Y_j = (V \Sigma^{-1} U^T) Y_j \tag{5.7}$$

5.3. EL MÉTODO LSM DE LONGSTAFF-SCHWARTZ

La regresión provee un estimador directo de los coeficientes a_k para la función de la esperanza condicional $\hat{F}(\omega, t_k)$, expresada como una combinación lineal de las funciones base como en (5.4). El algoritmo para el precio de continuación es mostrado en el siguiente pseudocódigo.

Algoritmo 2: Estimación del precio de continuación al tiempo t_j para cada trayectoria ω

entrada: Valores de todas las trayectorias al tiempo t_j , conjunto de m funciones base L_k , $k = 0, 2, \dots, m-1$, vector de precios de la opción Y_j .

salida: $\hat{F}(\omega, t_j)$

almacena los valores de cada trayectoria al tiempo t_j en el renglón S_j , es decir, $S_j(i) = S(t_j, \omega_i)$.

resuelve el sistema de ecuaciones $Y_j = \sum_{k=0}^{m-1} a_k L_k(S_j)$, usando regresión por mínimos cuadrados, para determinar a_k , $k = 0, 2, \dots, m-1$ (sólo las entradas no-cero de Y_j son consideradas aquí).

- **determina** SVD de la matriz $X_j [X_j(i, k) = L_k(S(t_j, \omega_i))]$, con representación $X_j = U\Sigma V^T$.
- **determina** a_k de U , V y Σ . $k = 1, 2, \dots, m-1$. $[a_k]_j = (V\Sigma^{-1}U^T)Y_j$.

regresa $\sum_{k=0}^{m-1} a_k L_k(S(t_j, \omega))$

Una optimización importante aplicada por Longstaff y Schwartz, es incluir únicamente aquellas trayectorias en la regresión para las cuales el precio de la opción es positivo o *in the money*. Esto incrementa de forma significativa la eficiencia del algoritmo y reduce el tiempo es su desarrollo computacional. Finalmente, Longstaff y Schwartz mostraron que el estimador $\hat{F}(\omega, t_k)$ converge en media cuadrada y en probabilidad a $F(\omega; t_k)$ cuando el número de trayectorias *in the money* tiende a infinito y, además, es el mejor estimador lineal insesgado basado en una métrica de media cuadrada.

5.3.3. Resultados de convergencia

El algoritmo LSM provee un simple y elegante procedimiento para encontrar la estrategia óptima de forma anticipada para una opción de estilo americano. Mientras que el último paso del algoritmo es aproximar el valor real de la opción usando un número realista de trayectorias y funciones base, es también práctico examinar qué puede decirse acerca de la convergencia teórica del algoritmo al verdadero valor $V(X)$ de la opción americana.

El primer resultado de convergencia se dirige a la parcialidad del algoritmo LSM y es aplicable incluso cuando la opción americana es continuamente ejercible.

Proposición 1. Para cualquier elección finita de M , K y del vector $\theta \in \mathbb{R}^{M \times (K-1)}$ representando los coeficientes para las M funciones base en cada una de las $K-1$ fechas de pronto ejercicio, sea $LSM^*(\omega; M, K)$ el flujo de efectivo descontado resultante, siguiendo la regla del algoritmo LSM de ejercer cuando el valor del ejercicio inmediato es positivo y más grande o igual a $\widehat{F}_M(\omega_i; t_k)$ tal y como se define por θ . Entonces la siguiente desigualdad es cierta:

$$V(X) \geq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N LSM^*(\omega_i; M, K) \quad (5.8)$$

Demostración. Ver el **Apéndice B** de la *Parte V*. ■

La intuición de este resultado es fácilmente entendible. El algoritmo LSM resulta en una regla de paro para una opción de estilo americano. El precio de ésta, sin embargo, está basado en una regla de paro que maximiza el valor de la opción; todas las demás, incluyendo la regla implementada por el algoritmo LSM, resultan en valores menores o iguales que el generado por la óptima regla de paro.

Este resultado es particularmente útil, debido a que provee un criterio objetivo para la convergencia. Por ejemplo, éste da pauta para determinar el número de funciones base necesarias para obtener una aproximación más precisa; simplemente incrementando M hasta que el valor obtenido por el algoritmo LSM ya no aumente. Esta propiedad útil e importante no es compartida por otros procedimientos, que simplemente descuentan funciones basadas en el valor de continuación estimado sin hacer comparación alguna con el valor del ejercicio inmediato.

Por esta naturaleza, es difícil proveer un resultado general de convergencia para el algoritmo LSM, debido a que es necesario considerar límites, como el número de puntos de discretización K , de funciones base M , y de trayectorias N cuando tienden a infinito. Además, se necesita considerar los efectos de propagación de la regla de paro hacia atrás, desde el punto t_{K-1} hasta t_1 . En el caso donde la opción americana sólo puede ser ejercida en $K=2$ puntos discretos en el tiempo, la convergencia del algoritmo es demostrada más fácilmente. Como un ejemplo, se tiene la siguiente proposición.

Proposición 2. Considérese que el valor de una opción americana depende de una única variable de estado X con soporte en $(0, \infty)$, la cual sigue un proceso de Markov. También, que la opción sólo puede ser ejercida en los tiempos t_1 y t_2 , y que la función de esperanza condicional $F(\omega; t_1)$ es absolutamente continua y

5.3. EL MÉTODO LSM DE LONGSTAFF-SCHWARTZ

$$\int_0^{\infty} e^{-X} F^2(\omega; t_1) dX < \infty,$$

$$\int_0^{\infty} e^{-X} F_X^2(\omega; t_1) dX < \infty$$

Entonces para cualquier $\varepsilon > 0$, existe un $M < \infty$ tal que

$$\lim_{N \rightarrow \infty} \Pr \left[\left| V(X) - \frac{1}{N} \sum_{i=1}^N LSM^*(\omega_i; M, K) \right| > \varepsilon \right] = 0$$

Demostración. Ver el **Apéndice B** de la *Parte V*. ■

Intuitivamente, este resultado significa que escogiendo M lo suficientemente grande, y haciendo $N \rightarrow \infty$, el algoritmo LSM resulta en un valor para la opción americana que difiere en ε de su valor real. En consecuencia, el algoritmo LSM converge a cualquier grado de precisión deseado debido a que ε es arbitrario. La clave para este resultado, es que la convergencia de $F_M(\omega; t_1)$ a $F(\omega; t_1)$ es uniforme sobre $(0, \infty)$ cuando las condiciones indicadas de integrabilidad son conocidas. Esto limita al error máximo cuando se estima la esperanza condicional que, a su vez, acota por arriba al error en el precio de la opción. Una importante implicación de este resultado, es que el número de funciones base necesario para obtener el nivel de precisión deseado, no necesita tender a infinito cuando $N \rightarrow \infty$. Mientras esta proposición esté limitada a una sola dimensión, se pueden hacer conjeturas sobre resultados similares que pueden ser obtenidos para problemas en dimensiones más grandes, encontrando las condiciones bajo las cuales la convergencia uniforme ocurre.

5.4. Ejemplos numéricos y resultados del algoritmo LSM

Considérese una Put americana que no paga dividendos sobre un bien subyacente, cuyo precio es gobernado por un proceso de movimiento browniano geométrico dado por

$$S_t = S_0 \exp(\sigma W_t + (r - \frac{\sigma^2}{2})t) \quad (5.9)$$

con r , como es usual, la tasa de interés libre de riesgo, σ denota la volatilidad constante; r y σ son constantes. S_0 representa el precio inicial del bien subyacente, mientras que W_t es el movimiento browniano. En este trabajo, como se señaló con anterioridad, las trayectorias del precio del bien subyacente son generadas mediante simulaciones

Monte Carlo. La forma en la que esto se realiza, es mostrada en los programas que constituyen los apéndices de este capítulo.

En una primera instancia, se considera un ejemplo numérico muy básico, explicado paso a paso, que ayudará a entender de la mejor forma las bases del algoritmo haciendo uso de las funciones (5.5) en la aproximación del valor de continuación. Después, se continua con el mismo ejemplo, pero esta vez considerando los primeros polinomios de Laguerre mostrados en (5.6). Se comentarán algunos resultados con base en estos dos ejemplos. Para finalizar, se desarrolla un ejemplo numérico mucho más real a los anteriores. El objetivo en todos los ejemplos, por supuesto, es aproximar el valor real $V(X)$ del precio de la opción. Todos los cálculos y desarrollos son ahora generados en *MATLAB*, buscando también la variedad de software en este trabajo.

5.4.1. ¿Por qué opciones Put americanas?

Como es usual, sean $S(t)$ el precio del bien subyacente al tiempo t y K el precio de ejercicio o precio *strike*. Considérese que el portador desea ejercer la opción en algún tiempo $t < T$. Esto vale la pena sólo si $S(t) > K$, otorgándole un *payoff* de $S(t) - K$ al tiempo t . En vez de eso, el portador podría hacer una venta en corto a precio de mercado al tiempo t , y entonces comprar el mismo bien al tiempo $t = T$. Con esta estrategia, el portador ha ganado un monto $S(t) - K$ al tiempo t y pagado un monto menor o igual a K al tiempo T . Esto es claramente mejor que ganar $S(t) - K$ al tiempo t . Debido a que nunca es óptimo ejercer una Call americana antes de la fecha de expiración, una opción Call americana debería tener el mismo valor que una Call europea. Esto no ocurre para las opciones Put. Por eso es importante analizar y valorar las opciones Put americanas.

5.4.2. Ejemplo con aproximación mediante monomios

Como se señaló antes, este primer ejemplo se hará de acuerdo al conjunto de ecuaciones base (5.5) para aproximar las funciones de continuación de la opción. Por simplicidad, se ilustra el algoritmo usando sólo ocho trayectorias para el precio del bien subyacente. La simulación de estas trayectorias, así como el resto del algoritmo, es mostrado en el programa del **Apéndice C.1** de este capítulo.

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

Considérese una opción Put americana cuyo precio es gobernado por el proceso (5.9). Se supone que la opción Put es ejercible a un precio *strike* de 1.10 en los tiempos 1, 2 y 3, donde el tiempo tres es la fecha de expiración. La tasa libre de riesgo es del 6%, mientras que la volatilidad es de 0.2. Se maneja un precio inicial de 1.0. Estos datos son introducidos en la primera parte del programa en el **Apéndice C.1**. La simulación de las trayectorias se muestra en la siguiente matriz de precios de mercado o precios *stock* $S(t)$.

Matriz de precios stock $S(t)$

Trayectoria	$t = 0$	$t = 1$	$t = 2$	$t = 3$
1	1.00	1.09	1.08	1.34
2	1.00	1.16	1.26	1.54
3	1.00	1.22	1.07	1.03
4	1.00	0.93	0.97	0.92
5	1.00	1.11	1.56	1.52
6	1.00	0.76	0.77	0.90
7	1.00	0.92	0.84	1.01
8	1.00	0.88	1.22	1.34

El primer objetivo es encontrar la *regla de paro* que maximiza el valor de la opción en cada tiempo a lo largo de cada trayectoria. Ya que el algoritmo es recursivo, se necesita generar primero un número de matrices intermedias de los flujos de efectivo. Condicionales sobre el no-ejercicio de la opción antes de la fecha final de expiración al tiempo tres, los flujos de efectivo o *payoffs* realizados por el portador de la opción siguiendo la estrategia óptima al tiempo tres son dados abajo. Evidentemente, se ejerce sólo si $S(3) > K$.

Matriz de los flujos de efectivo al tiempo $t = 3$

Trayectoria	$t = 1$	$t = 2$	$t = 3$
1	-	-	0.00
2	-	-	0.00
3	-	-	0.07
4	-	-	0.18
5	-	-	0.00
6	-	-	0.20
7	-	-	0.09
8	-	-	0.00

Estos *payoffs* son idénticos a aquellos que serían recibidos si la opción fuera europea en vez de americana.

PARTE V. APLICACIÓN

Si la Put está *in the money* en $t = 2$, entonces el portador de la opción debe decidir entre ejercer la opción inmediatamente, o continuar la vida de la opción hasta la fecha final de expiración en $t = 3$. De la matriz de precios *stock*, hay sólo cinco trayectorias para las cuales la opción se encuentra *in the money* en $t = 2$. Sea X el vector de los precios *stock* en $t = 2$ para estas cinco trayectorias, y sea Y el vector de sus correspondientes *payoffs* descontados con el factor de interés constante $r(t) = e^{-0.06} = 0.94176$, recibidos en $t = 3$ si la opción no fuese ejercida en $t = 2$. Como se señaló antes, utilizar sólo las trayectorias *in the money* incrementa de forma significativa la eficiencia del algoritmo en su intento por estimar las esperanzas condicionales. Los vectores X y Y están dados por las entradas no vacías de la siguiente matriz.

Regresión al tiempo $t = 2$

Trayectoria	Y	X
1	0.00×0.94176	1.08
2	-	-
3	0.07×0.94176	1.07
4	0.18×0.94176	0.97
5	-	-
6	0.20×0.94176	0.77
7	0.09×0.94176	0.84
8	-	-

Para estimar los flujos de efectivo esperados de continuar la vida de la opción, condicionados sobre el precio *stock* en $t = 2$, se aplica regresión en Y sobre las ecuaciones (5.5), es decir, $L_0 = 1$, $L_1 = X$ y $L_2 = X^2$. El vector de coeficientes B es encontrado por la regresión, y éstos son sustituidos en (4.4) para encontrar el estimador $\hat{F}(\omega; t) = E[Y | X]$ con $t = 2$ en este caso. La función de la esperanza condicional resultante en $t = 2$ está dada por $E[Y | X] = -1.070 + 2.9834X - 1.8136X^2$. La obtención de los coeficientes a_k se realiza por mínimos cuadrados, con base en el principio de la *descomposición del valor singular* (5.7). Estos coeficientes pueden ser verificados en el programa del **Apéndice C.1**.

El valor del ejercicio inmediato $E(\omega, t) = K - X$, definido como la diferencia entre el precio *strike* y el precio de mercado, es comparado con $E[Y | X]$. Por construcción, $E(\omega, t)$ será siempre positivo. Estos valores están dados en la siguiente matriz.

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

Decisión del ejercicio temprano óptimo en $t = 2$

Trayectoria	$E[Y X]$	$K - X$
1	0.02	0.0369
2	-	-
3	0.03	0.0461
4	0.13	0.1176
5	-	-
6	0.33	0.1520
7	0.26	0.1565
8	-	-

El valor de ejercicio inmediato es $1.10 - X$ para las trayectorias *in the money*, mientras que el valor de continuación es obtenido sustituyendo X en el estimador $E[Y | X]$. Esta comparación implica que es óptimo ejercer la opción en $t = 2$ para la cuarta, sexta y séptima trayectoria. Esto deja a la siguiente matriz, la cual muestra los flujos de efectivo recibidos por el portador de la opción condicionados al no-ejercicio antes de $t = 2$.

Matriz de los flujos de efectivo al tiempo $t = 2$

Trayectoria	$t = 1$	$t = 2$	$t = 3$
1	-	0.00	0.00
2	-	0.00	0.00
3	-	0.00	0.07
4	-	0.13	0.00
5	-	0.00	0.00
6	-	0.33	0.00
7	-	0.26	0.00
8	-	0.00	0.00

Observese que cuando la opción es ejercida en $t = 2$, el flujo de efectivo en la última columna se hace cero. Esto es porque una vez que la opción es ejercida, no hay más flujos de efectivo después, ya que el ejercicio sólo puede ocurrir una vez.

Procediendo de forma recursiva, se examina ahora lo que sucede en $t = 1$. De la matriz de precios *stock* $S(t)$, existen nuevamente cinco trayectorias donde la opción está *in the money* en $t = 1$. Para éstas, se define otra vez a Y como el valor descontado de los flujos de efectivo subsecuentes. Se aprecia que tomando Y para $t = 1$, se utilizan los flujos de efectivo actuales realizados a lo largo de cada trayectoria sin hacer uso del valor estimado de Y en $t = 2$.

PARTE V. APLICACIÓN

Debido a que la opción puede ser ejercida sólo una vez, futuros flujos de efectivo pueden ocurrir en $t = 2$ o en $t = 3$ pero no en ambos. Aquellos recibidos en $t = 3$ son descontados un período hasta $t = 2$ y ninguno es descontado dos períodos hasta $t = 1$. De forma similar, X representa el vector de precios *stock* en $t = 1$ para las trayectorias en donde la opción está *in the money*. Los vectores X y Y están dados por las entradas no vacías de la siguiente matriz.

Regresión al tiempo $t = 1$

Trayectoria	Y	X
1	0.00 x 0.94176	1.09
2	-	-
3	-	-
4	0.13 x 0.94176	0.93
5	-	-
6	0.33 x 0.94176	0.76
7	0.26 x 0.94176	0.92
8	0.00 x 0.94176	0.88

La función de la esperanza condicional en $t = 1$ es estimada nuevamente por regresión sobre una constante, X y X^2 . El estimador es $E[Y | X] = 2.0375 - 3.3354X + 1.3565X^2$. Los coeficientes pueden ser verificados en el mismo **Apéndice C.1**. Sustituyendo los precios simulados de X dentro de esta regresión, se obtienen los estimadores de los valores de continuación para $t = 1$. Estos valores estimados, así como los valores del ejercicio inmediato en $t = 1$, están dados en la siguiente matriz. La comparación de las dos columnas muestra que ejercer la opción en $t = 1$ es óptimo para la cuarta, sexta, séptima, y octava trayectoria.

Decisión del ejercicio temprano óptimo en $t = 1$

Trayectoria	$E[Y X]$	$K - X$
1	0.01	0.0139
2	-	-
3	-	-
4	0.17	0.1092
5	-	-
6	0.34	0.2866
7	0.18	0.1175
8	0.22	0.1533

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

Una vez identificada la estrategia de ejercicio en los tiempos 1, 2 y 3, *la regla de paro* puede ser ahora representada por la siguiente matriz, en donde los números uno denotan las fechas en las cuales la opción debe ser ejercida.

Regla de paro			
Trayectoria	$t = 1$	$t = 2$	$t = 3$
1	0	0	0
2	0	0	0
3	0	0	1
4	1	0	0
5	0	0	0
6	1	0	0
7	1	0	0
8	1	0	0

Con esta especificación de la regla de paro, es posible determinar los flujos de efectivo realizados siguiendo la regla, ejerciendo la opción en las fechas de ejercicio en donde existe un uno en la matriz anterior. Esto deja a la siguiente matriz con los flujos de efectivo de la opción en donde nuevamente, los valores pueden ser verificados en el programa del **Apéndice C.1**.

