



UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS

ESTUDIO Y APLICACIÓN DEL MODELO DE REGRESIÓN CUANTÍLICA

(Modalidad: Tesis de Grado)

RICHARD JOSÉ ROJAS

TRABAJO DE GRADO PRESENTADO COMO REQUISITO PARCIAL
PARA OPTAR AL TÍTULO DE LICENCIADO EN MATEMÁTICA

CUMANÁ, 2013

ESTUDIO Y APLICACIÓN DEL MODELO DE REGRESIÓN
CUANTÍLICA

APROBADO POR:

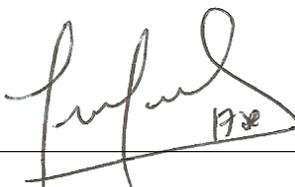


Profa. Nancy Ruiz

Asesor Académico



Jurado Principal



Jurado Principal

DEDICATORIA

A Dios, a Quien siempre he tenido presente en mi vida.

A mi madre, Julia, por darme la vida, la crianza, los valores y principios que día a día me inculcó.

A mis tías y tío, por su apoyo moral y económico durante la realización de mis estudios.

AGRADECIMIENTOS

Quiero agradecer a todo aquel que me ha proporcionado información durante mis investigaciones, en especial, al Lcdo. en Informática, Irving Martínez, a quien debo la realización del Capítulo IV y parte del Capítulo I. Sin la instrucción suministrada por él en cuanto al manejo del software R, me hubiese sido difícil realizar dichos Capítulos.

A los Profesores Tatjana Kecojević y Daniel Naiman, por su disposición para facilitarme los detalles de los códigos utilizados para realizar los ejemplos de minimización en el Capítulo I.

A la Profesora Nancy Ruiz, por su contribución y tiempo invertido durante la investigación.

Al Profesor Jacques Laforgue, por su colaboración en algunas demostraciones matemáticas y orientaciones durante la realización de la tesis.

Al Profesor Jaime Avendaño, por haber suministrado los datos de Sismología y recursos literarios acerca de la materia.

ÍNDICE

	Pág.
LISTA DE TABLAS	IX
LISTA DE FIGURAS	XI
RESUMEN	XII
INTRODUCCIÓN	1
1 DEFINICIONES BÁSICAS	7
1.1 Cuantil	8
1.2 Soluciones a Determinados Problemas de Minimización.	9
1.3 Función Chequeo	20
1.4 Cuantil como Solución a Problemas de Minimización.	23
1.5 Propiedades de los Cuantiles	31
1.6 Regresión Cuantílica	33
1.7 Usos de la Regresión Cuantílica.	35
2 ESTIMADORES DE LA REGRESIÓN CUANTÍLICA	36
2.1 Estimación de los Parámetros de la Regresión Cuantílica.	37
2.2 Propiedades de los Estimadores de la Regresión Cuantílica	41
2.3 Soluciones Básicas	45
2.4 Direcciones de la Pendiente	46
2.5 Condiciones para la Optimización	47
2.6 Holgura Complementaria	50
2.7 Dualidad	51
2.8 Método Simplex para la Regresión Cuantílica	52
2.9 Condición del Subgradiente	60
3 INFERENCIA SOBRE LOS ESTIMADORES DE LA REGRESIÓN CUANTÍLICA	63
3.1 Asintóticos de la Regresión Cuantílica con Errores I.I.D.	65

3.2	Asintóticos de la Regresión Cuantílica con Errores no I.I.D.	66
3.3	Estimación del Escalar Esparcido (<i>Sparsity</i>)	67
3.4	Estimación de la Matriz de Varianza-Covarianza para Modelos con Error no I.I.D.	72
3.5	Intervalos de Confianza para la Regresión Cuantílica Basados en la Transformación Rango (<i>Rank scores</i>)	74
3.6	Remuestreo del Residual.	78
3.7	Remuestreo del par (x,y).	80
3.8	Remuestreo sobre la Condición del Subgradiente.	81
3.9	Coefficiente de Determinación.	83
4	APLICACIÓN DE LA REGRESIÓN CUANTÍLICA	85
4.1	Ejemplo 1: Datos Engel (1857).	85
4.2	Estimación de los Datos Emitidos por las Agencias Sismológicas. .	95
4.3	Ejemplo 2: Magnitudes Sísmicas de ISC vs. USCGS.	96
4.4	Ejemplo 3: Magnitudes Sísmicas de ISC vs. NAO.	100
4.5	Ejemplo 4: Latitudes Sísmicas de ISCJB vs. TRN.	104
	CONCLUSIONES	108
	BIBLIOGRAFÍA	109
A	APÉNDICE	116
A.1	Cálculo de la derivada direccional de f en forma vectorial	117

LISTA DE TABLAS

4.1	Estimadores de la regresión cuantílica para las variables ingresos familiar y gastos de comida en el cuantil 0,5.	86
4.2	Estimadores e intervalos de confianza de la regresión cuantílica obtenidos usando el método Transformación Rango para las variables ingreso familiar y gastos de comida en los cuantiles 0,05 al 0,95.	89
4.3	Resumen de los estimadores de la regresión cuantílica para las variables ingresos familiar y gastos de comida en cada uno de los cuantiles presentados anteriormente.	90
4.4	Inferencias de la regresión cuantílica por el método del escalar esparcido con el parámetro de suavizado de Hall y Sheather (1988) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.	90
4.5	Inferencias de la regresión cuantílica por el método Remuestreo del residual para las variables ingresos familiar y gastos de comida en los taus 0,05 al 0,95.	91
4.6	Inferencias de la regresión cuantílica por el método Remuestreo del par (x, y) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.	92
4.6	Inferencias de la regresión cuantílica por el método Remuestreo del par (x, y) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95. (Continuación)	93
4.7	Inferencias de la regresión cuantílica por el método Remuestreo de Parzen, et al. (1994) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.	93

4.7	Inferencias de la regresión cuantílica por el método Remuestreo de Parzen, et al. (1994) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95. (Continuación)	94
4.8	Bondad del ajuste para las variables ingreso familiar y gasto familiar en la regresión cuantílica.	94
4.9	Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las magnitudes de las ondas sísmicas de ISC y USCGS en los cuantiles 0,05 al 0,95.	97
4.10	Inferencias de la regresión cuantílica obtenidas usando el método Remuestreo del residual para las magnitudes de las ondas sísmicas de ISC y USCGS en los cuantiles 0,05 al 0,95.	98
4.11	Bondad del ajuste para las magnitudes de las ondas sísmicas de ISC y USCGS en la regresión cuantílica.	98
4.12	Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95.	101
4.13	Inferencias de la regresión cuantílica obtenidas usando el método <i>Bootstrap</i> para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95.	101
4.13	Inferencias de la regresión cuantílica obtenidas usando el método Remuestreo del residual para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95. (Continuación)	102
4.14	Bondad del ajuste para las magnitudes de las ondas sísmicas de ISC y NAO en la regresión cuantílica.	102
4.15	Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las latitudes sísmicas de ISCJB y TRN en los cuantiles 0,05 al 0,95.	104

4.16 Inferencias de la regresión cuantílica obtenidas usando el método Re- muestreo del residual para las latitudes sísmicas de ISCJB y TRN en los cuantiles 0,05 al 0,95.	105
4.17 Bondad del ajuste para las latitudes sísmicas de ISCJB y TRN en la regresión cuantílica.	106

LISTA DE FIGURAS

1.1	Gráfica de la media muestral como solución al problema de minimización. Elaborada con el software R usando los 10 valores muestrales.	15
1.2	Gráfica de la mediana muestral como solución al problema de minimización. Elaborada con el <i>software</i> R usando los 10 valores muestrales.	18
1.3	Gráfica de la función forma de V usada para describir la mediana como la solución a un problema de minimización. Elaborada con el <i>software</i> R usando el mínimo $m = 2, 13$.	20
1.4	Gráfica de la función chequeo ρ para la regresión cuantílica	21
1.5	Gráfica de la función chequeo cuando $\tau = 0,25$. Elaborada con el <i>software</i> R usando los 21 valores muestrales.	23
1.6	Gráfica del cuantil muestral $\tau = 0,25$. como una solución al problema de minimización. Elaborada con el <i>software</i> R usando los 7 valores muestrales.	31
3.1	Parámetro de Suavizado h_n . Las curvas punteadas representan el parámetro de Hall & Sheather y las curvas sólidas el parámetro de Bofinger para los cuantiles 0,50; 0,75 y 0,95.	70
3.2	Diferentes elecciones del Parámetro de Suavizado h_n para una muestra de tamaño $n=500$, basada en la distribución gaussiana con diferentes niveles de significancia $\alpha_1 = 0,05$; $\alpha_2 = 0,01$ y $\alpha_3 = 0,001$.	71
4.1	Datos Engel (1857). La figura muestra el diagrama de dispersión de los gastos de comida vs. el ingreso familiar para una muestras de 235 observaciones.	87