Matriz de los flujos de efectivo de la opción			
Trayectoria	$t = 1$	$t = 2$	$t = 3$
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.07
4	0.17	0.00	0.00
5	0.00	0.00	0.00
6	0.34	0.00	0.00
7	0.18	0.00	0.00
8	0.22	0.00	0.00

Una vez identificados los *payoffs* generados por la Put americana en cada fecha de ejercicio a lo largo de cada trayectoria, es posible obtener su precio descontando cada valor de la matriz anterior hasta $t = 0$, y obteniendo el promedio sobre todas las trayectorias como se sugiere en (5.8). De esta forma, el precio estimado de la opción Put americana es de 0.1144. Esto es aproximadamente dos veces el valor de 0.0564 si esta opción fuera una Put europea -resultado que se obtiene descontando únicamente el *payoff* generado en $t = 3$ -. A pesar de que resulta simple, este ejemplo ilustra cómo los

mínimos cuadrados pueden usar la información de las trayectorias simuladas para estimar la esperanza condicional. A su vez, esta función es utilizada para identificar la decisión de ejercicio que maximiza el valor de la opción en cada fecha a lo largo de cada trayectoria. Como lo muestra este ejemplo, el algoritmo LSM puede ser fácilmente implementado, debido a que sólo utiliza regresión lineal simple. Un segundo ejemplo sigue a continuación en donde se aprecia, por otro lado, la precisión con la que el algoritmo LSM se acerca sin preocuparse demasiado por el conjunto de funciones base para realizar esta aproximación.

5.4.3. Ejemplo con aproximación mediante polinomios de Laguerre

Considérese ahora la misma opción Put americana usada en el ejemplo anterior, siguiendo el proceso (5.9) y con los mismos parámetros. Para realizar una adecuada comparación entre este ejemplo y el anterior, es necesario conservar los datos simulados para los precios *stock* $S(t)$ a lo largo de cada trayectoria, es decir, se usará la misma matriz de precios simulados:

Matriz de precios stock $S(t)$

Trayectoria	$t = 0$	$t = 1$	$t = 2$	$t = 3$
1	1.00	1.09	1.08	1.34
2	1.00	1.16	1.26	1.54
3	1.00	1.22	1.07	1.03
4	1.00	0.93	0.97	0.92
5	1.00	1.11	1.56	1.52
6	1.00	0.76	0.77	0.90
7	1.00	0.92	0.84	1.01
8	1.00	0.88	1.22	1.34

Como se planteó antes, se usará ahora el conjunto de los primeros tres polinomios de Laguerre considerados en (5.6) para estimar los flujos de efectivo esperados, de continuar viva la opción. Debido a que se realiza exactamente el mismo procedimiento que en el ejemplo anterior, sólo se mostrarán las esperanzas condicionales y los resultados finales generados. Los coeficientes de los polinomios de Laguerre, así como los flujos de efectivo generados por la opción, pueden ser verificados en el programa del **Apéndice C.2.**

Por ejemplo, el estimador de la esperanza condicional en $t = 2$ es $E[Y | X] = -1.7137 + 4.2709(1 - X) - 3.6272(1 - 2X + X^2 / 2)$. Sustituyendo los valores de

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

X dentro de esta función, y comparándolos con los valores del ejercicio inmediato en $t=2$ -que son los mismos que en el ejemplo anterior por construcción-, se encuentra la decisión de ejercicio óptimo para $t=2$ obteniendo la siguiente matriz de flujos de efectivo:

Matriz de los flujos de efectivo al tiempo $t=2$

Trayectoria	$t=1$	$t=2$	$t=3$
1	-	0.00	0.00
2	-	0.00	0.00
3	-	0.00	0.07
4	-	0.13	0.00
5	-	0.00	0.00
6	-	0.33	0.00
7	-	0.26	0.00
8	-	0.00	0.00

De la misma forma, se verá ahora lo que sucede en $t=1$. Esta vez el estimador de la esperanza condicional es $E[Y|X]=1.4150-2.0904(1-X)+2.7129(1-2X+X^2/2)$. Sustituyendo los valores de X dentro de esta función, y comparándolos con los del ejercicio inmediato en $t=1$, se encuentra la decisión del ejercicio óptimo para $t=1$ obteniendo finalmente la matriz de flujos de efectivo de la opción para este ejemplo.

Matriz de los flujos de efectivo de la opción

Trayectoria	$t=1$	$t=2$	$t=3$
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.07
4	0.17	0.00	0.00
5	0.00	0.00	0.00
6	0.34	0.00	0.00
7	0.18	0.00	0.00
8	0.22	0.00	0.00

Como puede observarse, las dos últimas matrices son iguales a las correspondientes en el ejemplo anterior. Es decir, se obtienen los mismos flujos de efectivo, a pesar de que se utilizaron conjuntos diferentes de funciones base en cada ejemplo, monomios en una primera instancia, y polinomios de Laguerre en una segunda. ¿Por qué sucede esto? La respuesta es evidente cuando se comparan las funciones estimadas de continuación. Por ejemplo, para $t=2$, la función estimada por monomios es

PARTE V. APLICACIÓN

$E[Y | X] = -1.070 + 2.9834X - 1.8136X^2$, mientras que la estimada mediante polinomios de Laguerre es $E[Y | X] = -1.7137 + 4.2709(1 - X) - 3.6272(1 - 2X + X^2 / 2)$. Estas dos funciones son básicamente la misma cuando se simplifica la segunda obteniendo entonces $E[Y | X] = -1.070 + 2.9835X - 1.8136X^2$. Lo mismo ocurre cuando se comparan las funciones estimadas para $t = 1$.

Extensas pruebas numéricas como las anteriores, han demostrado que los resultados del algoritmo LSM son notablemente robustos a la elección de funciones base. Por ejemplo, se utilizaron los primeros tres monomios habituales como base en una primera instancia. Se obtienen resultados virtualmente idénticos al momento de estimar la esperanza condicional cuando se usan los tres primeros polinomios de Laguerre, y sucedería lo mismo si se eligen los primeros tres polinomios de Hermite, o tres funciones trigonométricas. Como se señaló antes, algunas bases aproximan de forma virtualmente idéntica la función de esperanza condicional sobre cualquier trayectoria de precios cuando es necesario conocer la fecha en donde un temprano ejercicio debería ser óptimo.

Mientras que los resultados son robustos a la elección de funciones base, es importante tener cuidado con las implicaciones numéricas de esta elección. Construcciones con expresiones exponenciales o logarítmicas podrían resultar en inconvenientes computacionales cuando se trabaje con un número muy grande de trayectorias o de fechas de ejercicio.

Finalmente, la elección de funciones base también tiene implicaciones para la significancia estadística dentro de una regresión. En particular, algunas elecciones de funciones base están altamente correlacionadas una con la otra, lo que provoca dificultades para estimar los coeficientes individuales de la regresión, algo parecido al problema de multicolinealidad en econometría. Esta dificultad no afecta al algoritmo LSM, debido a que éste se enfoca únicamente sobre el valor ajustado de la regresión, en vez de los coeficientes individuales. Este valor no es afectado por el grado de correlación entre las variables. Sin embargo, si la elección de funciones base origina matrices cercanamente singulares, las imprecisiones numéricas en algunos algoritmos por mínimos cuadrados pueden afectar de alguna forma la estimación de la función de la esperanza condicional. Para evitar este tipo de problemas numéricos, se recomienda el uso de software que utilice el algoritmo de doble precisión en sus cálculos -*MATLAB* es uno de ellos-.

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

5.4.4. Un ejemplo más real

Ahora se aplicará el algoritmo en un ejemplo con datos reales. Supóngase por un momento que un inversionista europeo desea adquirir una opción Put el 1 de agosto de 2012, con barriles de Petróleos Mexicanos (PEMEX) como bien subyacente. También, que el precio de cada barril está gobernado por el mismo proceso estocástico que se ha usado anteriormente:

$$S_t = S_0 \exp(\sigma W_t + (r - \frac{\sigma^2}{2})t) \quad (5.10)$$

La siguiente tabla muestra los precios promedio de exportación del crudo mexicano proporcionada por PEMEX:



indicadores petroleros

Precio Promedio de Exportación de Petróleo Crudo

(dólares por barril)

	Total	Istmo	Maya ^a	Olmeca	Por Región		
					América	Europa	Lejano Oriente ^b
2007	61.64	69.92	60.38	70.89	61.21	64.74	65.54
2008	84.38	81.09	82.92	99.37	85.30	81.18	65.49
2009	57.40	63.38	56.22	65.79	56.73	63.15	61.02
2010	72.46	78.63	70.65	79.58	72.29	73.20	74.55
2011	101.00	106.22	98.80	109.83	100.93	101.62	101.03
2012	103.91	109.22	101.82	111.39	104.47	100.84	103.49
Enero	108.54	110.47	107.46	113.23	107.84	112.65	108.38
Febrero	110.23	116.05	108.45	117.60	110.22	109.02	111.74
Marzo	112.82	120.31	109.60	122.46	114.51	108.34	101.18
Abril	108.04	116.68	104.89	118.46	111.13	95.43	102.04
Mayo	^R 102.24	106.84	^R 100.23	109.25	102.92	^R 98.45	0.00
Junio	90.07	93.61	88.39	96.12	90.02	91.24	87.88
Julio	94.04	98.51	92.25	100.96	94.38	92.02	0.00

a. Excluye crudo Altamira.

b. Incluye otras regiones.

Tabla 5.1. Tabla de Precios Promedio de Exportación de PEMEX.

Fuente: http://www.ri.pemex.com/files/dcpe/petro/epreciopromedio_esp.pdf

PARTE V. APLICACIÓN

Se supondrá que el precio *strike* K se establece en 100.84 dólares por barril, es decir, el promedio de exportación a Europa en lo que va del año 2012. La tasa libre de riesgo para esta inversión es fijada en 2% mensual, mientras que la volatilidad se estima que sea del 0.15. Ya que la opción Put se desea adquirir el 1 de agosto de 2012, se toma como precio inicial el promedio del mes de julio de 2012 para una exportación europea, es decir 92.02. El tiempo de maduración T es de un año. La simulación de las trayectorias de precios *stock* S se realizará cada mes, mientras que se simularán 50 trayectorias. Como ya se señaló antes, el número de fechas de ejercicio, así como el de trayectorias, pueden tender a infinito.

Se utilizan aquí polinomios de Laguerre, aunque como se discutió anteriormente, pueden usarse otro tipo de funciones base obteniendo resultados virtualmente idénticos. Todas las simulaciones de precios *stock*, así como los coeficientes estimados y flujos de efectivo resultantes, son mostrados en el programa del **Apéndice C.3**. Desde luego, los parámetros de la Put, así como la base ortonormal, pueden ser modificados sin problema dentro del software.

Después de una iteración del programa, el precio estimado de la opción Put americana a un año expedida por PEMEX es de 17.3456 dólares, con un error estándar de 2.2592. Esto contrasta con el valor estimado de 10.5471 dólares de una Put europea obtenido en la misma simulación cuando se descuentan y promedian los flujos de efectivo generados únicamente en la fecha de expiración.

Otro aspecto importante es revisar lo que sucede cuando se realizan varias iteraciones. Después de ejecutar 50 veces el programa, éstos fueron los resultados obtenidos para el precio de la opción:

5.4. EJEMPLOS NUMÉRICOS Y RESULTADOS DEL ALGORÍTMO LSM

Valor =	ErrorStd =	Valor =	ErrorStd =
17.3456	2.2592	17.3383	2.3220
17.5578	2.0387	16.2807	2.1662
18.4020	2.0511	17.2504	2.1154
18.8603	2.3124	18.0448	2.1723
16.9703	1.9604	17.3783	1.8731
16.0446	1.8342	17.5640	1.9336
18.7047	2.3763	16.6962	2.1008
17.6961	1.9269	17.1220	1.9901
16.5758	2.0271	16.2339	2.4343
16.0323	2.0762	19.1885	2.1460
16.2797	1.8689	17.0959	2.0989
17.8330	2.1034	20.6640	2.5175
16.9287	1.9194	19.6145	2.5335
17.3415	2.5135	18.2998	1.8040
16.8197	2.0604	17.4861	1.8914
18.3771	1.6172	18.6444	2.1810
16.6936	1.7272	17.2575	2.5635
19.2171	2.0976	19.0485	2.0724
18.3387	1.8364	16.0604	1.8913
17.8127	2.0209	18.6234	1.8839
17.3456	2.2592	17.6349	2.7510
20.0517	2.5396	16.2948	1.7512
16.4533	2.0482	17.4376	2.3392
17.5356	2.1533	20.1265	1.9109
18.9317	2.5054	17.7475	2.3319

Tabla 5.2. Tabla de los Precios obtenidos para el ejemplo de PEMEX.

Los valores anteriores arrojan un promedio de 17.7056 dólares en el precio de la Put y una desviación estándar de 2.1182. Estos resultados sugieren que el algoritmo LSM de Longstaff y Schwartz es sumamente apropiado para problemas de varias simulaciones -reflejando los resultados de convergencia aquí expuestos-, otorgando una simple y elegante forma de aproximar el precio de una opción americana tomando el número de fechas de ejercicio tan grande como se desee, en un intento por hacer que sea continuamente ejercible -como supone la realidad-.

CONCLUSIONES

En la *Parte I* se han analizado los diversos problemas que dieron origen al método de mínimos cuadrados y a otros problemas de la época que, aunque no eran exactamente de mínimos cuadrados, posibilitaron las soluciones que se plantearían después.

Es tradicional decir que Gauss y Legendre inventaron el método de mínimos cuadrados. Normalmente, se atribuye a personas concretas las invenciones concretas, pero no los métodos generales, que suelen ser resultado de la evolución histórica de los problemas y de las soluciones particulares que se han ido dando a cada uno. Sin embargo, el método de mínimos cuadrados se atribuye a los dos matemáticos mencionados, habiendo sido el procedimiento que ha posibilitado la resolución de una infinidad de problemas dispares desde su descubrimiento.

Desde este punto de vista, el trabajo de Gauss y Legendre es extraordinario pero no es el único. Es una situación muy semejante a la del descubrimiento del cálculo diferencial por parte de Newton y Leibniz. Evidentemente, su trabajo es enorme, pero su labor, como la de los anteriores, tiene el mérito de haber surgido de una síntesis y de una imaginación inmensa para conseguir unir todos los problemas en uno y dar una única solución a todos ellos.

En la *Parte II* se presentaron los fundamentos del modelo de regresión lineal tradicional haciendo uso del principio de mínimos cuadrados. Este método estadístico es una de las técnicas más utilizadas para el análisis de datos. Sus resultados son de gran utilidad, provenientes del proceso conceptual lógico de usar una ecuación para representar la relación entre una variable de interés (la respuesta) y una variable regresora. Pero su desarrollo no termina en este capítulo como se vio en la *Parte IV*.

El análisis de regresión es sumamente interesante desde un punto de vista teórico, ya que es una elegante herramienta matemática que cuenta con sólidos fundamentos estadísticos. Su éxito también se debe a su aplicación en problemas con datos reales. Aquí se presentó un ejemplo de entre una infinidad.

Por otra parte, se aprecia el papel importante que juega el software estadístico en el análisis de regresión. Hoy en día, existe toda una gama de paqueterías capaces de ajustar una línea de regresión mediante mínimos cuadrados, así como realizar un completo análisis desplegando toda la información del ajuste. De esta forma, se han integrado aquí

CONCLUSIONES

muchas características de R en particular, útiles para realizar de forma práctica y precisa un análisis de regresión simple con base en la teoría mostrada en este capítulo.

En la *Parte III* se presentó la teoría de un método alternativo de estimación de los parámetros del modelo de regresión lineal por mínimos cuadrados, así como un procedimiento para medir y comparar la precisión entre éste y el tradicional visto en la *Parte II*. Se puede concluir que el método alternativo de regresión lineal por mínimos cuadrados ha demostrado ser legítimo, y una opción válida en la búsqueda de un método más preciso, sobre todo, cuando se tienen datos con errores presentes en ambas variables como se muestra en los resultados.

El método de graficar aleatoriamente puntos alrededor de la línea real y aplicar los métodos de regresión descritos, en un intento por acercarlos a la línea real tanto como sea posible, es similar a una situación experimental donde una línea de regresión es determinada utilizando datos observacionales y comparada con la ecuación teórica.

La definición de la línea “más precisa”, como línea de regresión con área más pequeña entre la pareja de rectas, es el único método estudiado para realizar la comparación de precisión.

En la *Parte IV* se estudió la extensión del modelo de regresión lineal simple al múltiple, haciendo uso del principio de mínimos cuadrados. Utilizar más de una variable explicativa ofrece la ventaja de utilizar más información en la construcción del modelo y, consecuentemente, realizar estimaciones más precisas haciendo uso únicamente de combinaciones lineales entre ellas.

Aquí se identifica lo verdaderamente útil y práctico que resulta la implementación de programas informáticos. En este tema, el objetivo principal no es enseñar las operaciones detalladas del software, sino verificar la información de salida.

En la *Parte V* se presenta una técnica simple, pero importante, para aproximar el valor de opciones Put americanas mediante simulación y regresión por mínimos cuadrados. Esta aproximación es intuitiva, precisa, fácil de aplicar y computacionalmente eficiente. Se muestra esta técnica usando un par de ejemplos sencillos, pero muy ilustrativos que incluyen la valuación, así como la estrategia de ejercicio de una Put americana con ciertos parámetros. También se desarrolla un ejemplo con base en datos reales para la valuación de una Put cuando se desea invertir en Petróleos Mexicanos. Los resultados sugieren que el algoritmo LSM implementado por Longstaff y Schwartz es muy conveniente para problemas de varias simulaciones.

CONCLUSIONES

Una ventaja importante de esta técnica es su precisión. Resultados importantes de convergencia fueron mostrados, así como la flexibilidad en la elección de funciones base para la aproximación de las funciones de continuación. No hay que olvidar también las ventajas con respecto al software. Los programas aquí elaborados son fácilmente traducibles a diversos lenguajes de carácter matemático, tales como *Maple*, *R*, o cualquier otro que utilice algoritmos de doble precisión para sus cálculos; además, la sustitución de funciones base se realiza sin problemas dentro de los programas.

La valuación de opciones americanas sigue siendo una interesante área de investigación, particularmente cuando técnicas de regresión son utilizadas. Esto es debido a la flexibilidad de este método cuando se trata de resolver problemas de dimensión finita o de múltiples factores. La aparición de esta herramienta para la valuación de opciones americanas abre el camino a técnicas de simulación semejantes prometedoras, particularmente en el mercado de derivados.

BIBLIOGRAFÍA

Se utilizan diversas fuentes en la elaboración de este trabajo: algunos libros y escritos sobre historia matemática, libros clásicos de estadística, así como un número importante de artículos sobre mínimos cuadrados y sus aplicaciones obtenidos de Jstor (la biblioteca virtual estadounidense).

- [1] ALDRICH, J. (1998). *“Doing Least Squares: Perspectives from Gauss and Yule”*. International Statistical Review / Revue Internationale de Statistique, Vol. 66, No. 1, pp. 61-81.
- [2] CANAVOS, G. C. (1984). *“Applied probability and statistical methods”*. Little; Brown, Michigan.
- [3] CHOUDHURY, A. R. (2007). *“Optimizations in Financial Engineering: The Least-Squares Monte Carlo method of Longstaff and Schwartz”*. India Research Lab, Plot 4, Block-C, Institutional Area; New Delhi, India.
- [4] DELAUBENFELS, R. (2006). *“The Victory of Least Squares and Orthogonality in Statistics”*. The American Statistician, Vol. 60, No. 4, pp. 315-321.
- [5] EGLOFF, D. (2005). *“Monte Carlo Algorithms for Optimal Stopping and Statistical Learning”*. The Annals of Applied Probability, Vol. 15, No. 2, pp. 1396-1432.
- [6] HARTER, W. L. (1974). *“The Method of Least Squares and Some Alternatives: Part I”*. International Statistical Review / Revue Internationale de Statistique, Vol. 42, No. 2, pp. 147-174.
- [7] HOGG, R. V. (2006); *“Introduction To Mathematical Statistics”*. Pearson Education.
- [8] HORN, R. A.; JOHNSON, C. R. (1990). *“Matrix Analysis”*. Cambridge University Press.
- [9] JIA, Q. (2009). *“Pricing American Options using Monte Carlo Methods”*. Department of Mathematics, Uppsala University; Suecia.
- [10] KNIGHT, K. (2010). *“Mathematical Statistics”*. Chapman & Hall/Crc Texts in Statistical Science.