4.2	Datos Engel (1857). La figura muestra la recta de regresión lineal y las rectas de regresión cuantílicas de los gastos de comida vs. el ingreso familiar.	88
4.3	Rectas de Regresión Cuantílica y Regresión Lineal de las magnitudes de las ondas sísmicas (mb) de USCGS y ISC.	99
4.4	Rectas de Regresión Cuantílica y Regresión Lineal de las magnitudes de las ondas sísmicas (mb) de NAO y ISC	103
4.5	Rectas de Regresión Cuantílica y Regresión Lineal de las latitudes sísmicas de ISCJB y TRN	106

RESUMEN

En este trabajo se introducen los conceptos y las propiedades básicas de la regresión cuantílica propuestos por Koenker y Bassett (1978). La regresión cuantílica, a diferencia de la regresión de la media, produce modelos que no sólo se basan en la estimación de la tendencia central de los datos sino que estiman toda la distribución de la variable dependiente. Para determinar los estimadores y los errores estándares de los modelos de regresión cuantílica, se requiere el uso de cualquier software con paquetes estadísticos que utilicen los métodos de regresión cuantílica. Se consideró el software R para determinar las tablas de los estimadores y los errores estándares de una muestra de los datos del *International Seismological Centre* suministrados por el Centro de Sismología de la Universidad de Oriente.

INTRODUCCIÓN

Girdano (2003), citado por Albright (2009), estableció que un modelo matemático es una idea matemática diseñada para estudiar un sistema particular o fenómeno del mundo real. Los modelos matemáticos pueden tomar muchas formas diferentes que involucran ecuaciones, inecuaciones, ecuaciones diferenciales, matrices, lógica y cualquier tipo de ideas matemáticas.

Estos autores se refieren a que un modelo matemático puede ser entendido como una fórmula que permite predecir el comportamiento de un fenómeno de la vida real y en esto se ve su importancia dado que el propósito, según Fowler (1997), es formular una descripción de un mecanismo en términos cuantitativos y el análisis del modelo resultante orienta hacia los resultados que pueden ser probados junto con las observaciones. El modelo también guía hacia las predicciones que en caso de ser probadas, dan la autenticidad del modelo.

Una definición general del modelo estadístico, según Gouriéroux (1995), es un par (\mathbf{Y}, \mathbf{P}) donde \mathbf{Y} es el conjunto de las observaciones posibles y \mathbf{P} es una familia de distribuciones de probabilidad sobre \mathbf{Y} .

Entre los modelos estadísticos están los modelos de regresión múltiple que tienen por finalidad encontrar las relaciones existentes entre una variable determinada que se considera dependiente y otro conjunto de variables que entre sí no son funcionalmente dependientes. Determinar la relación existente entre una variable y un conjunto de otras es bastante complejo; pues teóricamente pueden existir numerosas funciones que las enlacen.

Cuando el objetivo es encontrar la mejor función que sea lineal en los parámetros, estamos en presencia de un modelo de regresión lineal.

Draper y Smith (1998) definieron un modelo de regresión lineal como una ecuación matemática $Y = \beta_0 + \beta_1 X + \varepsilon$, que provee una relación lineal entre

dos variables continuas, una aleatoria dependiente Y y la otra independiente X , no aleatoria. En este modelo, β_0 es la intercepción y β_1 es la pendiente. Los parámetros β_0 y β_1 se determinan examinando las posibles incidencias de Y y X , mientras que ε es el error aleatorio con $\varepsilon \sim N(0, \sigma^2)$.

En la práctica, es necesario encontrar estimadores para β_0 y β_1 digamos b_0 y b_1 , para así poder determinar el modelo de regresión lineal estimado

$$\hat{Y} = b_0 + b_1 X,$$

por el método de mínimos cuadrados. Este método es un procedimiento óptimo en muchos casos cuando los errores son independientes y tienen una distribución gaussiana.

El modelo de regresión lineal anterior es un caso especial del modelo de regresión lineal múltiple. Según Draper y Smith (1998), en el modelo de regresión lineal múltiple se supone que existe una relación lineal entre algunas variables aleatorias Y , las cuales son dependientes, y k variables no aleatorias independientes X_i , $i = 1, 2, \dots, k$. Estas últimas se conocen como variables explicativas debido a que se utilizan para explicar la variación de Y .

El modelo de regresión múltiple puede expresarse:

$$\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{X}_{i1} + \dots + \beta_k \mathbf{X}_{ik} + \varepsilon_i, \quad i = 1, \dots, n,$$

en el cual β_0 es la intercepción y β_1 es un parámetro asociado con \mathbf{X}_{i1} y así sucesivamente. Este modelo puede representarse en forma matricial

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

en la cual \mathbf{Y} es una matriz $n \times 1$ de variables dependientes, \mathbf{X} es una matriz $n \times (k+1)$ de variables independientes, $\boldsymbol{\beta}$ es una matriz $(k+1) \times 1$ de estimadores del modelo y $\boldsymbol{\varepsilon}$ es una matriz $n \times 1$ de errores.

Al igual que la regresión lineal simple, usando el método de mínimos cuadrados se encuentra un vector de estimadores $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ que minimiza a

la suma de los errores al cuadrado de todas las observaciones. Este vector de estimadores tiene la fórmula

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i y_i\end{aligned}$$

En los modelos de regresión, se asume los errores $\boldsymbol{\varepsilon}$ como una sucesión de variables idénticamente distribuidas con medias iguales a cero. Generalmente, la distribución que se asume es la normal. Sin embargo, no siempre se cumple el supuesto de normalidad. En particular, no resulta conveniente aplicar este método cuando se tiene asimetría en la distribución y existe la presencia de datos atípicos.

Los econométricos Koenker y Bassett (1978) introdujeron y desarrollaron el concepto de Regresión Cuantílica como una solución a este tipo de problemas, sabiendo que la presencia de datos atípicos, en el caso de la regresión simple, puede alterar notablemente los resultados.

En el método de mínimos cuadrados el objetivo es minimizar la suma de los errores al cuadrado y utilizar la media como estimador; mientras que la regresión cuantílica busca minimizar la suma de los valores absolutos de los residuales ponderados asimétricamente y utiliza los cuantiles como estimadores. Un cuantil es un valor que corresponde a una proporción específica de una muestra o población, una vez que han sido ordenados los datos.

Una de las definiciones importantes utilizadas por Koenker y Bassett (1978), en la regresión cuantílica, es la de cuantiles muestrales debido a que estos tienen la propiedad de ser robustos en presencia de datos atípicos. Estos autores minimizaron la función objetivo

$$\sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i'\boldsymbol{\beta}), \tag{1}$$

en la cual ρ_{τ} es la característica central, la cual sirve como función de chequeo, $\{y_1, \dots, y_n\}$ es una muestra aleatoria de la variable dependiente \mathbf{Y} y $\mathbf{x}_i \in \mathbb{R}^p$

es el vector de covariables correspondiente a la i -ésima observación y_i . Los estimadores $\hat{\beta}$ de β se conocen como regresión cuantílica. Una herramienta útil en la minimización de funciones objetivos es la programación lineal.

Koenker y d'Orey (1987) afirmaron que para encontrar los estimadores $\hat{\beta}$ en la ecuación (1), es necesario escribir dicha función como un problema de programación lineal y luego expresarlo en forma básica

$$\beta(h) = \mathbf{X}(h)^{-1}\mathbf{y}(h) \quad (2)$$

en la cual h es un elemento de un conjunto indexado \mathcal{H} , $\mathbf{X}(h)$ es una matriz singular de \mathbf{X} con filas $\{x_i : i \in h\}$ y $\mathbf{y}(h)$ es un p vector de coordenadas $\{y_i : i \in h\}$.

Una vez escrito el problema en forma básica, el uso de un método como el algoritmo Simplex permite pasar de una solución básica del conjunto de restricciones a otra de modo que el valor de la función objetivo decrezca continuamente hasta alcanzar la minimización.

Según los mismos autores, el algoritmo Simplex de Barrodale y Roberts (1974) utilizado en la regresión de la media puede ser usado en la regresión cuantílica. Este algoritmo se desarrolla en dos fases; en la fase I, se busca un vértice factible y en la fase II, se procede de un vértice a otro adyacente y se continua aplicando iteraciones del método Simplex hasta que se logre la optimización.

En el caso de la regresión cuantílica podemos elegir cualquier subconjunto h en la ecuación (2) y definir un punto básico factible para el problema, y de ese modo completar la fase I, mientras que para la fase II del algoritmo, la función objetivo de la ecuación (1) debe satisfacer ciertas condiciones para lograr la optimización.

Existen muchos programas estadísticos que traen incorporado el algoritmo Simplex, uno de ellos es el Software R, una versión libre del lenguaje S, introducida por Becker, et al. (1988), y utilizado en la regresión cuantílica para hacer

gráficas y obtener el valor de los estimadores. Según Koenker (2012), el paquete *quantreg* almacenado en la librería de software R ejecuta por defecto una variante del algoritmo Simplex desarrollada por Barrodale y Roberts (1974).

De acuerdo a Koenker (2005), la teoría asintótica de inferencia estadística utilizada en la regresión de la media condicional puede extenderse a la regresión cuantílica siempre que se mantengan las nociones del error independiente e idénticamente distribuido (i.i.d). Uno de los procedimientos que se puede aplicar para hacer inferencias es el método asintótico con error i.i.d.

Alternativamente, este autor estableció varios métodos que no toman en cuenta la distribución de los residuales para producir las tablas de los errores estándares y construir los intervalos de confianza; entre estos, se tienen: El método Estimación del Escalar Esparcido (*Sparsity*), el cual se basa en la estimación de la función esparcida, que a su vez depende de un parámetro de suavizado y utiliza la distribución asintótica de los estimadores para construir los intervalos de confianza. El método de *Bootstrap*, el cual aplica varias veces el remuestreo sobre un conjunto de datos, calcula el estimador para cada muestra y encuentra la desviación estándar de la distribución de ese estimador. Por último, el método Transformación Rango (*Rankscore*), cuyo objetivo es minimizar una distancia que depende de la función *score* de Hájek y Šidák (1967) para encontrar un vector de estimadores robustos a los datos atípicos y luego invertir la prueba de Rango para determinar los intervalos de confianza. Este método asume que los errores son independientes e idénticamente distribuidos.

De acuerdo a Koenker (2012), el método utilizado en R por defecto para construir los intervalos de confianza, cuando la muestra es menor o igual a 1000, es el de la Transformación Rango. Pero cuando la muestra es superior a 1000, el método por defecto usado en R estima automáticamente a la matriz de covarianza y produce la tabla de los errores estándares para cada cuantil .

Hao & Naiman (2007) afirmaron que la técnica del *bootstrap* es preferible

a la del método asintótico con errores i.i.d. debido a que las condiciones de ese procedimiento asintótico usualmente no se mantienen, y en caso de mantenerse es complicado encontrar el error estándar y los cambios en la sesgidez. En cambio el método de *Bootstrap* ofrece flexibilidad para obtener el error estándar y el intervalo de confianza para cualquier cuantil.

De manera similar, el *software* estadístico Stata puede ser usado para obtener los estimadores, los errores estándares, los intervalos de confianza y la bondad del ajuste de la regresión cuantílica usando la función `qreg`. Los errores estándares e intervalos de confianza calculados con este comando asumen que la muestra tiene errores independientes e idénticamente distribuidos. Si se quiere obtener los errores estándares para una muestra donde la distribución no es i.i.d; se utiliza el método *Bootstrap* mediante el comando `bsreg`. El número de repeticiones del *Bootstrap* que se obtienen por defecto con este comando es 20, si se quiere un mayor número de repeticiones se puede usar la opción `reps(#)`.

En este trabajo se estudia el modelo de Regresión Cuantílica propuesto por Koenker y Bassett (1978); estos autores utilizaron los cuantiles condicionales para producir una familia de funciones condicionales cuantílicas que permitan obtener estimaciones más confiables que las del modelo de regresión de la media en presencia de datos atípicos. Para lograr este propósito se realizan los siguientes objetivos específicos:

- Enunciar los fundamentos básicos de la Regresión Cuantílica.
- Determinar los estimadores de la Regresión Cuantílica.
- Realizar inferencias sobre los estimadores de la Regresión Cuantílica.
- Aplicar el modelo de la Regresión Cuantílica a un conjunto de datos.

Cada objetivo se lleva acabo en un Capítulo propio, y la tesis finaliza con unas Conclusiones.

CAPITULO 1

DEFINICIONES BÁSICAS

En el ajuste de los modelos de regresión lineal, el método de mínimos cuadrados es útil cuando los errores tienen una distribución normal y existe poca dispersión. Sin embargo, los estimadores son muy sensibles a los datos atípicos. Estos se pueden encontrar en la variable dependiente o en la variable independiente, si los datos atípicos se encuentran en la variable independiente, el par (X, Y) se conoce como punto de influencia.

En el problema de la estimación de la elasticidad de La Tierra propuesto por Boskovich (1760), las condiciones para aplicar el método de mínimos cuadrados no se satisfacen. En este caso, para obtener los estimadores del modelo, el autor consideró reemplazar la minimización de la suma de los errores al cuadrado por la minimización de la suma de valores absolutos de los errores. Para finales del siglo XIX, Edgeworth (1888), siguiendo las ideas de Boskovich, propuso minimizar la suma de los residuales absolutos para la intercepción y la pendiente del modelo.

Los estimadores, de la recta de regresión de la mediana, obtenidos por Boskovich y Edgeworth resultaron ser robustos a los datos atípicos en la variable dependiente pero poco confiables si los datos atípicos se encuentran en la variable independiente, donde el efecto de los puntos de influencia es tan fuerte como en la recta de regresión por el método de mínimo cuadrados.

La idea de la regresión cuantílica tiene su origen en los estudios de la regresión de la mediana realizados por Boskovich (1760) y Edgeworth (1888). Koenker y Bassett (1978) extendieron las propiedades de los estimadores del cuantil 0,5 a los otros cuantiles condicionales. El modelo de la regresión cuantílica al utilizar los cuantiles condicionales como estimadores produce una familia de funciones condicionales cuantílicas que permiten modelar cualquier posición en la distribu-

ción de la variable dependiente mientras que el modelo de regresión lineal sólo provee información acerca de la media condicional. Además, cuando la distribución es muy sesgada, la media condicional ofrece poca información para estudiar el modelo, mientras que la regresión cuantílica permite un análisis más detallado de la variable dependiente en cada cuantil.

En el modelo de regresión lineal para obtener los estadísticos inferenciales se requiere de las nociones de normalidad, mientras que en la regresión cuantílica los estadísticos tienen distribución libre.

En este Capítulo se estudian los conceptos de Cuantil, Datos Atípicos, Regresión Cuantílica y Robustez. Además, se demuestra que la media, μ , minimiza el promedio al cuadrado de las desviaciones $E[(Y - \mu)^2]$, la mediana minimiza al promedio de las distancias absolutas $E[|Y - m|]$ y se generaliza a cualquier otro cuantil. Finalmente se mencionan los casos donde es útil la regresión cuantílica.

1.1 Cuantil

Definición 1.1. (Cuantil)

Un cuantil es un valor que corresponde a una proporción específica de una muestra o población, una vez que han sido ordenados los datos. Formalmente, para una variable aleatoria continua X , el τ -ésimo cuantil poblacional es aquel valor x_τ tal que X es menor o igual que x_τ con una probabilidad τ , así:

$$F_X(x_\tau) = P[X \leq x_\tau] = \tau \quad (1.1)$$

en la cual F_X es la función de distribución acumulada de X .

Otra forma de expresar el τ -ésimo cuantil, es la siguiente:

$$x_\tau = F_X^{-1}(\tau) \quad (1.2)$$

x_τ es tal que constituye la inversa de la función F_X para una probabilidad τ .

Se observan dos posibles escenarios. Un escenario sería, si la función de distribución acumulada F_X es monótona creciente, entonces los cuantiles están bien definidos para todo $\tau \in (0, 1)$. Sin embargo, si una función de distribución acumulada F_X no es estrictamente monótona creciente, existen algunos τ para el cual un único cuantil no puede ser definido. En este caso, se usa el valor más pequeño que x puede tomar para una probabilidad dada τ . Es decir, x_τ , puede ser descrito como el τ -ésimo cuantil de X :

$$x_\tau = F_X^{-1}(\tau) = \inf\{x \in \mathbb{R} \mid F_X(x) \geq \tau\}. \quad (1.3)$$

En consecuencia, un cuantil al igual que la media μ y la mediana m , puede ser considerado como una solución a ciertos problemas de minimización.