BIBLIOGRAFÍA

- [11] LONGSTAFF, F. A.; SCHWARTZ, E. S. (2001). *“Valuing American Options by Simulation: A Simple Least-Squares Approach”*. The Review of Financial Studies, Vol. 14, No. 1, pp. 113-147.
- [12] MAINDONALD, J. H; BRAUN W. J. (2003). *“Data Analysis and Graphics Using R: An Example-Based Approach”*. Cambridge University Press.
- [13] MARTIN, S.B. (1998). *“Less than the Least: An Alternative Method of Least Squares Linear Regression”*. McMurry University; Abilene, Texas.
- [14] MONTGOMERY, D. C.; PECK E. A.; VINING G. (2001). *“Introduction to Linear Regression Analysis”*. John Wiley & Sons, Inc., New Jersey.
- [15] MOOD, A. M.; GRAYBILL F. A. (1963). *“Introduction to the Theory of Statistics”*. McGraw-Hill.
- [16] PLACKETT, R. L. (1949). *“A Historical Note on the Method of Least Squares”*. Biometrika, Vol. 36, No. 3/4, pp. 458-460.
- [17] PLACKETT, R. L. (1972). *“Studies in the History of Probability and Statistics. XXIX: The Discovery of the Method of Least Squares”*. Biometrika, Vol. 59, No. 2. pp. 239-251.
- [18] STENTOFT, L. (2004). *“Convergence of the Least Squares Monte Carlo Approach to American Option Valuation”*. Management Science, Vol. 50, No. 9, pp. 1193-1203.
- [19] STIGLER, S. M. (1981). *“Gauss and the Invention of Least Squares”*. The Annals of Statistics, Vol. 9, No. 3, pp. 465-474.
- [20] STIGLER, S. M. (1986). *“The History of Statistics: The Measurement of Uncertainty Before 1900”*. Belknap, Press of Harvard University Press.
- [21] STIGLER, S. M. (2002). *“Statistics on the Table: The History of Statistical Concepts and Methods”*. Harvard University Press, pp. 320-331.
- [22] VERZANI, J. (2004). *“Using R for Introductory Statistics”*. Taylor & Francis.
- [23] WEINSBERG, S. (2005). *“Applied Linear Regression”*. John Wiley & Sons, Inc.; Hoboken, New Jersey.

APÉNDICES

Apéndice de la Parte I

A.1. Breve recuento sobre la controversia Gauss - Legendre

El método de mínimos cuadrados fue descubierto de forma independiente por Gauss y Legendre como se ha visto en la *Parte I*. A pesar de que es abordado por caminos distintos, de forma completamente analítica o haciendo uso de teoría probabilística, al final todo está basado en el mismo principio: minimizar la suma de los cuadrados de los errores. A pesar de que aparentemente Gauss había estado utilizando el método desde 1795, el primer tratado en donde se muestra de forma explícita fue publicado por Legendre en 1805. Cuatro años después, Gauss publicó su versión en la cual hace referencia a sus primeros trabajos. El fondo teórico de estas publicaciones ya ha sido abordado en este capítulo. Toca ahora revisar, a manera de complemento, la reacción de los involucrados así como el impacto que causó la controversia concerniente a la atribución del fabuloso invento. Las evidencias -fragmentos de cartas en su mayoría- son presentadas aquí.

Galle (astrónomo alemán nacido el 9 de junio de 1812) supone que las ideas básicas del método de mínimos cuadrados se le ocurrieron a Gauss en 1794 cuando preparaba sus estudios universitarios en el *Collegium Carolinum* en Brunswick (Alemania), leyendo un tratado sobre ecuaciones observacionales en el primer volumen de Lambert (1765) teniendo apenas diecisiete años. No hay evidencia de lo que Galle señala, sin embargo, el propio Gauss afirma que 1794 fue el año del descubrimiento. Éste llegó a Göttingen en octubre de 1795. Una lista de los libros que pidió prestados de la biblioteca de esta universidad fue recuperada por Dunnington (historiador norteamericano nacido el 15 de enero de 1906) en 1955, en donde se observa que Gauss revisó por mucho tiempo los tres volúmenes de Lambert. Era claro su interés en la teoría de ecuaciones observacionales, incluso, igual que en aquellos temas que desarrolló con creces.

El nombre de *mínimos cuadrados* -*les moindres quarrés* en francés- le fue asignado al método por parte de Legendre en su *Nouvelles méthodes pour la détermination des orbites des comètes* de 1805. Gauss no estaba consciente del trabajo de Legendre cuando

APÉNDICES

éste apareció. A continuación se muestran algunos fragmentos de cartas, una de Gauss a von Zach, y otra a su amigo, físico y astrónomo Olbers a quien Gauss le escribiría días después.

Gauss a von Zach, Brunswick, 8 de julio de 1806.

“... aún no he leído el trabajo de Legendre. Intencionalmente, he buscado evitarlo a fin de no intervenir con la sucesión de mis propias ideas mientras trabajo con mi propio método. Me encontré con un par de palabras de De la Lande en su reciente escrito las cuales son *moindres quarrés*. Tengo la impresión de que este principio, del cual he venido haciendo uso durante 20 años en muchos cálculos, y también, el que intento reafirmar en mi libro, no forma la parte esencial de mi análisis. Ha sido explicado de buena forma por Legendre...”

Gauss a Olbers, Brunswick, 30 de julio de 1806.

“... el Señor von Zach me escribió que te has ofrecido a revisar el trabajo de Legendre concerniente a las órbitas de los cometas. Te enviaré complacido a Bremen la obra que el señor von Zach me ha enviado; pero te pido por favor que me permitas tenerlo algunas semanas más. En una inspección preliminar, me parece que contiene muchas cosas agradables. Mucho de lo que es original en mi método, particularmente en su primera forma, lo encontré de nuevo en este libro. Parece ser mi destino el competir con Legendre en casi todos mis trabajos teóricos, como lo es en la alta aritmética, en la búsqueda de funciones trascendentales concernientes a la rectificación de la elipse, en los fundamentos de geometría, y ahora aquí de nuevo. En esta ocasión, el principio que he usado desde 1794, aquel en donde la suma de los cuadrados de los errores debe de ser minimizada para la adecuada representación de varias cantidades, las cuales no pueden ser representadas de forma exacta, es usado en el trabajo de Legendre y está más exhaustivamente desarrollado...”

Ocho meses después, Gauss le escribe nuevamente a Olbers en cuya carta hace referencia al mismo tema.

Gauss a Olbers, Brunswick, 24 de marzo de 1807.

“... me estoy ocupando ahora de tratar el problema: ‘Determinar los valores más probables de un número de incógnitas para un gran número de observaciones que dependen de ellas’, todo esto, haciendo uso del cálculo de probabilidades. El principio es que la suma de los cuadrados de las diferencias entre las cantidades teóricas y las observadas debe ser minimizada. Este principio lo he usado durante muchos años; te lo mencioné hace ya mucho tiempo, y ahora también está desarrollado por Legendre. En este contexto, el principio es preferible al de Laplace, en el cual la suma de las diferencias debe ser cero, y la suma de las mismas diferencias, cada una tomadas con signo positivo, debería ser un mínimo. Puedo demostrar que esto es inadmisibile en el cálculo de probabilidades, pero puede prestarse a contradicciones...”

Gauss terminó en 1806 una versión en alemán del libro que había estado preparando, pero encontró dificultades para publicarlo. Después de que lo tradujo al latín, el *Theoria Motus Corporum Coelestium* apareció hasta 1809. Una sección contenía una simple, pero concreta comparación del método de mínimos cuadrados con sus competidores en la cual, la idea de ‘inadmisible’ aparece brevemente. El siguiente texto es un extracto original del libro de Gauss:

“... por otro lado, el principio en que la suma de las diferencias entre las cantidades teóricas y observables debería ser mínima, tanto como fuese posible, podría ser considerado independientemente del *cálculo de probabilidades*.

Cuando el número de cantidades desconocidas es igual al número de cantidades observables que dependen de ellas, existe la forma de determinar sus valores de manera exacta sin problemas. Pero cuando el número de ecuaciones es más grande al de incógnitas, no puede ser determinado un procedimiento exacto; es entonces cuando las observaciones no cuentan con absoluta precisión. En este caso, se debe tener cuidado al establecer el mejor argumento para disminuir lo más posible los errores... denotando las diferencias entre las observaciones y los valores teóricos mediante Δ , Δ' , Δ'' , *etc.*, la condición se cumple no sólo si $\Delta\Delta + \Delta'\Delta' + \Delta''\Delta'' + \text{etc.}$ es tan pequeño como sea posible (lo cual es nuestro principio), sino también si $\Delta^4 + \Delta'^4 + \Delta''^4 + \text{etc.}$ o $\Delta^6 + \Delta'^6 + \Delta''^6 + \text{etc.}$, o en general, si la suma es de cualquier potencia con exponente par, la suma se vuelve tan pequeña como sea posible. Pero de entre todos estos principios, el nuestro es el más simple; esto debido a que en los otros caemos dentro de cálculos más complejos.

Por otro lado nuestro principio, del cual hacemos uso desde el año de 1795, a sido publicado más tarde por LEGENDRE en el tratado *Nouvelles méthodes pour la détermination des orbites des comètes* de 1805, donde han sido explicadas varias propiedades de este principio las cuales, por motivos de brevedad, omitimos aquí.

LAPLACE hace uso de otro principio para la solución de ecuaciones lineales cuando el número de ecuaciones es más grande que el número de cantidades desconocidas, el cual ha sido previamente propuesto por BOSCOVICH; llamado el *Método de Diferencias*, tomándolas todas ellas con signo positivo, deberían producir la suma tan pequeña como sea posible... por otro lado, LAPLACE mejora el principio, agregando una nueva condición: se requiere que en la suma de las diferencias, los signos permanezcan sin cambios y que la suma sea igual a cero...”

El *Theoria Motus* llegó a las manos de Legendre a través de Sophie Germain (matemática francesa nacida en París el 1 de abril de 1776). Siguió de inmediato una carta de Legendre hacia Gauss. Sólo algunos fragmentos fueron rescatados por Dunnington de los archivos personales que Gauss dejó en Göttingen. El interés de la carta es descubrir el papel que ésta juega en la relación entre Legendre y Gauss. Aquí se muestran algunos fragmentos traducidos de su francés original:

APÉNDICES

Legendre a Gauss. París, 31 de mayo de 1809.

“Me imagino Señor, que *Mademoiselle* Germain le ha hecho entrega del mensaje del cual ella estaba complacida de entregarle, en el cual le agradezco mucho a usted por el papel que me envió sobre su trabajo de series. Su escrito sobre un tema del cual yo he sido siempre aficionado despertó mi interés, y noté lo especial de su genio que le permitió demostrar las cuatro o cinco proposiciones a las cuales yo les di el nombre de *Ley de Reciprocidad* entre dos números primos.

Hace algunos días *Mademoiselle* Germain recibió de Alemania su *Theoria Motus Corporum Coelestium*; del cual me ha pasado algunos extractos que he tenido la oportunidad de leer. Puedo ver que este trabajo es digno de su reputación, y que usted ha reivindicado el análisis que algunos astrónomos han buscado para resolver el problema con exactitud sobre las órbitas de los planetas mediante observaciones.

En mis investigaciones para determinar las orbitas de los cometas, también tuve la intención de demostrar que este método, usado de forma adecuada, es capaz de proveer soluciones que son más exactas que aquellas obtenidas por otros métodos... pude ver en el transcurso de su exposición, que usted ha conseguido el mismo método, al que yo llamé *Méthode des moindres quarrés* en mi tratado sobre cometas. La idea de este método no responde a la llamada de un genio; sin embargo, cuando observo las dificultades e imprecisiones en los métodos que han sido implementados previamente, como aquel que Laplace justificó mediante Boscovich, coincido con usted en cuestionarlos... por otro lado, no le escondo a usted, Señor, que me siento decepcionado en ver que citando mi trabajo usted dice ‘*nuestro principio, del cual hacemos uso desde el año de 1795*’. No hay descubrimiento que no pueda ser reclamado por uno mismo diciendo que uno ha encontrado la misma cosa algunos años antes; pero si uno no suministra la evidencia citando el lugar donde lo ha publicado, esta aserción carece de sentido, y sólo sirve para desacreditar al verdadero autor del descubrimiento. En matemáticas, frecuentemente pasa que uno descubre las mismas cosas que habían sido descubiertas por otros y de las cuales ya se tenía conocimiento; esto me ha pasado en varias ocasiones, pero nunca lo he mencionado, y mucho menos llamado ‘*nuestro principio*’ a algo que ya ha sido previamente publicado por alguien más. Usted tiene tesoros suficientes, Señor, que no necesita de nadie más; y estoy perfectamente satisfecho, además, de la forma en que he expresado mi disgusto. Una vez tratado este punto, Señor, me permito señalarle que se ha equivocado en atribuirle algo a *Monsieur* Laplace que le pertenece a Euler, cuando usted menciona ‘por el elegante teorema descubierto por Laplace en donde $\int e^{-t^2} dt$ sobre $t = -\infty$ a $t = \infty$, $= \pi$ ’. Este teorema fue descubierto por Euler quien dio una

fórmula general para esta integral $\int dx \left(l \frac{1}{x} \right)^{\frac{1}{2}(2n+1)}$ de $x = 0$ a $x = 1$. Entonces, uno sólo pone

$e^{-t^2} = x$ para cambiar $\int e^{t^2}$ dentro de $\frac{1}{2} \int dx \left(l \frac{1}{x} \right)^{\frac{1}{2}}$.

Tengo el honor de ser, Señor, su obediente servidor.”

APÉNDICE DE LA PARTE I

Con respecto a las últimas líneas, De Moivre derivó la función *normal* en noviembre de 1733 usando una constante que casi pudo revelar: $\sqrt{\pi}$. Sin embargo, el reclamo hacia Gauss es fuerte en las líneas precedentes. Éste último escribió a Olbers poco después en el mismo año, buscando algún tipo de apoyo.

Gauss a Olbers. Göttingen, 4 de octubre de 1809.

“... ¿todavía recuerdas, querido amigo, que en una de mis primeras visitas a Bremen en 1803 hablé contigo sobre el principio del cual hice uso para representar las cantidades más exactas de las observaciones, en donde la suma de las diferencias debería ser minimizada cuando las observaciones tienen pesos iguales?... es importante para mí saber esto. Las razones del porqué pueden esperar...”

Laplace a Gauss. 15 de noviembre de 1811.

“Usted, *Monsieur* Gauss, dice en su trabajo sobre movimientos elípticos que ya estaba enterado de éste antes de que *Monsieur* Legendre lo publicara; estaré muy complacido de saber si antes de esta publicación, algún tratado concerniente a este método fue impreso en alemán, y le pido a usted, *Monsieur* Gauss, tener la amabilidad de informarme al respecto.”

Después de esto, Gauss expuso el caso a Olbers y a Laplace en dos cartas por separado en poco menos de una semana. En la primera, de nuevo pide apoyo a Olbers pero continua sin decirle las razones. En la segunda, abre con un complejo problema de métrica en *teoría de números* -que aquí se omite-, continuando con el asunto de los mínimos cuadrados.

Gauss a Olbers. Göttingen, 24 de enero de 1812.

“... ya te había mostrado las ideas esenciales en nuestro encuentro personal de 1803. Recuerdo entre mis papeles que datan de 1798, cuando el método era uno el cual ya había aplicado durante mucho tiempo, mi apreciación de que el método de Laplace era incompatible con los principios del *cálculo de probabilidades* en un diario donde escribo sobre mis ocupaciones matemáticas. En el otoño de 1802, escribí sobre este diario el conjunto de ocho estimadores en el problema de Ceres, encontrados por el método de mínimos cuadrados. Estos papeles, en los que apliqué el método en años anteriores, se encuentran ahora perdidos...”

Gauss a Laplace. Göttingen, 30 de enero de 1812.

“... he usado el método de mínimos cuadrados desde 1795, y encontré en mis papeles, que en el mes de junio de 1798, lo reivindicé con los principios del *cálculo de probabilidades* en un diario que había conservado desde el año de 1796, y que mostré a tiempo al Señor De Lindenau.

A pesar de mis frecuentes aplicaciones, fue a partir de 1802 que lo usé cada día en mis cálculos astronómicos sobre nuevos planetas. Después, intenté realizar un trabajo extenso que comencé

APÉNDICES

en 1805, y cuyo manuscrito original en alemán fue completado en 1806 pero publicado en latín hasta 1809. No estaba apurado en la publicación de algún fragmento de este trabajo, cuando Legendre se me adelantó. A pesar de ello, ya había platicado sobre el método con algunos amigos como Olbers, que seguramente lo recordará. Además, en mi *Theoria Motus Corporum Coelestium*, fui capaz de desarrollar la teoría concerniente al método sin haber revisado el trabajo de Legendre. ¿Podría yo discutir este principio, del cual informé a varios amigos en años anteriores, como parte de un trabajo que estaba preparando? ¿O como un método derivado del trabajo del Señor Legendre? No tenía idea de que el Señor Legendre fuera capaz de darle mucha importancia a una idea tan simple como esa... se debió sentir molesto conmigo cuando dije que lo había usado antes de que él lo hiciera. De hecho, tendría que ser fácil demostrar esto a todos con evidencias que no podrán ser refutadas. Pero pienso que todo aquel que me conoce debería creerme sólo por mi palabra, tal y como yo lo hubiera hecho con todo mi corazón si el Señor Legendre hubiera afirmado que él ya había conversado sobre el método antes de 1795. Tengo muchas cosas en mis papeles, de las cuales quizá pierdo la oportunidad de publicarlas por primera vez; pero sea lo que sea, prefiero dejar a que las cosas maduren..."

Olbers de inmediato estuvo de acuerdo con apoyar a Gauss, y cuatro años más tarde cumplió su promesa agregando un pie de página a uno de sus escritos sobre el período de las estrellas (Olbers, 1816, p.192).

Olbers a Gauss. Bremen, 10 de marzo de 1812.

"... acabo de recibir la edición de noviembre publicada por M.C., con su hermoso método de eliminación por mínimos cuadrados. Puedo atestiguar públicamente en la primera oportunidad, y lo haré con placer, que usted ya me había contado el principio básico en 1803. Esto lo recuerdo bastante como si hubiera pasado hoy. También debería haber algo escrito concerniente a esto en mis papeles. Además de lo señalado, también me informó sobre su fórmula de interpolación aquella vez..."

Una carta a Olbers en 1819 describe el progreso del trabajo que libera al método de mínimos cuadrados de suposiciones de normalidad y culmina con las líneas presentadas a la Real Sociedad de Göttingen en 1821, 1823 y 1826.