1.2 Soluciones a Determinados Problemas de Minimización.

Mediante el siguiente lema se mostrará la propiedad de minimización de la media, μ .

Lema 1.1. *Si Y es una variable aleatoria con segundo momento finito entonces*

$$f(b) = E[(Y - b)^2]$$

es mínimo cuando $b = E[Y] = \mu$, donde μ es la media de la distribución de Y .

Demostración.

$$\begin{aligned} \text{Sea } E[(Y - \mu)^2] &= E[Y^2] - 2E[Y]\mu + \mu^2 \\ &= \mu^2 - 2E[Y]\mu + (E[Y])^2 - (E[Y])^2 + E[Y^2] \\ &= (\mu - E[Y])^2 + (E[Y^2] - (E[Y])^2) \\ &= (\mu - E[Y])^2 + \text{Var}(Y). \end{aligned} \quad (1.4)$$

Debido a que la $\text{Var}(Y) \geq 0$, la minimización para la ecuación (1.4) se obtiene encontrando el valor de μ para el cual la expresión $(\mu - E[Y])^2$ es cero, en este caso $\mu = E[Y]$. Así, la media satisface el problema de minimización. \square

La mediana, m , tiene una propiedad de minimización similar a la media. Pero en vez de usar el cuadrado de la distancia, se mide la distancia entre los valores de Y y m , usando el promedio de las distancia absolutas de m a los valores de Y , es decir, $E|Y - m|$. Minimizando $E|Y - m|$ se encuentra el valor de m . La función $|Y - m|$ es convexa, así que para obtener la solución de minimización se debe encontrar un punto donde la derivada con respecto a m sea cero o donde las derivadas direccionales tengan signos diferentes pero con la misma magnitud. Para observar, por qué la mediana puede ser definida como un problema de minimización se enuncia el lema 1.2.

Lema 1.2. *Si Y es una variable aleatoria con primer momento finito y función de densidad de probabilidad f entonces $E|Y - b|$ es mínimo cuando $b = m$, donde m es la mediana de la distribución de Y .*

Demostración.

$$\begin{aligned}
 \text{Sea } E|Y - m| &= \int_{-\infty}^{+\infty} |y - m|f(y) dy \\
 &= \int_{-\infty}^m |y - m|f(y) dy + \int_m^{+\infty} |y - m|f(y) dy \\
 &= \int_{-\infty}^m (m - y)f(y) dy + \int_m^{+\infty} (y - m)f(y) dy \quad (1.5)
 \end{aligned}$$

$$\text{Con } |y - m| = \begin{cases} y - m & \text{si } y \geq m \\ m - y & \text{si } y < m. \end{cases}$$

Diferenciando con respecto a m y haciendo la derivada parcial igual a cero se obtiene la solución al problema de minimización.

Derivando parcialmente el primer término de (1.5) se obtiene:

$$\begin{aligned}
\frac{\partial}{\partial m} \left[\int_{-\infty}^m (m-y)f(y) dy \right] &= (m-y)f(y) \Big|_{y=m} + \int_{-\infty}^m \frac{\partial}{\partial m} (m-y)f(y) dy \\
&= \int_{-\infty}^m f(y) dy \\
&= F(m)
\end{aligned}$$

De manera análoga para el 2^{do} término (1.5)

$$\begin{aligned}
\frac{\partial}{\partial m} \left[\int_m^{+\infty} (y-m)f(y) dy \right] &= (y-m)f(y) \Big|_{y=m} + \int_m^{+\infty} \frac{\partial}{\partial m} (y-m)f(y) dy \\
&= - \int_m^{+\infty} f(y) dy
\end{aligned}$$

Puesto que

$$\int_m^{+\infty} f(y) dy = \int_{-\infty}^{+\infty} f(y) dy - \int_{-\infty}^m f(y) dy,$$

Por las propiedades de las funciones de densidad y distribución se tiene que

$$\int_{-\infty}^{+\infty} f(y) dy = 1 \quad \text{y} \quad \int_{-\infty}^m f(y) dy = F(m)$$

luego,

$$- \int_m^{+\infty} f(y) dy = -(1 - F(m)). \tag{1.6}$$

Sustituyendo (1.6) y (1.6) en (1.5) se obtiene

$$\begin{aligned}
\frac{\partial}{\partial m} \int_{-\infty}^{+\infty} |y-m| f(y) dy &= F(m) + F(m) - 1 \\
&= 2F(m) - 1
\end{aligned}$$

haciendo

$$2F(m) - 1 = 0$$

se tiene

$$F(m) = \frac{1}{2}$$

Luego la mediana satisface el problema de minimización. \square

Asimismo, la media y la mediana muestrales para una muestra de tamaño n , pueden ser vistas como soluciones a los problemas de minimización:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 \quad \text{y} \quad \frac{1}{n} \sum_{i=1}^n |y_i - m|,$$

respectivamente.

Se observa que el punto μ minimiza al promedio del cuadrado de la distancia

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2$$

En efecto,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \mu + \bar{y} - \bar{y})^2 \\ &= \frac{1}{n} \sum_{i=1}^n [(y_i - \bar{y}) + (\bar{y} - \mu)]^2 \\ &= \frac{1}{n} \sum_{i=1}^n [(y_i - \bar{y})^2 - 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2] \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{2}{n} \sum_{i=1}^n (y_i - \bar{y})(\bar{y} - \mu) \\ &\quad + \frac{1}{n} \sum_{i=1}^n (\bar{y} - \mu)^2 \end{aligned} \tag{1.7}$$

Desarrollando el 2^{do} y 3^{er} término de la ecuación (1.7) se tiene

$$\begin{aligned}
\frac{2}{n} \sum_{i=1}^n (y_i - \bar{y})(\bar{y} - \mu) + \frac{1}{n} \sum_{i=1}^n (\bar{y} - \mu)^2 &= \frac{2}{n} \left[\sum_{i=1}^n (y_i \bar{y} - y_i \mu - (\bar{y})^2 + \bar{y} \mu) \right] \\
&+ \frac{1}{n} \left[\sum_{i=1}^n (\bar{y})^2 - 2\mu \bar{y} + \mu \right] \\
&= \frac{2\bar{y}}{n} \sum_{i=1}^n y_i - \frac{2}{n} \sum_{i=1}^n y_i \mu - \frac{2n(\bar{y})^2}{n} \\
&+ \frac{2n\bar{y}\mu}{n} + \frac{n(\bar{y})^2}{n} - \frac{2\mu n\bar{y}}{n} + \frac{n\mu^2}{n} \\
&= (\bar{y})^2 - 2\mu\bar{y} + \mu^2 \\
&= \frac{1}{n} \sum_{i=1}^n (\mu - \bar{y})^2 \tag{1.8}
\end{aligned}$$

Sustituyendo (1.8) en la ecuación (1.7) se tiene

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 &= \frac{1}{n} \sum_{i=1}^n (\mu - \bar{y})^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\
&= (\mu - \bar{y})^2 + S_y^2,
\end{aligned}$$

en la cual \bar{y} es la media muestral y $S_y^2 \geq 0$ es la varianza muestral. Esta última expresión se minimiza cuando $(\mu - \bar{y})^2 = 0$, es decir, $\mu = \bar{y}$.

Para ilustrar gráficamente que la media muestral puede ser vista como un problema de minimización se utilizará el *software* R.