Gauss a Olbers. Göttingen, 22 de febrero de 1819.

"... estoy muy ocupado ahora con nuevas bases para el tan aclamado método de mínimos cuadrados. En mis primeras bases supuse que la probabilidad de un error observacional x estaba dada por e^{-hxx} , en el caso de que el método ofrezca el resultado más probable con completo rigor en todos los casos. Cuando la distribución de los errores es desconocida, es imposible establecer el resultado más probable a partir de observaciones ya realizadas. Laplace ha considerado el problema desde un ángulo diferente escogiendo un principio, el cual deja al método de mínimos cuadrados, y que es muy independiente de la distribución de los errores cuando el número de observaciones es indefinidamente grande.

APÉNDICE DE LA PARTE I

Con un número moderado de observaciones, sin embargo, uno se queda en la oscuridad si la distribución de los errores es desconocida; Laplace también notó esto diciendo que el método de mínimos cuadrados también sería aplicable en este caso, ya que ofrece un cálculo conveniente. Acabo de encontrar ahora que, escogiendo un principio algo diferente al de Laplace (y de hecho, uno de tal manera cuya hipótesis puede ser justificada al menos tan bien como la de Laplace, y en mi opinión, de una forma más natural), todas las ventajas permanecen, encontrando que en todos los casos para la distribución de los errores, el método de mínimos cuadrados será el más ventajoso, y la comparación de precisión de los resultados con respecto a las observaciones, en la cual me basé en mi *Theoria* usando la curva e^{-hxx} para los errores, continua siendo generalmente válida. Al mismo tiempo existe la ventaja de que todo puede ser demostrado y elaborado de manera clara, simple, analítica...”

En 1827, Legendre recibió una carta de Jacobi sobre el tema de funciones elípticas. Legendre tenía 75 años mientras que Jacobi 23. A pesar de la disparidad de edades, una correspondencia durante cinco años fue establecida. Aquí se encuentra un breve extracto de la primera carta, seguida por la parte final de la respuesta que Legendre hizo a esta carta.

Jacobi a Legendre. Königsberg de Prusia, 5 de agosto de 1827.

“... es apenas reciente cuando estas investigaciones han tomado forma. Sin embargo, no son las únicas investigaciones en Alemania con el mismo objetivo. El Señor Gauss, comenzó a informar sobre ellas, y me contó que él ya había desarrollado el método en 1808 y descubierto al mismo tiempo los nuevos alcances de éste. Me parece que esta noticia es muy interesante...”

Legendre a Jacobi. París, 30 de noviembre de 1827.

“... como puede *Monsieur* Gauss atreverse a decir que parte de mis teoremas eran conocidos por él y que los había descubierto antes de 1808?... Esta extrema impertinencia es increíble de un hombre que tiene el mérito suficiente para no tener que apropiarse de los descubrimientos de otros... pero es el mismo hombre quien, en 1801, quiso atribuirse el descubrimiento de *La ley de reciprocidad* publicada en 1785, y quien quiso apropiarse en 1809 del método de mínimos cuadrados publicado en 1805. Podría dar otros ejemplos, pero un hombre de honor debería abstenerse de ello...”

Mientras tanto, la controversia originada por la publicación de la *Theoria Motus* continuaba resonando incluso fuera de Alemania.

Schumacher a Gauss. Altona, 30 de noviembre de 1831.

“... creo haberle dicho que Zach publicó una carta suya en la edición de *G.E.* de octubre de 1799, en la cual usted evidentemente menciona al método de mínimos cuadrados, método que ya le había contado a Zach para ese entonces. Por la forma en la que obtuvo los resultados en uno de

APÉNDICES

sus estudios, me parece fácil demostrar que éstos fueron derivados por el método de mínimos cuadrados. Además, Zach sigue vivo, seguramente conservó su carta. ¿No le parece que podría despejar de una vez cualquier duda, incluso ante los reclamos del francés, los cuales encuentro particularmente ofensivos?”

Gauss a Schumacher. Göttingen, 3 de diciembre de 1831.

“... La carta que tú comentas es bien conocida por mí; ahí, la aplicación del método de mínimos cuadrados es sobre un extracto de la tabla de Ulugh Beingh de ecuaciones de tiempo, impresas antes en el mismo diario, a la cual le he dado varios resultados curiosos. Le hablé de estos resultados a Zach remarcando la conexión que existe entre éstos. Utilicé un método propio que he usado por años, de combinar cantidades involucrando errores aleatorios de una forma consistente de libre aleatoriedad. Sin embargo, no le informé a él sobre la naturaleza del método. Creo que le he escrito a usted antes, de que en ningún lado escribí esto... debería ser casi reconocido lo que digo en mi *Theoria Motus Corporum Coelestium*, que el método lo he usado desde 1794 sin necesidad de justificación, pero cuando Olbers afirmó, de que le hablé sobre el método en 1802, su palabra significó mucho...”

Después de 1827, Gauss no publicó nada más sobre las bases teóricas de los mínimos cuadrados, pero recordó brevemente su función para la distribución de los errores en una carta a Bessel.

Gauss a Bessel. Göttingen, 28 de febrero de 1839.

“... he leído con gran interés su trabajos en astronomía, y sobre como la función de probabilidad para los errores de las observaciones surge simultáneamente de distintas fuentes, aproximándose a la fórmula e^{-xxhh} ... esta forma me ha sido familiar por muchos años, aunque nunca llegué a desarrollarla por completo...”

La última carta en la cual se refiere al descubrimiento del método de mínimos cuadrados tiene esencialmente el mismo tema que la carta de Gauss a Laplace de 1812; pero esta vez había leído los papeles de Tobías Mayer y fue llevado poco a poco a una reconsideración del asunto.

Gauss a Schumacher. Göttingen, 6 de julio de 1840.

“... de entre otros papeles de Mayer que están ahora en mi posesión, puedo ver que al menos un cuarto de hojas es un borrador que contiene un pequeño número de pasajes que son cálculos numéricos comparando su teoría con observaciones. Puedo recordar cuatro o cinco de esos pasajes.

Estas notas en borrador, de paso, me han dejado algo de desacuerdo con respecto a otro asunto. Usted sabe que nunca he hecho gran historia de mí, sino hasta los procedimientos que he usado desde 1794, y a los que les di subsecuentemente el nombre de mínimos cuadrados. Por favor, no

APÉNDICE DE LA PARTE I

me malinterprete, no me refiero a los beneficios que estos ofrecen, esto es lo suficientemente claro; sino a la forma en que medí el valor de las cosas. Desde el principio la idea me pareció muy natural, muy accesible, que no dudé en lo mínimo que mucha gente, quien tuvo que enfrentarse a cálculos numéricos, por su propia voluntad han llegado a un procedimiento, usarlo, y sin pensar hacer mucho alboroto sobre una cosa tan natural. Para ser preciso, tuve a Tobías Mayer en mente de manera particular, y recuerdo claramente que, cuando solía discutir el método con otra persona (como, por ejemplo, pasó frecuentemente durante mis días de estudiante, 1795-1798), a menudo expresé que debería apostar cien a uno que Tobías Mayer ya había usado el mismo método en sus cálculos. Ahora que tengo estos papeles, yo hubiese perdido la apuesta. En el caso en el que considera eliminación, por ejemplo, para tres cantidades desconocidas en cuatro o cinco ecuaciones, lo hizo de tal manera que el más complaciente de los matemáticos pudo haberlo hecho sin ningún rasgo del más refinado estilo.

En todo caso, le digo esto en confidencia. Me siento realmente apenado al dar una opinión un poco baja sobre Mayer, pero detesto involucrarme en asuntos patrióticos...”

Gauss murió en 1855 y en 1856 su amigo Sartorius publicó su biografía, la cual ha sido una importante fuente de información para otros biógrafos. La sección que se refiere al método de mínimos cuadrados (p. 42-3) confirma muchos de los puntos antes mencionados, terminando con una cita del propio Gauss poco antes de morir:

“... El método de mínimos cuadrados no es uno de mis más grandes descubrimientos, pero todos aquellos que dudaron, debieron haber creído en mí.”

Apéndice de la Parte III

A.1. Programa del modelo general de la línea de regresión por el método alternativo y tradicional sobre Maple

Se introducen los datos:

```
[ > X:=[x1,x2,x3,x4];
[ > Y:=[y1,y2,y3,y4];
```

Proceso para el cálculo de la línea alternativa:

```
[ > n:=nops(X);
[ > media:=proc(X::list)
[ > local n, i, total;
[ > n:=nops(X);
[ > if n=0 then ERROR(`lista vacía`) fi;
[ > total:=add(i,i=X);
[ > total/n;
[ > end:
 $\bar{x}$  [ > medx:=media(X);
 $\bar{y}$  [ > medy:=media(Y);
 $\sum_{i=1}^n x_i$  [ > sx:=add(i,i=X);
 $\sum_{i=1}^n y_i$  [ > sy:=add(i,i=Y);
 $\sum_{i=1}^n x_i y_i$  [ > b:=medy-m*medx;
[ > sxy:=add(X[i]*Y[i], i=1..n);
 $\sum_{i=1}^n X_i^2$  [ > sx2:=add((X[i]-medx)^2, i=1..n);
 $\sum_{i=1}^n Y_i^2$  [ > sy2:=add((Y[i]-medy)^2, i=1..n);
 $\sum_{i=1}^n X_i Y_i$  [ > sxy:=add((X[i]-medx)*(Y[i]-medy), i=1..n);
 $m_+$  [ > ms:=evalf(sy2-sx2+sqrt((sx2-sy2)^2+4*sxy^2))/(2*sxy);
 $m_-$  [ > mr:=evalf(sy2-sx2-sqrt((sx2-sy2)^2+4*sxy^2))/(2*sxy);
[ > mp:=`if`(ms>0,ms,mr);mn:=`if`(mr<0,mr,ms);
[ > m:=`if`(n*sxy>sx*sy,mp,mn);
[ > S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
[ > b:=evalf(medy-m*medx);
```

Proceso para el cálculo de la línea tradicional:

$$\sum_{i=1}^n x_i^2$$

```
[> sx2:=add((X[i])^2,i=1..n);
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
> bt:=evalf(medy-mt*medx);
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n));
> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S;
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St;
```

Gráfica:

```
[> with(plots):
> p:=plot(linea_alt,x=0..10,y=0..10,color=blue):
> r:=plot(linea_trad,x=0..10,y=0..10,color=red):
> q:=plot([[x1,y1],[x2,y2],[x3,y4],[x4,y4]],style=POINT,
color=black,title=`Titulo`):
> display(p,r,q);
```

A.2. Ejemplo numérico 1

```

> X:=[1,2,4,5];
X:= [1, 2, 4, 5]
> Y:=[3,4,2,5];
Y:= [3, 4, 2, 5]
> n:=nops(X);
n:= 4
> media:=proc(X::list)
> local n, i, total;
> n:=nops(X);
> if n=0 then ERROR(`lista vacía`) fi;
> total:=add(i,i=X);
> total/n;
> end;
> medx:=media(X);
medx:= 3
> medy:=media(Y);
medy:=  $\frac{7}{2}$ 
> sx:=add(i,i=X);
sx:= 12
> sy:=add(i,i=Y);
sy:= 14
> b:=medy-m*medx;
> sxy:=add(X[i]*Y[i], i=1..n);
sxy:= 44
> sX2:=add((X[i]-medx)^2, i=1..n);
sX2:= 10
> sY2:=add((Y[i]-medy)^2, i=1..n);
sY2:= 5
> sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n);
sXY:= 2
> ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ms:= 0.3507810592
> mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
mr:= -2.850781060
> mp:=`if`(ms>0,ms,mr);mn:=`if`(mr<0,mr,ms);
mp:= 0.3507810592
mn:= -2.850781060

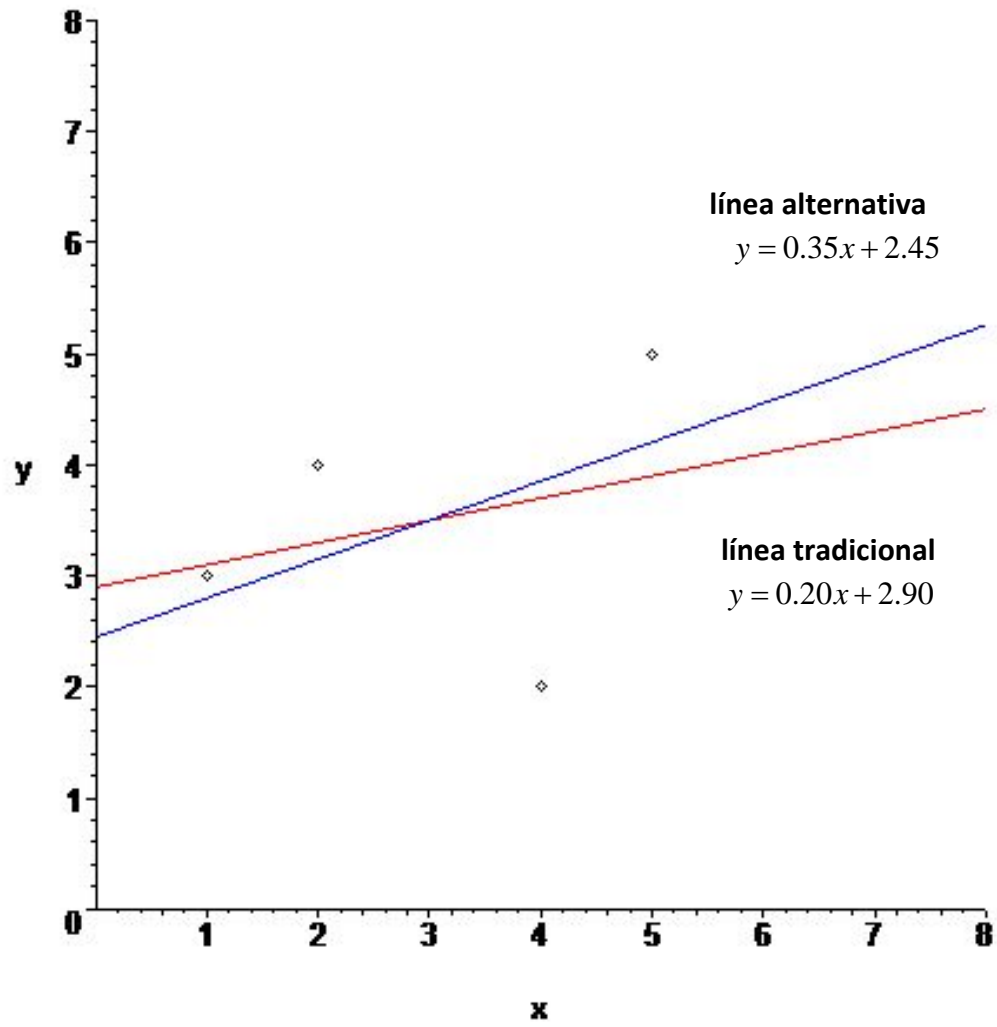
```

```

> m:=`if`(n*sxy>sx*sy,mp,mn);
           m := 0.3507810592
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
           S := 4.298437881
> b:=evalf(medy-m*medx);
           b := 2.447656822
> sx2:=add((X[i])^2,i=1..n);
           sx2 := 46
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
           mt := 0.2000000000
> bt:=evalf(medy-mt*medx);
           bt := 2.9000000000
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n));
           St := 4.6000000000
> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S;
           linea_alt := 0.3507810592 x + 2.447656822
           error_alt := 4.298437881
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St;
           linea_trad := 0.2000000000 x + 2.9000000000
           error_trad := 4.6000000000
> with(plots):
> p:=plot(linea_alt,x=0..8,y=0..8,color=blue):
> r:=plot(linea_trad,x=0..8,y=0..8,color=red):
> q:=plot([[1,3],[2,4],[4,2],[5,5]],style=POINT,color=black,
           title=`EJEMPLO 1`):
> display(p,r,q);

```


EJEMPLO 1



A.3. Ejemplo numérico 2

```

> X:=[0,1,2,3,4,5];
X:=[0,1,2,3,4,5]
> Y:=[6,4,3,4,2,1];
Y:=[6,4,3,4,2,1]
> n:=nops(X);
n:=6
> media:=proc(X::list)
> local n, i, total;
> n:=nops(X);
> if n=0 then ERROR(`lista vacía`) fi;
> total:=add(i,i=X);
> total/n;
> end:
> medx:=media(X);
medx:= $\frac{5}{2}$ 
> medy:=media(Y);
medy:= $\frac{10}{3}$ 
> sx:=add(i,i=X);
sx:=15
> sy:=add(i,i=Y);
sy:=20
> b:=medy-m*medx:
> sxy:=add(X[i]*Y[i], i=1..n);
sxy:=35
> sX2:=add((X[i]-medx)^2, i=1..n);
sX2:= $\frac{35}{2}$ 
> sY2:=add((Y[i]-medy)^2, i=1..n);
sY2:= $\frac{46}{3}$ 
> sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n);
sXY:=-15
> ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
ms:=-0.9303824105
> mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY);
mr:=1.074826856
> mp:='if'(ms>0,ms,mr);mn:='if'(mr<0,mr,ms);
mp:=1.074826856
mn:=-0.9303824105

```

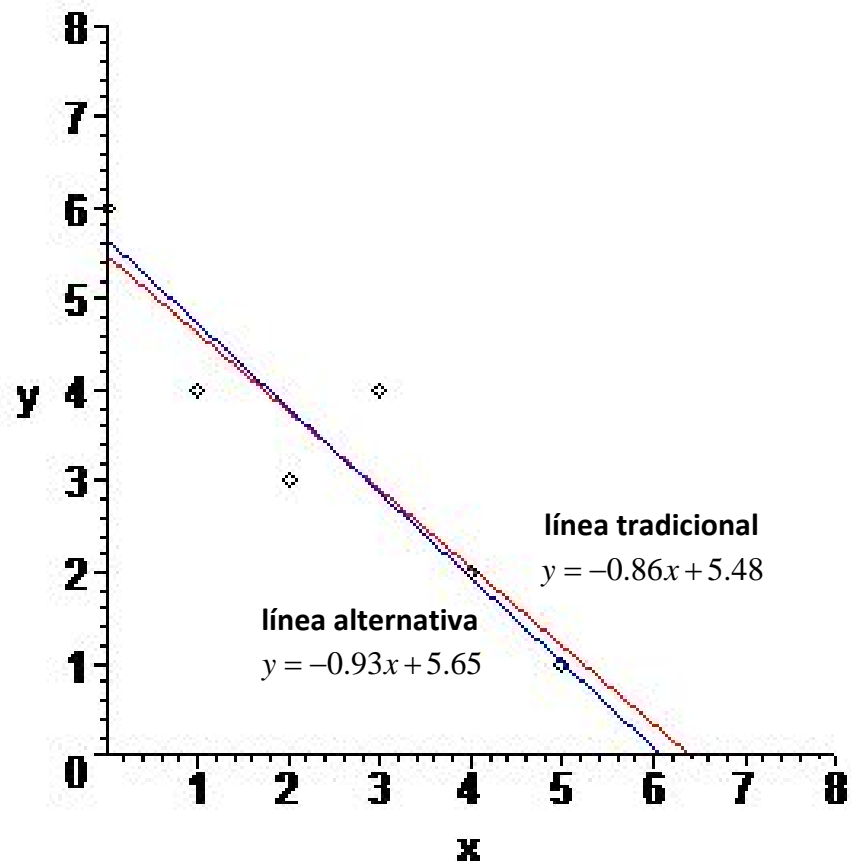
APÉNDICES

```

> m:='if`(n*sxy>sx*sy,mp,mn);
                                     m := -0.9303824105
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));
                                     S := 1.377597178
> b:=evalf(medy-m*medx);
                                     b := 5.659289359
> sx2:=add((X[i])^2,i=1..n);
                                     sx2 := 55
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));
                                     mt := -0.8571428571
> bt:=evalf(medy-mt*medx);
                                     bt := 5.476190476
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n));
                                     St := 2.476190476
> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S;
                                     linea_alt := -0.9303824105 x + 5.659289359
                                     error_alt := 1.377597178
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St;
                                     linea_trad := -0.8571428571 x + 5.476190476
                                     error_trad := 2.476190476
> with(plots):
> p:=plot(linea_alt,x=0..8,y=0..8,color=blue):
> r:=plot(linea_trad,x=0..8,y=0..8,color=red):
> q:=plot([[0,6],[1,4],[2,3],[3,4],[4,2],[5,1]],style=POINT,
          color=black,title='EJEMPLO 2'):
> display(p,r,q);

```

EJEMPLO 2



B.1. Comparación de precisión para datos con error presente en ambas variables

```
[> err:=(rand(0..10))/20:
[> m:=rand(1..30)/10:
[> b:=rand(0..10)/3:
[> yreal:=m()*x+b();


$$y_{real} = \frac{29x}{10} + 2$$

[> X:=[1+(err()*(-1)^(rand()), 2+(err()*(-1)^(rand()), 3+(err()*(-1)^(rand()), 4+(err()*(-1)^(rand()), 5+(err()*(-1)^(rand()))];


$$X = \left[ \frac{11}{20}, \frac{9}{4}, \frac{63}{20}, \frac{9}{2}, \frac{93}{20} \right]$$

[> Y:=[subs(x=1,yreal)+(err()*(-1)^(rand()), subs(x=2,yreal)+(err()*(-1)^(rand()), subs(x=3,yreal)+(err()*(-1)^(rand()), subs(x=4,yreal)+(err()*(-1)^(rand()), subs(x=5,yreal)+(err()*(-1)^(rand()))];


$$Y = \left[ \frac{22}{5}, \frac{159}{20}, \frac{53}{5}, \frac{273}{20}, \frac{323}{20} \right]$$

[> n:=nops(X);


$$n = 5$$

[> sx:=add(i,i=X):
[> medx:=sx/n:
[> sy:=add(i,i=Y):
[> medy:=sy/n:
[> b:=medy-m*medx:
[> sxy:=add(X[i]*Y[i], i=1..n):
[> sX2:=add((X[i]-medx)^2, i=1..n):
[> sY2:=add((Y[i]-medy)^2, i=1..n):
[> sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n):
[> ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
[> mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
[> mp:=`if`(ms>0,ms,mr):mn:=`if`(mr<0,mr,ms):
[> m:=`if`(n*sxy>sx*sy,mp,mn);


$$m = 2.756360272$$

[> b:=evalf(medy-m*medx);


$$b = 2.225791979$$

[> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1), i=1..n));


$$S = 0.3421221939$$

[> sx2:=add((X[i])^2, i=1..n):
[> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));


$$mt = 2.674770722$$

[> bt:=evalf(medy-mt*medx);


$$bt = 2.472192420$$

[> St:=evalf(add((Y[i]-mt*X[i]-bt)^2, i=1..n));


$$St = 2.864463137$$

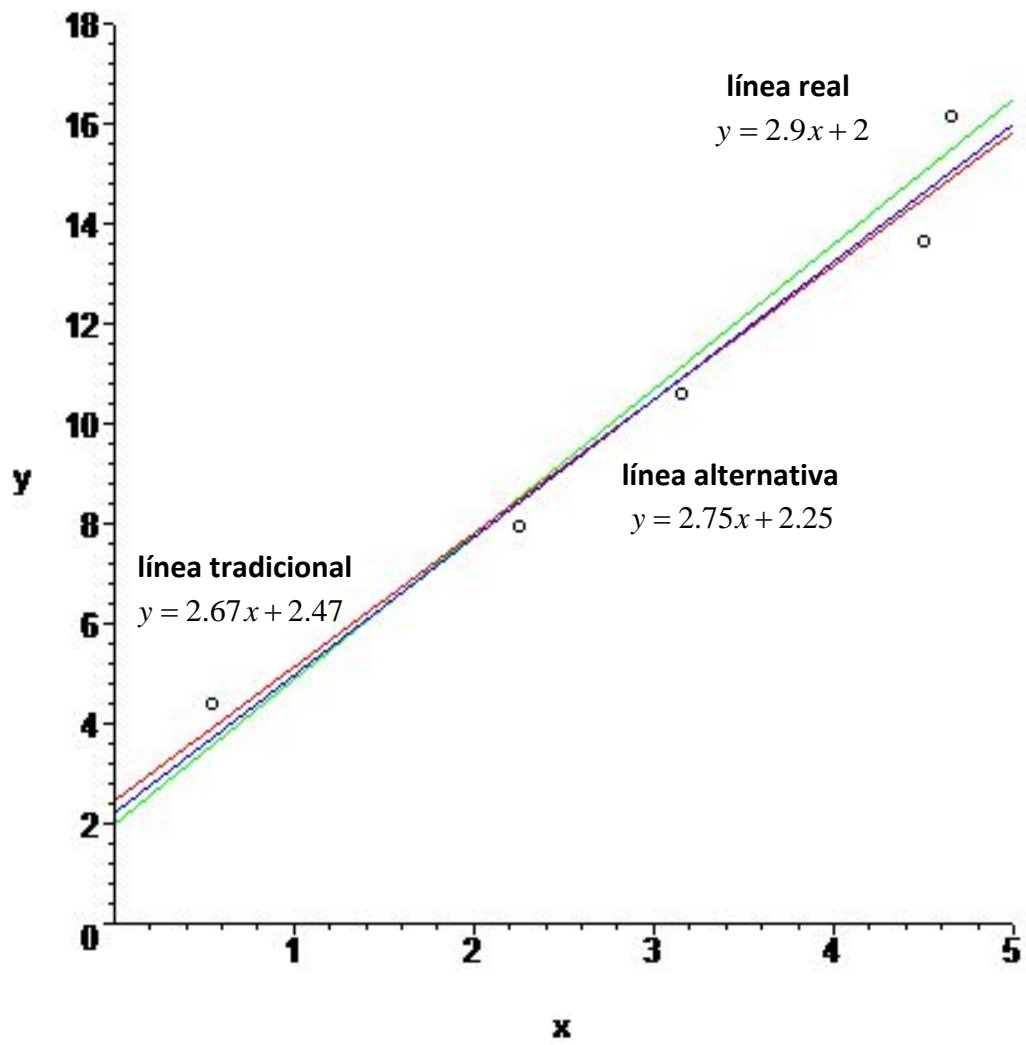
```

```

> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S;
      linea_alt := 2.756360272 x + 2.225791979
      error_alt := 0.3421221939
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St;
      linea_trad := 2.674770722 x + 2.472192420
      error_trad := 2.864463137
> errorAlt2:=int((linea_alt-yreal)^2,x=0..5);
      errorAlt2 := 0.3037747719
> errorTrad2:=int((linea_trad-yreal)^2,x=0..5);
      errorTrad2 := 0.5697156145
> errorAlt:=int(abs(linea_alt-yreal),x=0..5);
      errorAlt := 1.021466490
> errorTrad:=int(abs(linea_trad-yreal),x=0..5);
      errorTrad := 1.444353687
> mas_preciso:=`if`(errorAlt<errorTrad,alternativo,tradicional);
      mas_preciso := alternativo
> with(plots):
> a:=plot(linea_alt,x=0..5,y=0..18,color=blue):
> t:=plot(linea_trad,x=0..5,y=0..18,color=red):
> r:=plot(yreal,x=0..5,y=0..18,color=green):
> q:=plot(X,Y,x=0..5,y=0..18,color=black,style=point,symbol=circle,
      color=black,title=`EJEMPLO`):
> display(a,t,r,q);

```

EJEMPLO



B.2. Comparación de precisión para datos con error presente en ambas variables

(Versión de iteración múltiple)

```
[> numA:=0:numT:=0:
> for i from 1 to 100 do
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> yreal:=m()*x+b():
> X:=[1+(err())*(-1)^(rand()),2+(err())*(-1)^(rand()),3+(err())*(-1)
  ^ (rand()),4+(err())*(-1)^(rand()),5+(err())*(-1)^(rand())]:
> Y:=[subs(x=1,yreal)+(err())*(-1)^(rand()),subs(x=2,yreal)+(err())*
  (-1)^(rand()),subs(x=3,yreal)+(err())*(-1)^(rand()),subs(x=4,yreal)
  )+(err())*(-1)^(rand()),subs(x=5,yreal)+(err())*(-1)^(rand())]:
> n:=nops(X):
> sx:=add(i,i=X):
> medx:=sx/n:
> sy:=add(i,i=Y):
> medy:=sy/n:
> b:=medy-m*medx:
> sxy:=add(X[i]*Y[i], i=1..n):
> sX2:=add((X[i]-medx)^2, i=1..n):
> sY2:=add((Y[i]-medy)^2, i=1..n):
> sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n):
> ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mp:=`if`(ms>0,ms,mr):mn:=`if`(mr<0,mr,ms):
> m:=`if`(n*sxy>sx*sy,mp,mn):
> b:=evalf(medy-m*medx):
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n)):
> sx2:=add((X[i])^2,i=1..n):
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2)):
> bt:=evalf(medy-mt*medx):
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n)):
> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S:
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St:
> errorAlt2:=int((linea_alt-yreal)^2,x=0..5):
> errorTrad2:=int((linea_trad-yreal)^2,x=0..5):
> errorAlt:=int(abs(linea_alt-yreal),x=0..5):
> errorTrad:=int(abs(linea_trad-yreal),x=0..5):
> mas_preciso:=`if`(errorAlt<errorTrad,alternativo,tradicional):
> numA:=`if`(mas_preciso=alternativo,numA+1,numA):
> numT:=`if`(mas_preciso=tradicional,numT+1,numT):
> od:
> numA;
50
> numT;
50
```


B.3. Comparación de precisión para datos con error presente sólo en la variable y

```
[ > err:=(rand(0..10))/20:
[ > m:=rand(1..30)/10:
[ > b:=rand(0..10)/3:
[ > yreal:=m()*x+b();


$$y_{real} = \frac{4x}{5} + 3$$


[ > X:=[1,2,3,4,5];


$$X = [1, 2, 3, 4, 5]$$


[ > Y:=[subs(x=1,yreal)+(err())*(-1)^(rand()),subs(x=2,yreal)+(err())*(-1)^(rand()),subs(x=3,yreal)+(err())*(-1)^(rand()),subs(x=4,yreal)+(err())*(-1)^(rand()),subs(x=5,yreal)+(err())*(-1)^(rand())];


$$Y = \begin{bmatrix} \frac{39}{10} & \frac{47}{10} & \frac{99}{20} & \frac{57}{10} & \frac{143}{20} \end{bmatrix}$$


[ > n:=nops(X);


$$n = 5$$


[ > sx:=add(i,i=X);


$$sx = 15$$


[ > medx:=sx/n:
[ > sy:=add(i,i=Y):
[ > medy:=sy/n:
[ > b:=medy-m*medx:
[ > sxy:=add(X[i]*Y[i], i=1..n):
[ > sX2:=add((X[i]-medx)^2, i=1..n):
[ > sY2:=add((Y[i]-medy)^2, i=1..n):
[ > sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n):
[ > ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
[ > mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
[ > mp:=`if`(ms>0,ms,mr):mn:=`if`(mr<0,mr,ms):
[ > m:=`if`(n*sxy>sx*sy,mp,mn);


$$m = 0.7694176245$$


[ > b:=evalf(medy-m*medx);


$$b = 2.971747126$$


[ > S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n));


$$S = 0.2523678164$$


[ > sx2:=add((X[i])^2,i=1..n):
[ > mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2));


$$mt = 0.7500000000$$


[ > bt:=evalf(medy-mt*medx);


$$bt = 3.0300000000$$


[ > St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n));


$$St = 0.3980000000$$

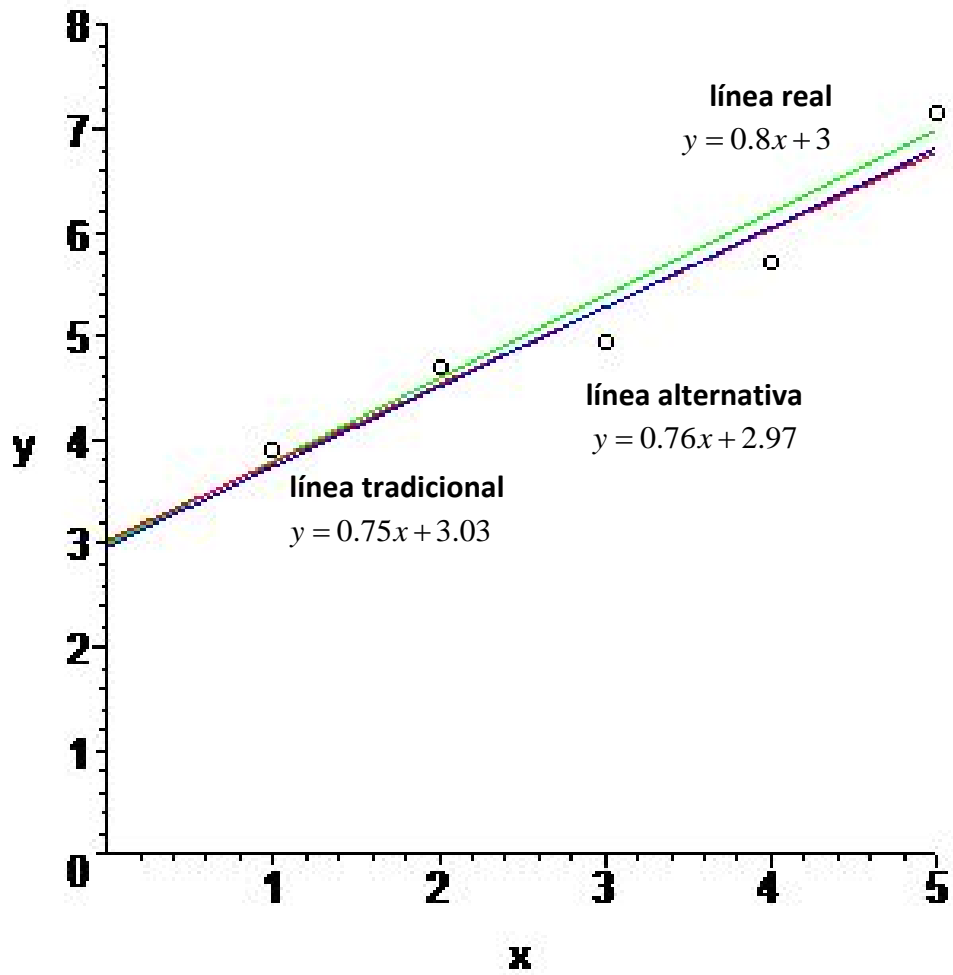
```

```

> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S;
      linea_alt := 0.7694176245 x + 2.971747126
      error_alt := 0.2523678164
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St;
      linea_trad := 0.7500000000 x + 3.0300000000
      error_trad := 0.3980000000
> errorAlt2:=int((linea_alt-yreal)^2,x=0..5);
      errorAlt2 := 0.06456219495
> errorTrad2:=int((linea_trad-yreal)^2,x=0..5);
      errorTrad2 := 0.07116666667
> errorAlt:=int(abs(linea_alt-yreal),x=0..5);
      errorAlt := 0.5235440638
> errorTrad:=int(abs(linea_trad-yreal),x=0..5);
      errorTrad := 0.4930000000
> mas_preciso:=`if`(errorAlt<errorTrad,alternativo,tradicional);
      mas_preciso := tradicional
> with(plots):
> a:=plot(linea_alt,x=0..5,y=0..8,color=blue):
> t:=plot(linea_trad,x=0..5,y=0..8,color=red):
> r:=plot(yreal,x=0..5,y=0..8,color=green):
> q:=plot(X,Y,x=0..5,y=0..8,color=black,style=point,symbol=circle,
      color=black,title=`EJEMPLO`):
> display(a,t,r,q);

```

EJEMPLO



B.4. Comparación de precisión para datos con error presente sólo en la variable y

(Versión de iteración múltiple)

```
[> numA:=0:numT:=0:
> for i from 1 to 100 do
> err:=(rand(0..10))/20:
> m:=rand(1..30)/10:
> b:=rand(0..10)/3:
> yreal:=m()*x+b():
> X:=[1,2,3,4,5]:
> Y:=[subs(x=1,yreal)+(err()*(-1)^(rand()),subs(x=2,yreal)+(err()*
(-1)^(rand()),subs(x=3,yreal)+(err()*(-1)^(rand()),subs(x=4,yreal
)+(err()*(-1)^(rand()),subs(x=5,yreal)+(err()*(-1)^(rand()))]:
> n:=nops(X):
> sx:=add(i,i=X):
> medx:=sx/n:
> sy:=add(i,i=Y):
> medy:=sy/n:
> b:=medy-m*medx:
> sxy:=add(X[i]*Y[i], i=1..n):
> sX2:=add((X[i]-medx)^2, i=1..n):
> sY2:=add((Y[i]-medy)^2, i=1..n):
> sXY:=add((X[i]-medx)*(Y[i]-medy), i=1..n):
> ms:=evalf(sY2-sX2+sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mr:=evalf(sY2-sX2-sqrt((sX2-sY2)^2+4*sXY^2))/(2*sXY):
> mp:=`if`(ms>0,ms,mr):mn:=`if`(mr<0,mr,ms):
> m:=`if`(n*sxy>sx*sy,mp,mn):
> b:=evalf(medy-m*medx):
> S:=evalf(add((m*X[i]-Y[i]+b)^2/(m^2+1),i=1..n)):
> sx2:=add((X[i])^2,i=1..n):
> mt:=evalf((n*sxy-sx*sy)/(n*sx2-sx^2)):
> bt:=evalf(medy-mt*medx):
> St:=evalf(add((Y[i]-mt*X[i]-bt)^2,i=1..n)):
> linea_alt:=evalf(m)*x+evalf(b);error_alt:=S:
> linea_trad:=evalf(mt)*x+evalf(bt);error_trad:=St:
> errorAlt2:=int((linea_alt-yreal)^2,x=0..5):
> errorTrad2:=int((linea_trad-yreal)^2,x=0..5):
> errorAlt:=int(abs(linea_alt-yreal),x=0..5):
> errorTrad:=int(abs(linea_trad-yreal),x=0..5):
> mas_preciso:=`if`(errorAlt<errorTrad,alternativo,tradicional):
> numA:=`if`(mas_preciso=alternativo,numA+1,numA):
> numT:=`if`(mas_preciso=tradicional,numT+1,numT):
> od:
> numA;
49
> numT;
51
]>
```

APÉNDICES

B.5. Resultados de precisión entre ambos métodos

Múltiple: error en X y Y		Múltiple: error sólo en Y	
numA	numT	numA	numT
52	48	48	52
45	55	45	55
54	46	47	53
56	44	49	51
50	50	52	48
53	47	41	59
59	41	45	55
51	49	56	44
55	45	52	48
43	57	43	57
54	46	49	51
46	54	46	54
62	38	45	55
48	52	42	58
46	54	58	42
59	41	48	52
56	44	43	57
50	50	52	48
57	43	43	57
43	57	49	51
1039	961	953	1047
51.95%	48.05%	47.65%	52.35%

NumA: Número de veces más preciso por el método alternativo.