Considere una muestra de 10 valores

$$y_i : \{0, 50; 1, 10; 1, 60; 1, 73; 2, 13; 2, 93; 3, 34; 4, 06; 5, 49; 7, 12.\}$$

1. Definir un vector de valores muestrales con la función c.

```
y<-c(0.5, 1.10, 1.6, 1.73, 2.13, 2.93, 3.34, 4.06, 5.49, 7.12)
```

- Determinar la media muestral y mostrar su valor en pantalla .

```
ymean<-mean(y)
ymean
3
```

- Crear una función para calcular el promedio de la distancia cuadrada de los puntos a μ .

```
mean.squared.dist<-function(u)
{
sum((y-u)^2)/10
}
```

- Crear un conjunto de valores uniformemente separados entre cero y siete coma dos.

```
u<-seq(0, 7.2, by = 0.001)
```

- Aplicar la función definida anteriormente para cada uno de los puntos generados en paso anterior.

```
w<-sapply(u, mean.squared.dist)
```

- Graficar cada uno de los puntos con su valor correspondiente evaluado en la función.

```
plot(u, w, type = "l", lwd = 2, ylim = c(0,20), xlab = "Valores Muestrales", ylab = "Distancia cuadrada de la media a los puntos de u")
```

- Graficar los puntos que corresponden a la distancia absoluta de la media para cada uno de los valores en el vector "y".

```
for(i in seq(1,10))
{
points(y[i], mean.squared.dist(y[i]), pch=19)
}
```

8. Crear un segmento de recta para visualizar el punto donde ocurre la minimización.

```
segments(ymean, 0, ymean, 20)
```

9. Añadir una etiqueta a manera de leyenda mostrando el valor de la media en la gráfica.

```
text(3.3, 0, "y = 3")
```

La gráfica del promedio de la distancia cuadrada de los puntos muestrales a un punto dado μ se muestra en la figura (1.1)

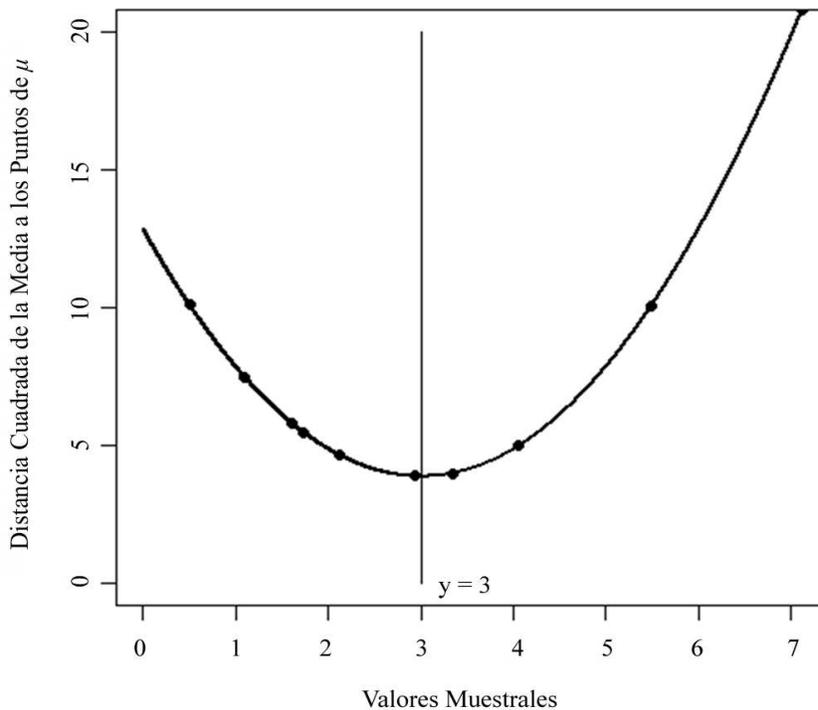


Figura 1.1: Gráfica de la media muestral como solución al problema de minimización. Elaborada con el *software* R usando los 10 valores muestrales.

De manera similar, se puede demostrar a nivel muestral que la mediana

minimiza el promedio de la distancia absoluta de los puntos muestrales a m usando

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

Entonces, para cualquier muestra, la función definida por

$$f(m) = \frac{1}{n} \sum_{i=1}^n |y_i - m|$$

es la suma de las funciones que tienen forma de V

$$f_i(m) = \frac{|y_i - m|}{n},$$

donde m es la mediana.

La función f_i toma un valor mínimo igual a cero cuando $m = y_i$. Pero no es diferenciable en ese punto mínimo; tiene una derivada direccional igual a $-1/n$ en la dirección negativa y $1/n$ en la dirección positiva. Haciendo la sumas de estas funciones, la derivada direccional de f con respecto a m es $\frac{r-s}{n}$ en la dirección negativa y $\frac{s-r}{n}$ en la dirección positiva, (ver Apéndice (A.1)) donde s es el número de puntos a la izquierda de m y r es el número de puntos a la derecha de m . Se sigue que el mínimo de f ocurre cuando el número de puntos a la izquierda es igual al número de puntos a la derecha, es decir, cuando m es una mediana.

Con el mismo vector de valores muestrales y_i definido en el ejemplo anterior, se ilustrará gráficamente que la mediana muestral puede ser vista como un problema de minimización.

1. Determinar la mediana muestral y mostrar su valor en pantalla.

```
ymedian<-median(y)
```

```
ymedian
```

```
2.53
```

2. Crear una función para calcular el promedio de la distancia absoluta de los puntos a m .

```

mean.absolute.dist<-function(u)
{
sum(abs(y-u))/10
}

```

3. Crear un conjunto de valores uniformemente separados entre cero y siete.

```
u<-seq(0, 7, by = 0.001)
```

4. Aplicar la función definida anteriormente para cada uno de los puntos generados en paso anterior.

```
w<-sapply(u, mean.absolute.dist)
```

5. Construir la región de la gráfica con sus ejes etiquetados.

```
plot(u, w, type = "null", xlab = "Valores Muestrales", ylab = "Distancia absoluta de la media a los puntos de m")
```

6. Graficar los puntos que corresponden al promedio de la distancia absoluta para cada uno de los valores en el vector "y".

```

for (i in seq(1,10))
{
points(y[i], mean.absolute.dist(y[i]), pch = 19)
}

```

7. Graficar los segmentos de recta comprendidos entre cada par de puntos adyacentes.

```

for (i in seq(2,10))
{
u<-seq(y[i-1], y[i], length = 10)
w<-sapply(u, mean.absolute.dist)
lines(u, w)
}

```

8. Crear un segmento de recta para visualizar el punto donde ocurre la minimización.

```
segments(ymedian, 0, ymedian, 4.5)
```

9. Añadir una etiqueta a manera de leyenda mostrando el valor de la mediana en la gráfica.

```
text(3.2, 2, "m = 2.53")
```

La gráfica de esta función para los diez puntos muestrales dados se muestra en la figura (1.2)

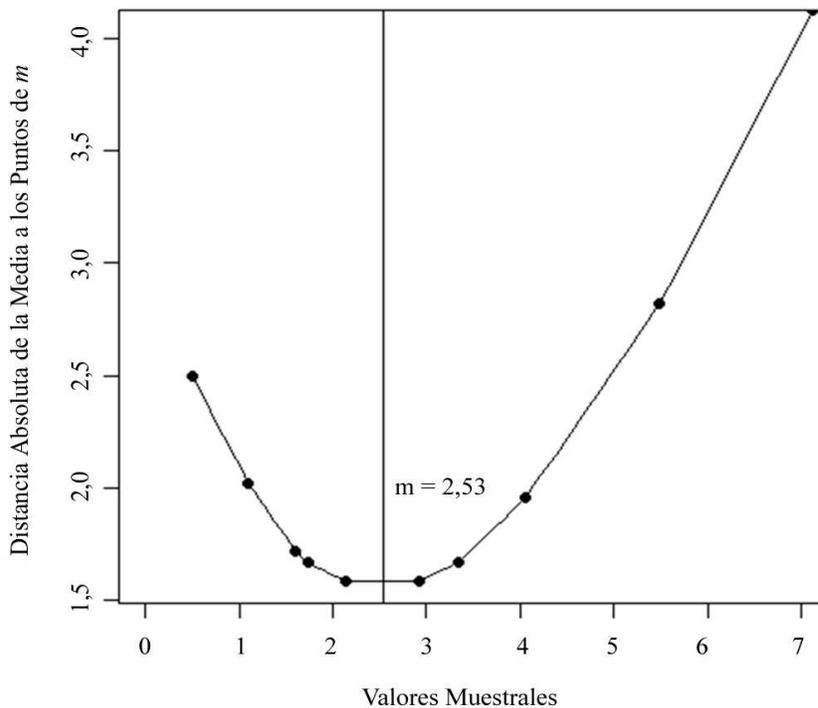


Figura 1.2: Gráfica de la mediana muestral como solución al problema de minimización. Elaborada con el *software* R usando los 10 valores muestrales.

En comparación con la función graficada en (1.1), la figura (1.2) permanece convexa y parabólica; la función de la figura (1.2) es lineal por segmentos y presenta una pendiente que cambia en cada punto muestral. Un valor mínimo de la

función mostrada en la figura (1.2) coincide con el valor muestral de la mediana 2,53.