NumT: Número de veces más preciso por el método tradicional.

Apéndice de la Parte IV

Bases de las pruebas estadísticas

Sea Y una variable aleatoria que sigue una distribución normal con media μ y varianza σ^2 , es decir

$$Y \sim N(\mu, \sigma^2)$$

A.1. Distribuciones Centrales

1. Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \sigma_i^2$. Sean a_1, a_2, \dots, a_n constantes conocidas. Si se define la combinación lineal de variables Y_i por

$$U = \sum_{i=1}^n a_i Y_i$$

entonces

$$U \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

Esto es, una combinación de variables aleatorias normalmente distribuidas, también sigue una distribución normal.

2. Si $Y \sim N(\mu, \sigma^2)$, entonces

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Z es la llamada variable aleatoria normal estándar.

3. Sea $Z = (Y - \mu) / \sigma$. Si $Y \sim N(\mu, \sigma^2)$, entonces Z^2 sigue una distribución χ^2 , la cual se denota por

$$Z^2 \sim \chi_1^2$$

Esto es, el cuadrado de una variable aleatoria normal estándar es una variable aleatoria χ^2 con un grado de libertad.

APÉNDICES

4. Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \sigma_i^2$, y sean

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}$$

entonces

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2$$

Esto es: (1) la suma de n variables aleatorias normales estándar sigue una distribución χ^2 con n grados de libertad, y (2) la suma de variables aleatorias χ^2 también sigue una distribución χ^2 .

5. **El teorema del límite central.** Si Y_1, Y_2, \dots, Y_n son variables aleatorias independientes e idénticamente distribuidas con $E(Y_i) = \mu_i$ y $\text{Var}(Y_i) = \sigma_i^2 < \infty$, entonces

$$\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

converge en distribución a una normal estándar cuando $n \rightarrow \infty$.

Esto es, si n es suficientemente grande, entonces \bar{Y} sigue aproximadamente una distribución normal.

6. Si $Z \sim N(0,1)$, $V \sim \chi_v^2$, y Z y V son independientes, entonces

$$\frac{Z}{\sqrt{V/v}} \sim t_v$$

donde t_v es una distribución t con v grados de libertad.

7. Sea $V \sim \chi_v^2$, y sea $W \sim \chi_\eta^2$. Si V y W son independientes, entonces

$$\frac{V/v}{W/\eta} \sim F_{v,\eta}$$

donde $F_{v,\eta}$ es una distribución F con v y η grados de libertad. Esto es, el cociente de dos variables aleatorias χ^2 , cada una dividida entre sus respectivos grados de libertad, sigue una distribución F .

A.2. Distribuciones no centrales

1. Sea $X \sim N(\delta, 1)$, y sea $V \sim \chi_v^2$. Si X y V son independientes, entonces

$$\frac{X}{\sqrt{V/v}} \sim t'_{v,\delta}$$

donde $t'_{v,\delta}$ es la distribución t no central con v grados de libertad y parámetro de no centralidad δ .

2. Si $X \sim N(\delta, 1)$, entonces

$$X^2 \sim \chi_{1,\delta^2}^2$$

donde χ_{1,δ^2}^2 es la distribución χ^2 no central con un grado de libertad y parámetro de no centralidad δ^2 .

3. Si X_1, X_2, \dots, X_n , son variables aleatorias independientes y normalmente distribuidas con $E(X_i) = \delta_i$ y $\text{Var}(X_i) = 1$, entonces

$$\sum_{i=1}^n X_i^2 \sim \chi_{n,\lambda}^2$$

donde el parámetro de no centralidad, λ , es

$$\lambda = \sum_{i=1}^n \delta_i^2$$

4. Sea $V \sim \chi_{n,\lambda}^2$, y sea $W \sim \chi_\eta^2$. Si V y W son independientes, entonces

$$\frac{V/v}{W/\eta} \sim F'_{v,\eta,\lambda}$$

donde $F'_{v,\eta,\lambda}$ es una distribución F no central con v y η grados de libertad y parámetro de no centralidad λ .

Bases de la teoría de modelos lineales

B.1. Definiciones básicas para matrices

1. Rango de una matriz. El rango de una matriz \mathbf{A} , es el número de columnas linealmente independientes, que también es el número de renglones linealmente independientes.

2. Matriz identidad. La matriz identidad de orden k , denotada por \mathbf{I} o por \mathbf{I}_k , es una matriz cuadrada de $k \times k$ cuyos elementos diagonales son 1, y cuyos elementos no diagonales son 0. Esto es:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \cdots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

3. Matriz inversa. Sea \mathbf{A} una matriz de $k \times k$. La inversa de \mathbf{A} , denotada por \mathbf{A}^{-1} , es otra matriz de $k \times k$ tal que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Si la inversa existe, entonces es única.

4. Matriz transpuesta. Sea \mathbf{A} una matriz de $n \times k$. La transpuesta de \mathbf{A} , denotada por \mathbf{A}' o por \mathbf{A}^T , es una matriz de $k \times n$ tal que si

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix}, \text{ entonces } \mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix}$$

Nota: Si \mathbf{A} es una matriz de $m \times n$ y \mathbf{B} es una matriz de $n \times p$, entonces

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

5. Matriz simétrica. Sea \mathbf{A} una matriz de $k \times k$. Se dice que \mathbf{A} es simétrica si $\mathbf{A} = \mathbf{A}'$.

6. Matriz idempotente. Sea \mathbf{A} una matriz de $k \times k$. Se dice que \mathbf{A} es idempotente si $\mathbf{A} = \mathbf{A}\mathbf{A}$. Si \mathbf{A} también es simétrica, entonces se dice que \mathbf{A} es simétrica idempotente. Si \mathbf{A} es simétrica idempotente, entonces $\mathbf{I} - \mathbf{A}$ también es simétrica idempotente.

7. Matriz ortogonal. Sea \mathbf{A} una matriz de $k \times k$. Si \mathbf{A} es ortogonal, entonces $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Como una consecuencia, si \mathbf{A} es una matriz ortogonal, entonces $\mathbf{A}^{-1} = \mathbf{A}'$.

8. Forma cuadrática. Sea \mathbf{y} un vector de $k \times 1$, y sea \mathbf{A} una matriz de $k \times k$. La función

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_{i=1}^k \sum_{j=1}^k a_{ij}y_iy_j$$

es una forma cuadrática. \mathbf{A} es llamada la matriz de la forma cuadrática.

9. Matrices definidas positivas y matrices semidefinidas positivas. Sea \mathbf{A} una matriz de $k \times k$. Se dice que \mathbf{A} es definida positiva si cumple las siguientes condiciones:

- (a) $\mathbf{A} = \mathbf{A}'$ (\mathbf{A} es simétrica) y
- (b) $\mathbf{y}'\mathbf{A}\mathbf{y} > 0 \quad \forall \mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq \mathbf{0}$

\mathbf{A} es definida semipositiva si cumple la siguiente condición:

- (a) $\mathbf{y}'\mathbf{A}\mathbf{y} = 0$ para alguna $\mathbf{y} \neq \mathbf{0}$

10. Traza de una matriz. Sea \mathbf{A} una matriz de $k \times k$. La traza de \mathbf{A} , denotada por $\text{traza}(\mathbf{A})$ o $\text{tr}(\mathbf{A})$, es la suma de los elementos diagonales de \mathbf{A} ; esto es

$$\text{traza}(\mathbf{A}) = \sum_{i=1}^k a_{ii}$$

Nota:

- (a) Si \mathbf{A} es una matriz de $m \times n$, y \mathbf{B} es una matriz de $n \times m$, entonces

$$\text{traza}(\mathbf{AB}) = \text{traza}(\mathbf{BA})$$

- (b) Si las matrices son perfectamente conformables, entonces

$$\text{traza}(\mathbf{ABC}) = \text{traza}(\mathbf{CAB})$$

- (c) Si \mathbf{A} y \mathbf{B} son matrices de $k \times k$, y a y b son escalares, entonces

$$\text{traza}(a\mathbf{A} + b\mathbf{B}) = a \text{traza}(\mathbf{A}) + b \text{traza}(\mathbf{B})$$

APÉNDICES

11. *Rango de una matriz idempotente.* Sea \mathbf{A} una matriz idempotente. El rango de \mathbf{A} es su traza.

12. *Una identidad importante para una matriz particionada.* Sea \mathbf{X} una matriz de $n \times p$ particionada tal que

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$$

Nótese que

$$\begin{aligned}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} &= \mathbf{X} \\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}_1, \mathbf{X}_2] &= \mathbf{X} \\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}_1, \mathbf{X}_2] &= [\mathbf{X}_1, \mathbf{X}_2]\end{aligned}$$

Por consecuencia,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1 = \mathbf{X}_1 \quad y$$

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2 = \mathbf{X}_2$$

De forma similar

$$\mathbf{X}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' = \mathbf{X}'_1 \quad y$$

$$\mathbf{X}'_2 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' = \mathbf{X}'_2$$

B.2. Derivadas de las matrices

Sea \mathbf{A} una matriz de constantes de $k \times k$, \mathbf{a} un vector de constantes de $k \times 1$, y \mathbf{y} un vector de variables de $k \times 1$.

1. Si $\mathbf{z} = \mathbf{a}'\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{a}'\mathbf{y}}{\partial \mathbf{y}} = \mathbf{a}$$

2. Si $\mathbf{z} = \mathbf{y}'\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{y}'\mathbf{y}}{\partial \mathbf{y}} = 2\mathbf{y}$$

3. Si $\mathbf{z} = \mathbf{a}'\mathbf{A}\mathbf{y}$, entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{a}' \mathbf{A} \mathbf{y}}{\partial \mathbf{y}} = \mathbf{A}' \mathbf{a}$$

4. Si $\mathbf{z} = \mathbf{y}' \mathbf{A} \mathbf{y}$ entonces

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \frac{\partial \mathbf{y}' \mathbf{A} \mathbf{y}}{\partial \mathbf{y}} = \mathbf{A} \mathbf{y} + \mathbf{A}' \mathbf{y}$$

Si \mathbf{A} es simétrica, entonces

$$\frac{\partial \mathbf{y}' \mathbf{A} \mathbf{y}}{\partial \mathbf{y}} = 2 \mathbf{A} \mathbf{y}$$

B.3. Esperanzas

Sea \mathbf{A} una matriz de constantes de $k \times k$, \mathbf{a} un vector de constantes de $k \times 1$, y \mathbf{y} un vector aleatorio de $k \times 1$ con media $\boldsymbol{\mu}$ y con matriz \mathbf{V} de varianzas y covarianzas no singular.

1. $E(\mathbf{a}' \mathbf{y}) = \mathbf{a}' \boldsymbol{\mu}$
2. $E(\mathbf{A} \mathbf{y}) = \mathbf{A} \boldsymbol{\mu}$
3. $\text{Var}(\mathbf{a}' \mathbf{y}) = \mathbf{a}' \mathbf{V} \mathbf{a}$
4. $\text{Var}(\mathbf{A} \mathbf{y}) = \mathbf{A} \mathbf{V} \mathbf{A}'$

Nota: Si $\mathbf{V} = \sigma^2 \mathbf{I}$, entonces $\text{Var}(\mathbf{A} \mathbf{y}) = \sigma^2 \mathbf{A} \mathbf{A}'$

5. $E(\mathbf{y}' \mathbf{A} \mathbf{y}) = \text{traza}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}$

Nota: Si $\mathbf{V} = \sigma^2 \mathbf{I}$, entonces $E(\mathbf{y}' \mathbf{A} \mathbf{y}) = \sigma^2 \text{traza}(\mathbf{A}) + \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}$

B.4. Teoría de distribución

Sea \mathbf{A} una matriz de constantes de $k \times k$, y \mathbf{y} un vector aleatorio normal multivariado de $k \times 1$ con media $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas no singular \mathbf{V} ; entonces,

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$$

Sea U la forma cuadrática definida por $U = \mathbf{y}' \mathbf{A} \mathbf{y}$.

5. Si $\mathbf{A} \mathbf{V}$ o $\mathbf{V} \mathbf{A}$ es una matriz idempotente de rango p , entonces

$$U \sim \chi_{p, \lambda}^2$$

donde $\lambda = \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu}$.

APÉNDICES

6. Sea $\mathbf{V} = \sigma^2 \mathbf{I}$, la cual es una típica suposición. Si \mathbf{A} es idempotente con rango p , entonces

$$\frac{U}{\sigma^2} \sim \chi_{p,\lambda}^2$$

donde $\lambda = \boldsymbol{\mu}' \mathbf{A} \boldsymbol{\mu} / \sigma^2$.

7. Sea \mathbf{B} una matriz de $q \times k$, y sea \mathbf{W} la forma lineal dada por $\mathbf{W} = \mathbf{B} \mathbf{y}$. La forma cuadrática $U = \mathbf{y}' \mathbf{A} \mathbf{y}$ y \mathbf{W} son independientes si

$$\mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}$$

Nota: Si $\mathbf{V} = \sigma^2 \mathbf{I}$, entonces U y \mathbf{W} son independientes si $\mathbf{B} \mathbf{A} = \mathbf{0}$.

8. Sea \mathbf{B} una matriz de $k \times k$. Sea $V = \mathbf{y}' \mathbf{B} \mathbf{y}$. Las dos formas cuadráticas, U y V , son independientes si

$$\mathbf{A} \mathbf{V} \mathbf{B} = \mathbf{0}$$

Nota: Si $\mathbf{V} = \sigma^2 \mathbf{I}$, entonces U y V son independientes si $\mathbf{A} \mathbf{B} = \mathbf{0}$.

Resultados importantes de SC_R y SC_{Res}

C.1. Resultados importantes de SC_R

Por definición,

$$SC_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

nótese que $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ y que

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \mathbf{1}'\mathbf{y}$$

donde $\mathbf{1}$ es un vector de $n \times 1$ con todos sus elementos unos. Además, $n = \mathbf{1}'\mathbf{1}$; entonces, $\bar{y} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}$. Por consecuencia, se puede escribir SC_R como

$$\begin{aligned} SC_R &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= [\hat{\mathbf{y}} - \mathbf{1}\bar{y}]' [\hat{\mathbf{y}} - \mathbf{1}\bar{y}] \\ &= [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}]' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y}] \\ &= \mathbf{y}' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] \mathbf{y} \end{aligned}$$

Nótese ahora que $\mathbf{X} = [\mathbf{1}, \mathbf{X}_R]$ donde \mathbf{X}_R es la matriz formada por los valores reales de las regresoras. De esta forma, SC_R involucra un caso especial de una matriz particionada. Entonces, se usa la identidad especial para matrices particionadas para mostrar que

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{1} = \mathbf{1} \quad \text{y} \quad \mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{1}'$$

Consecuentemente, se puede demostrar que $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']$ es idempotente. Bajo la suposición de que $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, entonces

$$\frac{SC_R}{\sigma^2} = \frac{1}{\sigma^2} \mathbf{y}' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] \mathbf{y}$$

sigue una distribución χ^2 no central con parámetro de no centralidad λ y grados libertad igual al número del rango de $[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}']$. Ya que esta matriz es idempotente, el rango es su traza. Nótese que

$$\begin{aligned} \text{traza}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] &= \text{traza}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] - \text{traza}[\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] \\ &= \text{traza}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] - \text{traza}[\mathbf{1}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}] \\ &= \text{traza}(\mathbf{I}_p) - \text{traza}(\mathbf{1}) \\ &= p - 1 = k \end{aligned}$$

Bajo la suposición de que el modelo es correcto,

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} = [\mathbf{1}, \mathbf{X}_R] \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_R \end{bmatrix} = \beta_0\mathbf{1} + \mathbf{X}_R\boldsymbol{\beta}_R$$

Entonces, el parámetro de no centralidad es

$$\begin{aligned} \lambda &= \frac{1}{\sigma^2} E(\mathbf{y})' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] E(\mathbf{y}) \\ &= \frac{1}{\sigma^2} \boldsymbol{\beta}' \mathbf{X}' [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] \mathbf{X}\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2} [\beta_0 \quad \boldsymbol{\beta}'_R] \begin{bmatrix} \mathbf{1}' \\ \mathbf{X}'_R \end{bmatrix} [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'] [\mathbf{1} \quad \mathbf{X}_R] \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_R \end{bmatrix} \\ &= \frac{1}{\sigma^2} [\beta_0 \quad \boldsymbol{\beta}'_R] \begin{bmatrix} \mathbf{1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{1}'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \\ \mathbf{X}'_R\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\mathbf{X}'_R\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \end{bmatrix} [\mathbf{1} \quad \mathbf{X}_R] \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_R \end{bmatrix} \\ &= \frac{1}{\sigma^2} [\beta_0 \quad \boldsymbol{\beta}'_R] \begin{bmatrix} \mathbf{0}' \\ \mathbf{X}'_R-\mathbf{X}'_R\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' \end{bmatrix} [\mathbf{1} \quad \mathbf{X}_R] \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_R \end{bmatrix} \\ &= \frac{1}{\sigma^2} [\beta_0 \quad \boldsymbol{\beta}'_R] \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{X}'_R\mathbf{X}_R-\mathbf{X}'_R\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{X}_R \end{bmatrix} \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_R \end{bmatrix} \\ &= \frac{1}{\sigma^2} \boldsymbol{\beta}'_R [\mathbf{X}'_R\mathbf{X}_R-\mathbf{X}'_R\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{X}_R] \boldsymbol{\beta}_R \end{aligned}$$

Si ahora se define la matriz de los valores no *centrados* de las regresoras,

$$\mathbf{X}_C = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2k} - \bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{nk} - \bar{x}_k \end{bmatrix}$$

donde \bar{x}_1 es el promedio de la primera regresora, \bar{x}_2 de la segunda, y así sucesivamente, entonces es fácil comprobar que se puede reescribir el parámetro de no centralidad como

$$\lambda = \frac{1}{\sigma^2} \boldsymbol{\beta}'_R [\mathbf{X}'_C \mathbf{X}_C] \boldsymbol{\beta}_R$$

El valor esperado para SC_R es

$$\begin{aligned} E(SC_R) &= E\left(\mathbf{y}' \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] \mathbf{y}\right) \\ &= \text{traza} \left(\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] \sigma^2 \mathbf{I} \right) \\ &\quad + E(\mathbf{y})' \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] E(\mathbf{y}) \\ &= k\sigma^2 + \boldsymbol{\beta}'_R \mathbf{X}'_C \mathbf{X}_C \boldsymbol{\beta}_R \end{aligned}$$

Y como un resultado,

$$E(CM_R) = E\left(\frac{SC_R}{k}\right) = \sigma^2 + \frac{\boldsymbol{\beta}'_R \mathbf{X}'_C \mathbf{X}_C \boldsymbol{\beta}_R}{k}$$

C.2. Resultados importantes de SC_{Res}

Por definición,

$$SC_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Se puede escribir SC_{Res} como

$$\begin{aligned} SC_{Res} &= (\mathbf{y} - \hat{\mathbf{y}})' (\mathbf{y} - \hat{\mathbf{y}}) \\ &= \left[\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \right]' \left[\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \right] \\ &= \mathbf{y}' \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y} \end{aligned}$$

Es fácil ver que $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ es simétrica idempotente. De esta forma,

$$\frac{SC_{\text{Res}}}{\sigma^2} = \frac{1}{\sigma^2} \mathbf{y}' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{y}$$

sigue una distribución χ^2 . Los grados de libertad vienen del rango de $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$, que a su vez es su traza. Se puede demostrar que ésta es $n - p$. Bajo la suposición de que el modelo es correcto,

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

Entonces, el parámetro de no centralidad es

$$\begin{aligned} & \frac{1}{\sigma^2} E(\mathbf{y})' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] E(\mathbf{y}) \\ &= \frac{1}{\sigma^2} \boldsymbol{\beta}' \mathbf{X}' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{X}\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2} \boldsymbol{\beta}' [\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \boldsymbol{\beta} = 0 \end{aligned}$$

Y como un resultado,

$$\frac{SC_{\text{Res}}}{\sigma^2} \sim \chi_{n-p}^2$$

El valor esperado de SC_{Res} es

$$\begin{aligned} E(SC_{\text{Res}}) &= E\left(\mathbf{y}' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \mathbf{y}\right) \\ &= \text{traza}\left([\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \sigma^2 \mathbf{I}\right) + E(\mathbf{y})' [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] E(\mathbf{y}) \\ &= (n - p)\sigma^2 \end{aligned}$$

Como un resultado,

$$E(CM_{\text{Res}}) = E\left(\frac{SC_{\text{Res}}}{n - p}\right) = \sigma^2$$

C.3. La prueba F

Una estadística F es el cociente de dos variables aleatorias χ^2 independientes, cada una dividida por sus respectivos grados de libertad. Ya se vio que SC_R / σ^2 y SC_{Res} / σ^2 siguen una distribución χ^2 . La idea es mostrar ahora que son independientes. De la teoría básica de modelos lineales, SC_R y SC_{Res} lo son bajo la suposición de que $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ si

$$\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] \sigma^2 \mathbf{I} \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] = \mathbf{0}$$

Nótese que

$$\begin{aligned} & \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] \sigma^2 \mathbf{I} \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \\ &= \sigma^2 \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \\ &= \sigma^2 \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right. \\ & \quad \left. + \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \\ &= \sigma^2 \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' + \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}' \right] = \mathbf{0} \end{aligned}$$

Por lo que SC_R y SC_{Res} son independientes. Ahora se ve que

$$\frac{SC_R}{k\sigma^2} = \frac{CM_R}{\sigma^2} \quad \text{y} \quad \frac{SC_{Res}}{(n-p)\sigma^2} = \frac{CM_{Res}}{\sigma^2}$$

son variables aleatorias χ^2 , cada una dividida por sus respectivos grados de libertad. Como resultado se tiene,

$$\frac{SC_R}{CM_{Res}} \sim F'_{k, n-p, \lambda}$$

donde

$$\lambda = \frac{1}{\sigma^2} \boldsymbol{\beta}'_R \mathbf{X}'_C \mathbf{X}_C \boldsymbol{\beta}_R$$

En el caso especial de la regresión lineal simple, se tiene una sola regresora; por lo que $\boldsymbol{\beta}_R = \beta_1$, y

APÉNDICES

$$\mathbf{X}'_C \mathbf{X}_C = \sum_{i=1}^n (x_i - \bar{x})^2$$

Y como resultado para la regresión lineal simple,

$$\frac{SC_R}{CM_{\text{Res}}} \sim F'_{1, n-2, \lambda}$$

donde $\lambda = \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$.

D.1. El teorema de Gauss – Markov cuando $\text{Var}(\varepsilon) = \sigma^2 \mathbf{I}$

El teorema de Gauss – Markov establece que el estimador por mínimos cuadrados ordinarios (estimador MCO) de β , $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$, es el mejor estimador lineal insesgado (BLUE por sus siglas en inglés). Por “el mejor” significa que $\hat{\beta}$ tiene la varianza más pequeña en comparación a la clase de estimadores insesgados que son combinaciones lineales de los datos. Un problema es que $\hat{\beta}$ es un vector; por lo que su varianza es realmente una matriz. De esta forma, se desea probar que $\hat{\beta}$ tiene la varianza mínima para cualquier combinación lineal de los coeficientes estimados, $\mathcal{L}'\hat{\beta}$. Nótese que

$$\begin{aligned} \text{Var}(\mathcal{L}'\hat{\beta}) &= \mathcal{L}'\text{Var}(\hat{\beta})\mathcal{L} \\ &= \mathcal{L}'\left[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right]\mathcal{L} \\ &= \sigma^2\mathcal{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{L} \end{aligned}$$

es un escalar. Sea $\tilde{\beta}$ otro estimador insesgado de β que es una combinación lineal de \mathbf{y} . El objetivo es mostrar entonces que $\text{Var}(\mathcal{L}'\tilde{\beta}) \geq \sigma^2\mathcal{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{L}$ con al menos una \mathcal{L} tal que $\text{Var}(\mathcal{L}'\tilde{\beta}) > \sigma^2\mathcal{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{L}$. Primero, nótese que se puede escribir cualquier otro estimador de β como combinación lineal de \mathbf{y} , esto es

$$\tilde{\beta} = \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{B} \right] \mathbf{y} + \mathbf{b}_0$$

donde \mathbf{B} es una matriz de $p \times n$, y \mathbf{b}_0 es un vector de constantes de $p \times 1$ que ajusta de forma apropiada los estimadores MCO para formar el alternativo. También se ve que si el modelo es correcto, entonces

$$\begin{aligned} E(\tilde{\beta}) &= E\left(\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{B} \right] \mathbf{y} + \mathbf{b}_0\right) \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{B} \right] E(\mathbf{y}) + \mathbf{b}_0 \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{B} \right] \mathbf{X}\beta + \mathbf{b}_0 \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta + \mathbf{B}\mathbf{X}\beta + \mathbf{b}_0 \\ &= \beta + \mathbf{B}\mathbf{X}\beta + \mathbf{b}_0 \end{aligned}$$

De esta manera, $\tilde{\beta}$ es insesgado si y sólo si $\mathbf{b}_0 = \mathbf{0}$ y $\mathbf{BX} = \mathbf{0}$. La varianza de $\tilde{\beta}$ es

$$\begin{aligned}\text{Var}(\tilde{\beta}) &= \text{Var}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]\mathbf{y}\right) \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]\text{Var}(\mathbf{y})\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]', \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]\sigma^2\mathbf{I}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]', \\ &= \sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{B}\right]\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}+\mathbf{B}'\right] \\ &= \sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1}+\mathbf{BB}'\right]\end{aligned}$$

ya que $\mathbf{BX} = \mathbf{0}$, lo cual a su vez $(\mathbf{BX})' = \mathbf{X}'\mathbf{B}' = \mathbf{0}$. Como un resultado,

$$\begin{aligned}\text{Var}(\mathcal{L}'\tilde{\beta}) &= \mathcal{L}'\text{Var}(\tilde{\beta})\mathcal{L} \\ &= \mathcal{L}'\left(\sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1}+\mathbf{BB}'\right]\right)\mathcal{L} \\ &= \sigma^2\mathcal{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathcal{L} + \sigma^2\mathcal{L}'\mathbf{BB}'\mathcal{L} \\ &= \text{Var}(\mathcal{L}'\hat{\beta}) + \sigma^2\mathcal{L}'\mathbf{BB}'\mathcal{L}\end{aligned}$$

Primero, nótese que \mathbf{BB}' es al menos una matriz definida semipositiva; ya que $\sigma^2\mathcal{L}'\mathbf{BB}'\mathcal{L} \geq \mathbf{0}$. Ahora, se puede definir la combinación lineal $\mathcal{L}^* = \mathbf{B}'\mathcal{L}$, y como resultado

$$\mathcal{L}'\mathbf{BB}'\mathcal{L} = \mathcal{L}^*{}'\mathcal{L}^* = \sum_{i=1}^p \mathcal{L}_i^{*2}$$

el cual es estrictamente más grande que 0 para alguna $\mathcal{L} \neq \mathbf{0}$ a menos de que $\mathbf{B} = \mathbf{0}$. Entonces, el estimador MCO de β es el mejor estimador lineal insesgado.

Apéndice de la Parte V

A.1. Black-Scholes PDE (Ecuación diferencial parcial de Black-Scholes)

Considérese que se tiene una opción cuyo valor $V(S,t)$ depende únicamente de S el precio de mercado, y de t el tiempo. No es necesario especificar si V pertenece a una Call o a una Put; en efecto, V puede ser el valor de un portafolio de diferentes opciones, pero piense por simplicidad que se trata de una simple Call o Put. Usando el lema de Itô [la ecuación (5.1)], para cualquier función dos veces diferenciable $f(S,t)$ se obtiene:

$$df = \sigma S \frac{\partial f}{\partial S} dx + \left(\mu S \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + \frac{\partial f}{\partial t} \right) dt \quad (1)$$

Por lo que se puede escribir

$$dV = \sigma S \frac{\partial V}{\partial S} dX + \left(\mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} \right) dt \quad (2)$$

Nótese que se requiere que V tenga al menos una derivada en t y dos derivadas en S .

A continuación se construye un portafolio consistente de una opción y una cantidad $-\Delta$ del bien subyacente. Este número no es aún especificado. El valor de este portafolio es

$$\Pi = V - \Delta S$$

El salto de su precio en un paso de tiempo es

$$d\Pi = dV - \Delta dS \quad (3)$$

Aquí Δ se mantiene fijo durante todo el período de tiempo; si no fuera así, entonces $d\Pi$ podría contener términos en $d\Delta$. Poniendo todo junto, (5.1), (2) y (3), se encuentra que Π sigue la caminata aleatoria

$$d\Pi = \sigma S \left(\frac{\partial V}{\partial S} - \Delta \right) dX + \left(\mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \mu \Delta S \right) dt \quad (4)$$

Se pueden eliminar los componentes aleatorios en esta caminata escogiendo

$$\Delta = \frac{\partial V}{\partial S} \quad (5)$$

Esta es la tasa de cambio del valor de la opción o portafolio de opciones con respecto a S .

APÉNDICES

Nótese ahora que Δ es el valor de $\partial V / \partial S$ al inicio del período de tiempo dt . Esto resulta en un portafolio cuyo incremento es totalmente determinista:

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt \quad (6)$$

Considérese ahora los conceptos de arbitraje, suministro y demanda, con la suposición de que no existen costos de transacción. El retorno de una cantidad Π invertida en valores libres de riesgo debería verse como un incremento de $r\Pi dt$ en un tiempo dt . Si el lado derecho de la ecuación (6) fuera más grande que este monto, un inversionista podría hacer arbitraje garantizando un profit libre de riesgo, pidiendo prestado una cantidad Π para invertir en el portafolio. En este caso el profit sería más grande que el costo del préstamo. A la inversa, si el lado derecho de la ecuación (6) fuera menor que $r\Pi dt$, entonces un inversionista podría adoptar la posición corta en el portafolio e invertir Π en el banco. En cualquiera de las dos formas se realiza arbitraje. Dado que se supone el principio del no-arbitraje, esto conduce a la siguiente igualdad:

$$r\Pi dt = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt \quad (7)$$

Sustituyendo (1) y (5) dentro de (7) y dividiendo todo entre dt se llega a

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0 \quad (8)$$

Esta es la famosa *ecuación diferencial parcial de Black-Scholes* (PDE). Es una relación entre V , S y t , y algunas derivadas parciales de V . Nótese que (8) no contiene el parámetro de deriva μ . En otras palabras, el valor de una opción es independiente a la velocidad en que el precio del bien aumenta. El único elemento que aparece de (5.1) es la volatilidad σ . Una consecuencia de esto, es que dos personas podrían diferir en sus estimados para μ , pero estarían de acuerdo en el precio de la opción. Otro aspecto importante es que no se especifica el tipo de opción que está siendo evaluada. El PDE se satisface para cualquier tipo de opción sobre S cuyo valor puede ser expresado mediante una función suave $V(S, t)$. Por ejemplo, la fórmula de *Black-Scholes* calcula de forma exacta el precio de las opciones Call y Put europeas bajo condiciones específicas finales y de frontera. Para la valuación de opciones americanas, *Black-Scholes* se vuelve intratable. Complejas técnicas binomiales y de diferencia finita deben ser implementadas para aproximar el valor de la opción. Afortunadamente, en 1999, Longstaff y Schwartz comenzaron el desarrollo de un algoritmo que implementa regresión simple con base en la *fórmula de valuación neutral al riesgo*: El algoritmo LSM de mínimos cuadrados.

B.1. Demostración de las proposiciones

Demostración de la proposición 1

Por definición, el valor del bien subyacente X_{t_k} es \mathcal{F}_{t_k} – medible. De forma similar, el valor de ejercicio inmediato de la opción es \mathcal{F}_{t_k} – medible. Por construcción, $F_M(\omega; t_k)$ es una función lineal de funciones \mathcal{F}_{t_k} – medibles de X_{t_k} , y por lo tanto es \mathcal{F}_{t_k} – medible. Entonces, el caso donde el valor de ejercicio inmediato es más grande que cero y más grande o igual a $F_M(\omega; t_k)$ es en \mathcal{F}_{t_k} , y por lo tanto, la regla del algoritmo LSM para ejercer cuando este evento ocurre define un tiempo de paro. Se denota entonces el valor presente de la opción siguiendo esta regla de paro por E_θ . Debido a que $V(X)$ es el supremo de los valores presentes obtenidos sobre todo el conjunto de tiempos de paro, $V(X) \geq E_\theta$. Ya que la forma funcional de $F_M(\omega; t_k)$ es idéntica a través de todas las trayectorias, los flujos de efectivo descontados $LSM^*(\omega; M, K)$ son independientes e idénticamente distribuidos, y la ley fuerte de los grandes números implica que

$$\Pr \left[\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N LSM^*(\omega_i; M, K) = E_\theta \right] = 1$$

Esta igualdad combinada con la desigualdad $V(X) \geq E_\theta$ conllevan al resultado. ■

Demostración de la proposición 2

Al tiempo $t = 2$, la estrategia de paro del algoritmo LSM es la misma que la óptima; la opción es ejercida si se encuentra *in the money*. Bajo las suposiciones dadas, la esperanza condicional $F(\omega; t_1)$ es una función sólo para X_{t_1} . Si $F(\omega; t_1)$ satisface las condiciones indicadas, entonces el teorema de Sansone (IV.9.1, 1959) implica que la convergencia de $F_M(\omega; t_1)$ a $F(\omega; t_1)$ es uniforme en M sobre el intervalo $(0, \infty)$, donde los primeros polinomios de Laguerre son usados como el conjuntos de funciones base. Esto implica que para un ε dado, existe un M tal que $\sup_{X_{t_1}} \left[|F(\omega; t_1) - F_M(\omega; t_1)| < \varepsilon / 2 \right] = 0$. Por las condiciones de integrabilidad y el teorema 3.5 de White (1984), el valor ajustado $\hat{F}_M(\omega; t_1)$ de la regresión LSM converge en probabilidad a $F_M(\omega; t_1)$ cuando $N \rightarrow \infty$,

APÉNDICES

$$\lim_{N \rightarrow \infty} \Pr \left[\left| F_M(\omega; t_1) - \hat{F}_M(\omega; t_1) \right| > \varepsilon / 2 \right] = 0$$

Entonces, para cualquier ε , existe un M tal que

$$\lim_{N \rightarrow \infty} \Pr \left[\left| F(\omega; t_1) - \hat{F}_M(\omega; t_1) \right| > \varepsilon \right] = 0$$

Para completar la prueba, se particiona a Ω en cinco conjuntos: 1) el conjunto de trayectorias donde la opción es ejercida al tiempo t_1 bajo la estrategia óptima y la del algoritmo LSM, 2) el conjunto de trayectorias donde la opción es ejercida al tiempo t_1 , bajo la estrategia óptima o la del algoritmo LSM, 3) el conjunto de trayectorias donde la opción es ejercida al tiempo t_1 , bajo la estrategia del algoritmo LSM pero no bajo la óptima, 4) el conjunto de trayectorias donde la opción es ejercida al tiempo t_1 , bajo la estrategia óptima pero no bajo la del algoritmo LSM, y 5) un conjunto de trayectorias con probabilidad cero para las cuales la diferencia entre $F(\omega; t_1)$ y $\hat{F}_M(\omega; t_1)$ es más grande que ε cuando $N \rightarrow \infty$. Ahora, considérese un portafolio consistente de una posición larga en una opción ejercida usando la estrategia LSM, una inversión de ε en una cuenta del mercado de dinero, y una posición corta en una opción ejercida usando la estrategia óptima. Para las trayectorias en el conjunto 3), al tiempo t_1 , se usa el fondo de la cuenta y el resultado del ejercicio para adquirir una opción europea con fecha de maduración t_2 . Para las trayectorias del conjunto 4), al tiempo t_1 , se usan los fondos generados en la cuenta y la posición corta en una opción europea con fecha de maduración t_2 para financiar la salida de flujos de efectivo del ejercicio. Se puede ver que esta estrategia resulta en flujos de efectivo que son más grandes o iguales a cero para cada trayectoria de los conjuntos 1), 2), 3) y 4). Ya que éstos descontados son no-negativos, los promedios sobre todas las trayectorias son no-negativos, y el resultado se sigue de un argumento de no-arbitraje estándar, de la definición de $V(X)$, y de la ley de los grandes números. ■

C.1. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana (con monomios como funciones base)

```

close all
clear all

%=====
%Parámetros de la opción americana
%=====

%Stock
sigma=0.2;           %Volatilidad
S0=1.0;             %Precio inicial

%Tasas
r=0.06;            %Tasa libre de riesgo
D=0;              %Tasa de dividendos

%Opciones
T=3;              %Tiempo de maduración
K=1.1;           %Precio strike

%=====
%Simulaciones Monte Carlo para los precios
%=====

dt=1;             %Longitud del intervalo de tiempo (1 día, por ejemplo)
N=(T+1)/dt;      %Número de fechas de ejercicio contando t=0
NSim=8;          %Número de simulaciones

dBt=sqrt(dt)*randn(NSim,N); %Movimiento browniano
St=zeros(NSim,N); %Inicializamos la matriz
St(:,1)=S0*ones(NSim,1); %Vector de precio iniciales

%Simulación de precios
for t=2:N;
    St(:,t)=St(:,t-1).*exp((r-D-.5*sigma^2)*dt+sigma*dBt(:,t));
end

S=St;            %Sólo un cambio de nombre
S
NSim=size(St,1); %Número de simulaciones