Para elegir el valor del mínimo se observa que en la figura (1.2) cualquier valor comprendido entre 2,13 y 2,93 sirve como mínimo. En este caso, se puede elegir $m = 2,13$. Luego, utilizando el vector de valores muestrales y_i se realiza un proceso similar al del ejemplo anterior.

1. Crear una función para calcular el valor de $f_i(m)$.

```
valorabs<-function(y)
{
  abs(y-2.13)/10
}
```

2. Crear un conjunto de valores uniformemente separados entre cero y siete.

```
u<-seq(0, 7, by = 0.001)
```

3. Aplicar la función definida anteriormente para cada uno de los puntos generados en paso anterior.

```
w<-sapply(u, valorabs)
```

4. Graficar cada uno de los puntos con su valor correspondiente evaluado en la función.

```
plot(u, w, type = "l", lwd = 2, ylim = c(0,0.5), xlab = "Valores Muestrales", ylab = "fi(m)")
```

5. Añadir una etiqueta a manera de leyenda mostrando el valor del mínimo seleccionado.

```
text(2.8, 0, "y = 2.13")
```

La gráfica de $f_i(m)$ vs. los valores muestrales del vector y_i se muestra en la figura (1.3)

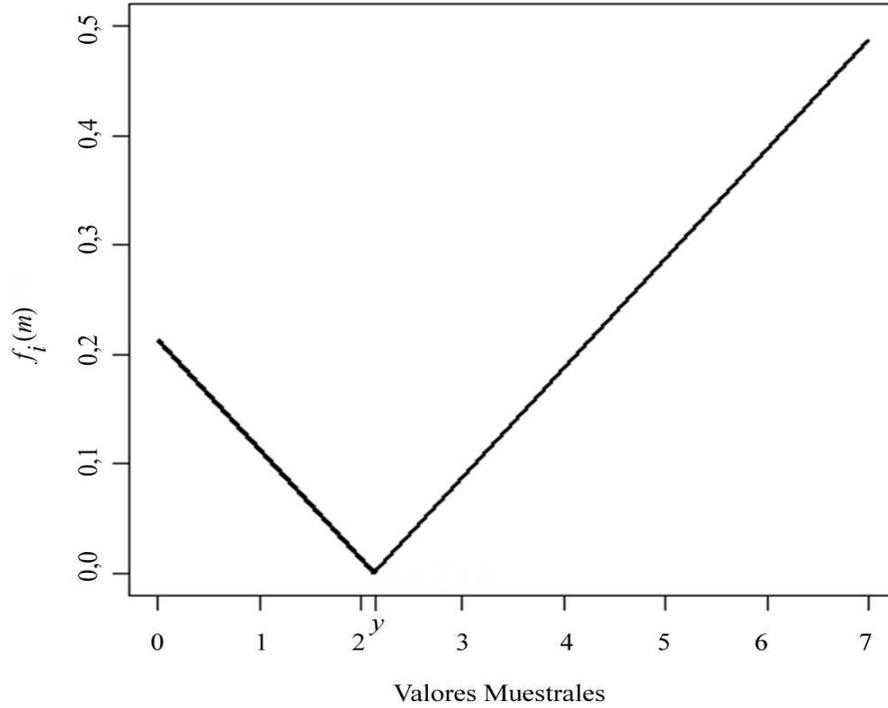


Figura 1.3: Gráfica de la función forma de V usada para describir la mediana como la solución a un problema de minimización. Elaborada con el *software* R usando el mínimo $m = 2, 13$.

En la gráfica (1.3) un valor mínimo para f_i se presenta cuando $y_i = 2, 13$.

1.3 Función Chequeo

Definición 1.2. (Función Chequeo)

Para $\tau \in (0, 1)$ se define la función chequeo de las siguientes formas:

$$\rho_\tau(u) = [\tau I(u \geq 0) + (1 - \tau)I(u < 0)] |u| = [\tau - I(u < 0)] u \quad (1.9)$$

donde u es la predicción del error, $I(u)$ es la función indicatriz definida mediante

$$I(u \geq 0) = \begin{cases} 1 & \text{si } u \geq 0 \\ 0 & \text{si } u < 0 \end{cases}$$

$$I(u < 0) = \begin{cases} 1 & \text{si } u < 0 \\ 0 & \text{si } u \geq 0 \end{cases}$$

y

$$\rho_\tau(u) = \begin{cases} (1 - \tau)|u| & \text{si } u < 0 \\ \tau|u| & \text{si } u \geq 0 \end{cases} \quad (1.10)$$

La función $\rho_\tau(\cdot)$ es continua y lineal por segmentos, asimétrica cuando $\tau \neq 0,5$ y no es diferenciable en $u = 0$. Su gráfica se ilustra en la figura 1.4:

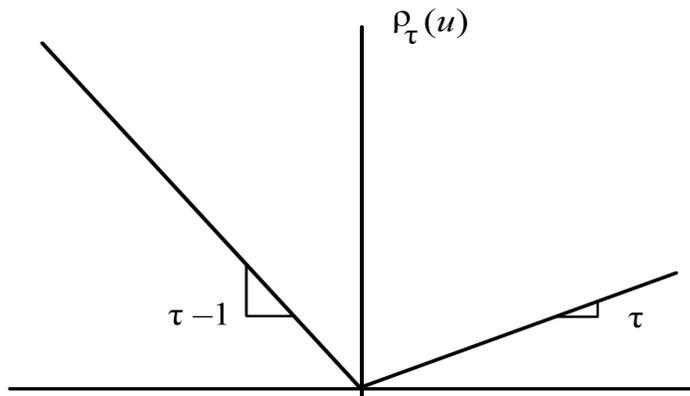


Figura 1.4: Gráfica de la función chequeo ρ para la regresión cuantílica

Gráficamente se puede observar cómo es el comportamiento de la función ρ_τ cuando $\tau = 0,25$ usando el *software* R. Considere una muestra de 21 valores

$y_i : \{-10; -9; -8; -7; -6; -5; -4; -3; -2; -1; 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10\}$

1. Definir un vector de valores muestrales comprendidos entre -10 y 10.

```
y <- seq(-10, 10)
```

2. Crear una función para calcular el vector de la función de chequeo para un f y un τ dados.

```
funct_rho<-function(u,tau){
```

Crear una variable para guardar la función indicatriz.

```
var<-0
```

Crear un vector vacío para almacenar el vector de valores de ρ_τ para cada u .

```
vec_rhos<-vector()
```

Para cada valor en el vector u , se le aplica la función y se obtiene un valor de la función indicatriz dependiendo del valor del $y[i]$.

```
for (i in 1:length(u)){
```

```
  if (u[i]>0){
```

```
    var<- 0
```

```
  }
```

```
  else{
```

```
    var<- 1
```

```
  }
```

```
  vec_rhos[i]<-u[i]*(tau-var)
```

```
}
```

```
return(vec_rhos)
```

```
}
```

3. Se aplica la función ρ_τ al vector para $\tau = 0,25$.

```
rhos<-funct_rho(y,0.25)
```

4. Graficar los puntos de y con su respectivo valor de la función ρ_τ .

```
plot(y,rhos, type = "l", lwd = 1, xlim = c(-10,10), ylim = c(0,8))
```

5. Para cada elemento de y se grafica con un círculo relleno el valor correspondiente a la función objetivo.

```
for (i in 1:length(y)){
```

```

points(y[i], rhos[i], pch = 20)
}

```

La grafica de la función chequeo cuando $\tau = 0,25$ se ilustra en la figura (1.5)

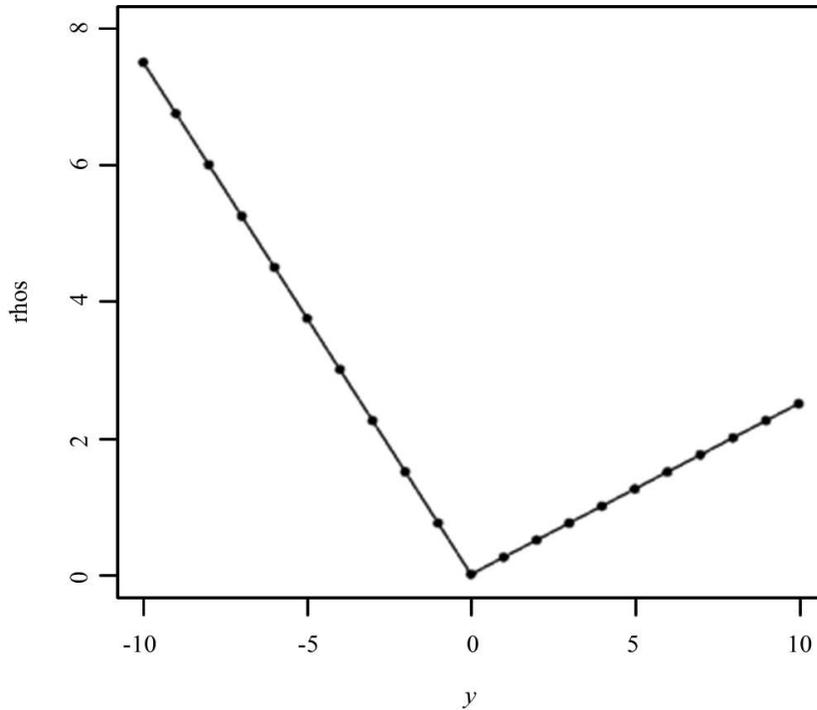


Figura 1.5: Gráfica de la función chequeo cuando $\tau = 0,25$. Elaborada con el *software* R usando los 21 valores muestrales.