```

APÉNDICES

```

%=====
%Estimación del precio de la opción americana
%=====

%Inicialización de la matrix de flujos de efectivo
M=NaN*ones(NSim,N);

M(:,N)=max(K-S(:,N),0);

for t=N:-1:3;

    %Paso 1: Selección de las trayectorias in the money al tiempo t-1
    I=find(K-S(:,t-1)>0);
    ISize=length(I);

    %Paso 2: Proyección de los flujos de efectivo de t a t-1
    if t==N
        Y=(ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N);
    else
        Y=sum(((ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N))')';
    end

    S1=S(I,t-1);
    X=[ones(ISize,1),S1,S1.^2];
    B=inv(X'*X)*X'*Y;
    B
    S2=S(:,t-1);
    X2=[ones(NSim,1),S2,S2.^2];

    %Encontrar cuando la opción es ejercida:
    IAlto=find(K-S(:,t-1)>=max(X2*B,0));

    %Encontrar cuando la opción no es ejercida:
    ICont=setdiff([1:NSim],IAlto);

    %Reemplazar los payoffs en la matrix (valores cuando la opción es
    %ejercida y ceros cuando no lo es)
    M(IAlto,t-1)=K-S(IAlto,t-1);
    M(IAlto,t:N)=zeros(length(IAlto),N-t+1);
    M(ICont,t-1)=zeros(length(ICont),1);

end

%=====
%Resultados
%=====

M
Ydesc=sum(((ones(NSim,1)*exp(-r*[1:N-1]*dt)).*M(:,2:N))')';
Ydesc
Valor=mean(Ydesc)
ErrorStd=std(Ydesc)/sqrt(NSim)

```

Resultados del programa

```

S =
    1.0000    1.0900    1.0800    1.3400
    1.0000    1.1600    1.2600    1.5400
    1.0000    1.2200    1.0700    1.0300
    1.0000    0.9300    0.9700    0.9200
    1.0000    1.1100    1.5600    1.5200
    1.0000    0.7600    0.7700    0.9000
    1.0000    0.9200    0.8400    1.0100
    1.0000    0.8800    1.2200    1.3400

B =
   -1.0700
    2.9834
   -1.8136

B =
    2.0375
   -3.3354
    1.3565

M =
    NaN         0         0         0
    NaN         0         0         0
    NaN         0         0    0.0700
    NaN    0.1700         0         0
    NaN         0         0         0
    NaN    0.3400         0         0
    NaN    0.1800         0         0
    NaN    0.2200         0         0

Ydesc =
     0
     0
    0.0585
    0.1601
     0
    0.3202
    0.1695
    0.2072

Valor =
    0.1144

Errorstd =
    0.0419

>>

```

APÉNDICES

C.2. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana (con polinomios de Laguerre como funciones base)

```

close all
clear all

%=====
%Simulación de precios del ejemplo anterior
%=====

S=[1 1.09 1.08 1.34; 1 1.16 1.26 1.54; 1 1.22 1.07 1.03; 1 0.93 0.97 ...
    0.92; 1 1.11 1.56 1.52; 1 0.76 0.77 0.90; 1 0.92 0.84 1.01; ...
    1 0.88 1.22 1.34]

%=====
%Parámetros de la opción americana
%=====

N=4;           %Número de fechas de ejercicio contando t=0
NSim=size(S,1); %Número de simulaciones
dt=1;         %Longitud del intervalo de tiempo
r=0.06;       %Tasa libre de riesgo
K=1.1;        %Precio strike

%=====
%Estimación del precio de la opción americana
%=====

%Inicialización de la matriz de flujos de efectivo
M=NaN*ones(NSim,N);
M(:,N)=max(K-S(:,N),0);

for t=N:-1:3;

    %Paso 1: Selección de las trayectorias in the money al tiempo t-1
    I=find(K-S(:,t-1)>0);
    ISize=length(I);

    %Paso 2: Proyección de los flujos de efectivo de t a t-1
    if t==N
        Y=(ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N);
    else
        Y=sum(((ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N))'')';
    end

    S1=S(I,t-1);
    X=[ones(ISize,1),1-S1,1-2*S1+0.5*S1.^2]
    B=inv(X'*X)*X'*Y;
    B

```

APÉNDICE DE LA PARTE V

```
S2=S(:,t-1);
X2=[ones(NSim,1),1-S2,1-2*S2+0.5*S2.^2];

%Encontrar cuando la opción es ejercida:
IAlto=find(K-S(:,t-1)>=max(X2*B,0));

%Encontrar cuando la opción no es ejercida:
ICont=setdiff([1:NSim],IAlto);

%Reemplazar los payoffs en la matriz (valores cuando la opción es
%ejercida y ceros cuando no lo es)
M(IAlto,t-1)=K-S(IAlto,t-1);
M(IAlto,t:N)=zeros(length(IAlto),N-t+1);
M(ICont,t-1)=zeros(length(ICont),1);

end

%=====
%Resultados
%=====

M
Ydesc=sum((ones(NSim,1)*exp(-r*[1:N-1]*dt)).*M(:,2:N))';
Ydesc
Valor=mean(Ydesc)
ErrorStd=std(Ydesc)/sqrt(NSim)
```

APÉNDICES

Resultados del Programa

S =

1.0000	1.0900	1.0800	1.3400
1.0000	1.1600	1.2600	1.5400
1.0000	1.2200	1.0700	1.0300
1.0000	0.9300	0.9700	0.9200
1.0000	1.1100	1.5600	1.5200
1.0000	0.7600	0.7700	0.9000
1.0000	0.9200	0.8400	1.0100
1.0000	0.8800	1.2200	1.3400

B =

-1.7137
4.2709
-3.6272

B =

1.4150
-2.0904
2.7129

M =

NaN	0	0	0
NaN	0	0	0
NaN	0	0	0.0700
NaN	0.1700	0	0
NaN	0	0	0
NaN	0.3400	0	0
NaN	0.1800	0	0
NaN	0.2200	0	0

ydesc =

0
0
0.0585
0.1601
0
0.3202
0.1695
0.2072

valor =

0.1144

ErrorStd =

0.0419

>>

C.3. Programa del algoritmo LSM de Longstaff y Schwartz para aproximar el valor de una Put americana expedida por PEMEX

```

close all
clear all

%=====
%Parámetros de la opción americana
%=====

%Stock
sigma=0.15;           %Volatilidad
S0=92.02;             %Precio Inicial

%Tasas
r=0.02;              %Tasa libre de riesgo
D=0;                 %Tasa de dividendos

%Opciones
T=12;                %Tiempo de maduración
K=100.84;            %Precio strike

%=====
%Simulaciones Monte Carlo para los precios
%=====

dt=1;                %Longitud del intervalo de tiempo; 1 mes en el ejemplo
N=(T+1)/dt;         %Número de fechas de ejercicio contando t=0
NSim=50;             %Número de simulaciones

dBt=sqrt(dt)*randn(NSim,N); %Movimiento browniano
St=zeros(NSim,N);   %Inicializamos la matriz
St(:,1)=S0*ones(NSim,1); %Vector de precio iniciales

%Simulación de precios
for t=2:N;
    St(:,t)=St(:,t-1).*exp((r-D-.5*sigma^2)*dt+sigma*dBt(:,t));
end

S=St;                %Sólo un cambio de nombre
S
NSim=size(St,1);    %Número de simulaciones

```


APÉNDICES

```

%=====
%Estimación del precio de la opción americana
%=====

%Inicialización de la matrix de flujos de efectivo
M=NaN*ones(NSim,N);

M(:,N)=max(K-S(:,N),0);

for t=N:-1:3;

    %Paso 1: Selección de las trayectorias in the money al tiempo t-1
    I=find(K-S(:,t-1)>0);
    ISize=length(I);

    %Paso 2: Proyección de los flujos de efectivo de t a t-1
    if t==N
        Y=(ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N);
    else
        Y=sum(((ones(ISize,1)*exp(-r*[1:N-t+1]*dt)).*M(I,t:N))')';
    end

    S1=S(I,t-1);
    X=[ones(ISize,1),1-S1,1-2*S1+0.5*S1.^2];
    B=inv(X'*X)*X'*Y;
    B
    S2=S(:,t-1);
    X2=[ones(NSim,1),1-S2,1-2*S2+0.5*S2.^2];

    %Encontrar cuando la opción es ejercida:
    IAlto=find(K-S(:,t-1)>=max(X2*B,0));

    %Encontrar cuando la opción no es ejercida:
    ICont=setdiff([1:NSim],IAlto);

    %Reemplazar los payoffs en la matrix (valores cuando la opción es
    %ejercida y ceros cuando no lo es)
    M(IAlto,t-1)=K-S(IAlto,t-1);
    M(IAlto,t:N)=zeros(length(IAlto),N-t+1);
    M(ICont,t-1)=zeros(length(ICont),1);

end

%=====
%Resultados
%=====

M
Ydesc=sum(((ones(NSim,1)*exp(-r*[1:N-1]*dt)).*M(:,2:N))')';
Ydesc
Valor=mean(Ydesc)
ErrorStd=std(Ydesc)/sqrt(NSim)

```

Resultados del Programa

S =

92.0200	89.8661	106.3337	94.3280	100.7685	89.8427	80.9835	78.5590	69.6566	79.2679	82.4811	101.8555	101.9895
92.0200	113.6324	137.2350	108.9409	107.0455	114.4360	118.9199	101.3337	118.9532	95.2885	109.2506	111.9163	108.5204
92.0200	104.0764	103.9987	101.1651	75.2655	75.3864	74.4093	79.4538	83.2663	69.3863	111.3225	142.1511	138.1765
92.0200	100.8612	111.9819	151.9399	207.1107	190.7312	148.0562	139.8846	166.2429	183.6217	176.9906	172.3076	156.5122
92.0200	73.1206	56.0966	57.3184	58.7889	45.8325	36.5886	39.5320	39.7145	38.5681	42.4665	40.5880	41.0163
92.0200	84.6637	91.8935	77.9355	73.9087	88.3954	81.7018	72.5498	68.9202	55.7046	40.4686	45.0419	37.9141
92.0200	108.8986	114.2840	107.6630	111.6794	120.8750	153.3823	182.9032	192.2801	207.1894	239.6650	279.2870	239.3357
92.0200	100.3450	95.9588	96.4718	117.1001	127.9663	120.5946	107.6236	103.9880	92.7439	100.4877	96.4566	87.5642
92.0200	91.0605	103.6378	114.2280	114.0332	135.2307	148.6110	152.5445	212.5133	190.9635	185.6451	222.9670	252.6644
92.0200	101.4590	88.7758	74.1380	78.9030	76.6825	85.6650	79.4009	77.6956	73.3988	86.5839	87.6698	82.9762
92.0200	98.5263	107.2789	111.6760	135.5806	144.2987	174.3054	199.4509	186.5964	141.2540	139.4076	135.7948	137.0800
92.0200	79.1580	75.2964	118.3227	101.9312	115.0284	131.3467	124.0856	104.5565	103.7729	145.1959	137.3818	150.9491
92.0200	87.0719	66.7872	65.1961	60.6393	63.3165	70.1014	86.3992	88.2884	75.7548	75.7666	87.1438	70.6228
92.0200	67.1714	75.4167	76.1822	76.3893	95.4005	104.7742	90.2977	70.2778	76.3978	77.6796	93.2173	88.4343
92.0200	117.5933	120.2713	114.5541	129.5060	107.8549	112.5561	139.1466	175.4313	183.6562	150.5714	178.7055	195.9670
92.0200	76.8336	64.2087	55.6687	53.7689	57.2114	81.9760	86.0613	102.6222	105.7752	98.9732	99.2481	110.4187
92.0200	83.8342	82.0257	90.4209	75.3473	65.2171	71.8258	76.5691	75.8909	78.4949	86.8155	84.3271	72.4292
92.0200	99.4797	100.1645	108.5870	131.6084	110.1632	107.0558	92.2038	80.1441	79.4763	65.1526	77.5001	77.9373
92.0200	82.3449	81.2572	79.2088	71.7598	63.7516	60.9444	64.4526	78.7550	77.2440	74.1856	70.7740	76.2252
92.0200	93.5493	106.9606	114.2666	107.2499	120.2684	148.4802	195.7616	174.5941	159.7295	179.7404	162.9445	213.2536
92.0200	83.1508	77.6790	72.7743	87.9439	73.8113	83.9707	97.5543	100.0643	129.6245	129.6993	112.7514	124.1679
92.0200	94.7428	113.6892	105.8575	123.5283	147.7772	137.3676	130.5886	135.8200	166.2885	133.0513	160.9442	180.4984
92.0200	91.9339	106.5833	111.4100	116.3989	121.7947	119.2333	133.9659	123.5023	152.9353	163.4568	197.9977	206.0660
92.0200	111.4184	98.6027	75.4237	75.9278	67.6246	59.4339	54.3967	51.1985	55.1387	47.8855	48.2377	32.7410
92.0200	81.3055	82.5167	104.2496	81.2171	96.6903	91.6767	90.0475	102.9414	89.7395	80.9487	78.1564	70.3480
92.0200	110.6157	111.0826	124.3771	111.6110	103.2768	106.7868	127.1154	132.7597	123.1827	108.9676	120.2684	174.7080
92.0200	90.1311	80.6328	77.3798	76.5107	72.9005	66.8326	67.8736	81.8466	81.3953	83.5355	77.0649	76.0294
92.0200	111.1371	110.4921	98.8344	100.0796	105.9000	97.9534	112.5448	118.9816	123.0486	129.7803	123.8380	158.2230
92.0200	88.7891	86.8228	91.9796	96.6096	96.2294	82.8345	78.7404	60.2748	64.0521	49.8891	49.3520	65.5930
92.0200	119.1710	97.8142	136.9544	137.0223	177.7344	189.6081	176.4152	146.9734	131.4835	107.1811	101.0571	113.0831
92.0200	92.0458	131.5208	139.4567	136.7715	127.1410	124.4408	152.9808	152.6093	161.2668	154.6650	140.3662	160.4949
92.0200	94.6091	100.7911	84.9302	85.8911	107.6024	87.2599	81.4324	98.4511	90.3520	73.7563	91.2419	130.7452
92.0200	74.4742	75.7523	98.4711	86.1107	92.8926	100.5596	101.3336	114.2245	110.8352	148.8380	190.0230	206.8274
92.0200	94.9796	111.3911	100.1965	98.8659	108.0910	123.9543	134.3051	117.1178	138.5981	161.7434	118.1510	129.1213
92.0200	81.0449	61.2582	54.1745	47.2066	45.1956	37.1277	35.7298	35.8923	38.8548	38.2267	40.9419	38.9894
92.0200	102.8138	137.1511	133.0924	143.3881	172.7632	150.5398	132.2849	113.4772	110.2082	96.4796	92.3040	87.3632
92.0200	93.8299	59.5064	56.9395	58.9521	61.3350	55.5754	48.9321	42.7666	32.5576	32.9386	25.8913	22.8937
92.0200	85.7302	76.4588	73.4951	60.8967	62.6644	40.6007	32.5620	34.8345	32.8168	38.9168	37.9684	36.7333
92.0200	117.7467	118.7715	108.2375	100.5265	93.7600	96.9461	95.1466	89.7344	81.7899	85.6636	74.0027	96.1807
92.0200	89.1404	67.9500	55.9759	66.4561	77.1896	86.3873	81.1826	72.8632	68.6872	75.6588	77.0025	70.3642
92.0200	82.4953	74.4290	80.5468	70.2153	57.7425	70.4527	66.5069	54.9906	59.7382	59.6180	61.6485	76.0927
92.0200	91.1141	112.7258	119.2074	156.3446	140.0547	185.1283	190.4804	274.8149	314.7730	340.9501	370.4679	329.3493
92.0200	95.7980	98.0868	92.1607	102.3360	114.0294	101.9244	105.9324	101.2242	72.1834	67.3893	90.0881	80.6005
92.0200	82.9654	89.5819	103.2192	85.4727	94.1387	77.2338	82.8633	93.5628	122.5595	83.6659	70.3222	70.5164
92.0200	79.0550	81.0168	120.1184	154.4665	150.8338	137.4588	96.5465	103.7131	116.5859	117.5004	123.9159	154.4125
92.0200	75.0080	84.5486	83.7720	62.7128	56.9940	69.8183	61.5715	78.5247	61.6560	48.4500	67.9675	57.6478
92.0200	91.1953	88.8916	74.8011	86.4670	69.1813	74.3360	77.6688	90.5802	98.2683	153.6183	205.7537	143.6589
92.0200	81.6953	66.7397	70.3289	76.2641	132.8844	121.4260	155.9688	159.6369	187.1105	213.8772	209.1734	196.4052
92.0200	88.2625	99.6449	125.9690	136.0106	122.2679	97.5400	115.3078	100.7659	120.3203	128.3590	113.4790	111.2859
92.0200	96.9008	87.0288	111.4694	150.3143	146.9011	168.3245	172.0058	216.3429	224.3953	213.3826	275.4776	228.7173

APÉNDICES

B =

112.8085
1.5194
0.0095

B =

133.1407
2.3492
0.0235

B =

96.1375
0.8913
-0.0009

B =

102.7259
1.1356
0.0039

B =

101.0444
1.1039
0.0033

B =

116.3272
1.6287
0.0110

B =

178.0820
3.7364
0.0459

B =

142.1923
2.2969
0.0210

B =

-0.7843
-1.4901
-0.0285

B =

156.9140
2.9286
0.0323

B =

-212.1636
-6.1767
-0.0820

NaN	0	0	0	0	0	0	0	0	18.3589	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	25.4536	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	27.7194	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	60.3714	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	13.2758	0
NaN	0	0	0	0	24.1575	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	34.0528	0	0	0	0	0	0	0	0	0
NaN	33.6686	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	36.6313	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	35.6229	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	35.6874	0	0
NaN	0	0	0	0	37.0884	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	27.0287	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	33.2154	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	19.8913	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	27.9395	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	50.9509	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	27.0837	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	39.5818	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	41.3336	0	0	0	0	0	0	0	0	13.4768
NaN	0	0	0	0	38.1756	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	32.8900	0	0	0	0	0	0	0	26.8373	0
NaN	0	0	0	0	43.0975	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	33.4507	0	0
NaN	0	0	0	0	0	23.6062	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	43.8460	0	0	0	0	0	0
NaN	0	0	0	0	31.6587	0	0	0	0	0	0
NaN	0	34.1003	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0
NaN	0	0	0	0	0	0	0	0	0	0	0

M =

APÉNDICES

```
Ydesc =  
 15.0310  
    0  
 23.0313  
    0  
 27.1706  
 49.4280  
    0  
 10.4431  
    0  
 21.8586  
    0  
    0  
 32.7176  
 33.0019  
    0  
 35.1950  
 32.2330  
 29.2184  
 33.5590  
    0  
 24.4565  
    0  
    0  
 30.0545  
 16.2856  
    0  
 25.2807  
    0  
 41.7151  
    0  
    0  
 22.1743  
    0  
    0  
 38.0298  
 10.6012  
 39.7129  
 34.5427  
 21.5375  
 31.6004  
 38.9962  
    0  
 27.3871  
 20.9368  
    0  
 39.6735  
 28.6459  
 32.7632  
    0  
    0
```

```
valor =  
 17.3456
```

```
ErrorStd =  
 2.2592
```

```
>>
```