1.4 Cuantil como Solución a Problemas de Minimización.

Sea Y una variable aleatoria, para cualquier $\tau \in (0, 1)$, la distancia de Y a un \hat{y} dado está medida por la distancia absoluta pero se aplican diferentes ponderaciones dependiendo de si Y se ubica a la derecha o la izquierda de \hat{y} . En este caso, la distancia ponderada viene dada por

$$d_{\tau}(Y, \hat{y}) = \rho_{\tau}|Y - \hat{y}| \tag{1.11}$$

en la cual ρ_{τ} es la función chequeo definida en (1.10).

Aplicando la definición de la función ρ_τ a la ecuación (1.11), se obtiene que la distancia de Y a un \hat{y} está dada por

$$d_\tau(Y, \hat{y}) = \begin{cases} \tau|y - \hat{y}| & \text{si } y \geq \hat{y} \\ (1 - \tau)|y - \hat{y}| & \text{si } y < \hat{y}. \end{cases}$$

Se buscará el valor \hat{y} que minimiza la esperanza de la función de chequeo:

$$E[\rho_\tau(Y - \hat{y})].$$

Esta minimización ocurre cuando \hat{y} es el τ -ésimo cuantil.

Aplicando un procedimiento similar utilizado en la minimización de la mediana, se tiene

$$\begin{aligned} E[\rho_\tau(Y - \hat{y})] &= \int_{-\infty}^{\hat{y}} (1 - \tau)|y - \hat{y}|f(y) dy + \int_{\hat{y}}^{+\infty} \tau|y - \hat{y}|f(y) dy \\ &= -(1 - \tau) \int_{-\infty}^{\hat{y}} (y - \hat{y})f(y) dy + \tau \int_{\hat{y}}^{+\infty} (y - \hat{y})f(y) dy \end{aligned}$$

Diferenciando con respecto a \hat{y} y haciendo la derivada parcial igual a cero se obtiene la solución al problema de minimización.

La derivada parcial del 1^{er} término es

$$\begin{aligned} -(1 - \tau) \frac{\partial}{\partial \hat{y}} \left[\int_{-\infty}^{\hat{y}} (y - \hat{y})f(y) dy \right] &= -(1 - \tau) \frac{\partial}{\partial \hat{y}} \left[\int_{-\infty}^{\hat{y}} yf(y) dy - \int_{-\infty}^{\hat{y}} \hat{y}f(y) dy \right] \\ &= -(1 - \tau) \frac{\partial}{\partial \hat{y}} \int_{-\infty}^{\hat{y}} yf(y) dy \\ &\quad + (1 - \tau) \frac{\partial}{\partial \hat{y}} \int_{-\infty}^{\hat{y}} \hat{y}f(y) dy \end{aligned}$$

$$\begin{aligned}
-(1-\tau) \frac{\partial}{\partial \hat{y}} \left[\int_{-\infty}^{\hat{y}} (y - \hat{y}) f(y) dy \right] &= -(1-\tau) y f(y) \Big|_{y=\hat{y}} + \int_{-\infty}^{\hat{y}} \frac{\partial}{\partial \hat{y}} y f(y) dy \\
&+ (1-\tau) \hat{y} f(y) \Big|_{y=\hat{y}} + (1-\tau) \int_{-\infty}^{\hat{y}} \frac{\partial}{\partial \hat{y}} \hat{y} f(y) dy \\
&= -(1-\tau) \hat{y} F(\hat{y}) + (1-\tau) \hat{y} F(\hat{y}) \\
&+ (1-\tau) \int_{-\infty}^{\hat{y}} f(y) dy \\
&= (1-\tau) \int_{-\infty}^{\hat{y}} f(y) dy
\end{aligned}$$

Por lo tanto,

$$-(1-\tau) \frac{\partial}{\partial \hat{y}} \left[\int_{-\infty}^{\hat{y}} (y - \hat{y}) f(y) dy \right] = (1-\tau) F(\hat{y}) \tag{1.12}$$

Ahora para el segundo término

$$\begin{aligned}
\tau \frac{\partial}{\partial \hat{y}} \left[\int_{\hat{y}}^{\infty} (y - \hat{y}) f(y) dy \right] &= \tau \frac{\partial}{\partial \hat{y}} \left[\int_{\hat{y}}^{\infty} y f(y) dy - \int_{\hat{y}}^{\infty} \hat{y} f(y) dy \right] \\
&= \tau \frac{\partial}{\partial \hat{y}} \int_{\hat{y}}^{\infty} y f(y) dy - \tau \frac{\partial}{\partial \hat{y}} \int_{\hat{y}}^{\infty} \hat{y} f(y) dy \\
&= \tau y f(y) \Big|_{y=\hat{y}} + \int_{\hat{y}}^{\infty} \frac{\partial}{\partial \hat{y}} y f(y) dy - \tau y f(y) \Big|_{y=\hat{y}} \\
&\quad - \tau \int_{\hat{y}}^{\infty} \frac{\partial}{\partial \hat{y}} \hat{y} f(y) dy
\end{aligned}$$

$$\begin{aligned} \tau \frac{\partial}{\partial \hat{y}} \left[\int_{\hat{y}}^{\infty} (y - \hat{y}) f(y) dy \right] &= \tau \hat{y} F(\hat{y}) - \tau \hat{y} F(\hat{y}) - \tau \int_{\hat{y}}^{\infty} f(y) dy \\ &= -\tau \int_{\hat{y}}^{\infty} f(y) dy \end{aligned}$$

Como

$$\int_{\hat{y}}^{\infty} f(y) dy = \int_{-\infty}^{\infty} f(y) dy - \int_{-\infty}^{\hat{y}} f(y) dy,$$

Por las propiedades de la función de distribución se tiene que

$$\int_{-\infty}^{\infty} f(y) dy = 1 \quad \text{y} \quad \int_{-\infty}^{\hat{y}} f(y) dy = F(\hat{y}),$$

luego,

$$-\tau \int_{\hat{y}}^{\infty} f(y) dy = -\tau (1 - F(\hat{y})). \quad (1.13)$$

Combinando las ecuaciones (1.12) y (1.13) se obtiene

$$\begin{aligned} \frac{\partial}{\partial \hat{y}} \int_{-\infty}^{\infty} d_{\tau}(Y, \hat{y}) f(y) dy &= (1 - \tau) F(\hat{y}) - \tau (1 - F(\hat{y})) \\ &= F(\hat{y}) - \tau F(\hat{y}) - \tau + \tau F(\hat{y}) \\ &= F(\hat{y}) - \tau \end{aligned}$$

Luego, igualando la derivada parcial a cero se tiene

$$F(\hat{y}) - \tau = 0$$

$$F(\hat{y}) = \tau.$$

Esto satisface el problema de minimización.

Definición 1.3. (Cuantil Muestral)

Dada una variable aleatoria $\{Y_1, \dots, Y_n\}$ se define el cuantil muestral mediante

$$F_n(Y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$$

donde F_n es la función de distribución acumulada e I es la función indicatriz definida mediante

$$I(Y_i \leq y) = \begin{cases} 1 & \text{si } Y_i < y \\ 0 & \text{si } Y_i > y. \end{cases} \quad (1.14)$$

Se observa que por (1.14), la función F_n es continua y constante por segmentos.

Si F_n es una función escalonada, luego

$$F_n^{-1}(y) = \inf\{y : F_n(y) \geq \tau\} \quad \text{para } \tau \in (0, 1)$$

toma un número finito de valores diferentes.

Similarmente, el τ -ésimo cuantil muestral de Y puede ser visto como una solución al problema de minimización.

El valor y minimiza el promedio de la función de chequeo

$$\frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - y) \quad (1.15)$$

Aplicándole la definición de la función de chequeo a (1.15) se obtiene

$$\frac{1-\tau}{n} \sum_{y_i < y} |y_i - y| + \frac{\tau}{n} \sum_{y_i > y} |y_i - y| \quad (1.16)$$

Considere ahora una muestra de tamaño $n = 7$ para ilustrar gráficamente que los cuantiles muestrales pueden definirse como un problema de minimización

$$y_i : \{1, 62234; -0, 16122; -0, 46060; -0, 61184; 0, 67386; 0, 71067; -0, 86516\}$$

1. Definir un vector de valores muestrales con la función c .

$$y < -c(1.62234, -0.16122, -0.46060, -0.61184, 0.67386, 0.71067, -0.86516)$$

2. Crear una función para calcular el vector de la función de chequeo para un f y un τ dados.

```
funct_loss<-function(f,tau){
```

Crear un vector vacío para almacenar cada valor de la función objetivo.

```
val<-vector()
```

Para cada valor en el vector y , se le aplica la función correspondiente dependiendo del valor de los $y[i]$ con respecto al valor de f .

```
for (i in 1:length(y)){
  if (y[i]<f){
    val[i]<- (1-tau)*abs(y[i]-f)
  }
  else{
    val[i]<- tau*abs(y[i]-f)
  }
}
return(val)
}
```

3. Crear un conjunto de valores entre -2 y 2 separados uniformemente por 0.001.

```
f<-seq(-2,2,by = .001)
```

4. Crear un vector vacío para almacenar el valor de la función objetivo correspondiente al conjunto de valores definido anteriormente.

```
suma<-vector()
```

5. Para cada valor en el conjunto f se le aplica la función chequeo para crear luego un bosquejo de la gráfica que pasa por los puntos de y .

```
for (i in 1:length(f)){
```

```
w<-sapply (f[i], funct_loss, tau=0.25)
```

Guardar el valor de la suma para el i -ésimo valor de f .

```
suma[i]<-sum(w)
```

```
}
```

6. Crear un vector vacío para almacenar los valores de la función objetivo para los puntos del vector y .

```
k<-vector()
```

7. Para cada valor en el conjunto y se le aplica la función `funct_loss` con $\tau = 0,25$.

```
for (i in 1:length(y)){
```

```
w<-sapply(y[i], funct_loss, tau=0.25)
```

Guardar el valor de la suma para el i -ésimo valor de y .

```
k[i]<-sum(w)
```

```
}
```

8. Calcular el mínimo valor del vector.

```
min_valor<-min(k)
```

9. Obtener el índice del vector `suma` cuyo valor es el mínimo calculado arriba.

```
indice_minimo<-match(c(min_valor),k)
```

10. Obtener el valor de y cuyo índice es el calculado arriba.

```
minimo<-y[indice_minimo]
```

11. Graficar los puntos de f con su respectivo valor de la función objetivo (`suma`).

```
plot(f,suma, type = "l", lwd = 1, xlim = c(-2.5,2.5), ylim = c(0,12))
```

12. Para cada elemento de y se grafica con un círculo relleno el valor correspondiente a la función objetivo.

```
for (i in seq(1,7)){  
  points(y[i], k[i], pch = 20)  
}
```

13. Dibujar un segmento de recta para representar la ubicación del valor mínimo ver gráfica (1.6).

```
segments(minimo,0,minimo,12)
```

14. Mostrar en pantalla el valor de la función objetivo para cada uno de los puntos de y .

```
k  
7.836247 1.963087 1.588863 1.551052 3.842018 3.961650 1.741043
```

15. Mostrar en pantalla el valor mínimo de las sumas para cada uno de los puntos de y .

```
min_valor  
1.551052
```

16. Mostrar en pantalla el valor del elemento de y con el que se obtuvo el valor mínimo.

```
minimo  
-0.61184
```

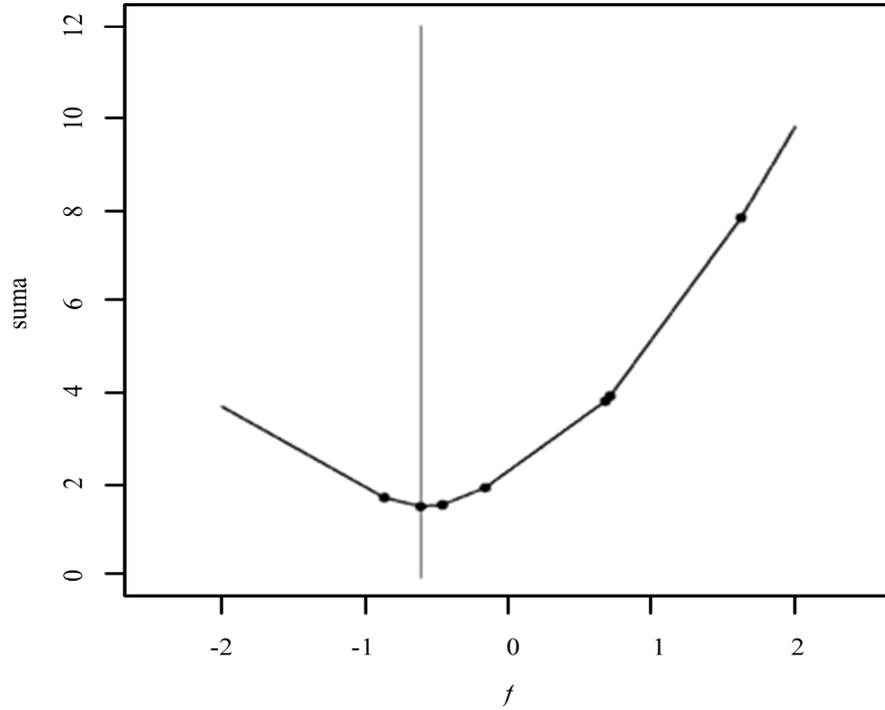


Figura 1.6: Gráfica del cuantil muestral $\tau = 0,25$. como una solución al problema de minimización. Elaborada con el *software* R usando los 7 valores muestrales.

1.5 Propiedades de los Cuantiles

- i. Para cualquier $\tau \in (0, 1)$, los estimadores del τ -ésimo cuantil se encuentran resolviendo el problema de minimización

$$\text{mín } E [\rho_\tau(y - \hat{y})]$$

- ii. Transformación Monótona: Para una función monótona $h \in \mathbb{R}$ y cualquier variable aleatoria Y

$$Q_{h(Y)}(\tau) = h(Q_Y(\tau)) \tag{1.17}$$

Por Ejemplo, si a la función h se le aplica una transformación logarítmica, luego la ecuación (1.17) puede expresarse

$$Q_{\log(Y)}(\tau) = \log(Q_Y(\tau))$$

y equivalentemente,

$$Q_Y(\tau) = \exp^{Q_{\log(Y)}(\tau)}$$

iii. Equivarianza Lineal: Suponga que Y es una variable aleatoria y a, c son constantes luego,

$$Q_{c+aY} = c + aQ_Y(\tau)$$

iv. Robustez a los datos atípicos en la variable dependiente Y . (Ver definición (1.8))

Definición 1.4. (Cuantil Condicional)

Suponga que (X, Y) es un vector aleatorio, donde Y es la variable dependiente y X es la variable independiente. El τ -ésimo cuantil condicional se define

$$F_{Y/X} = P(Y \leq \tau/x) = \tau$$

donde $F_{Y/X}$ es la función condicional de Y dado $X = x$.

Otra forma de expresar el τ -ésimo cuantil condicional es

$$F_{Y/X}^{-1}(\tau/x) = \inf\{y : F_{Y/X}(\tau/x) \geq \tau\}$$

La demostración de que el τ -ésimo cuantil condicional minimiza la esperanza de la función de chequeo

$$E[\rho_\tau(Y - \hat{y})/X],$$

se realiza de forma similar a la del cuantil.

Definición 1.5. (Datos Atípicos)

Un dato atípico es una observación que parece ser inconsistente con el resto del conjunto de datos.

Si un valor en la variable dependiente se encuentra distante de la mayoría de los valores de \mathbf{Y} , ese valor se conocen como dato atípico en la dirección de \mathbf{Y} , en la variable independiente como dato atípico en la dirección de \mathbf{X} .

En el contexto de la regresión lineal, los datos atípicos se conocen como puntos que tienen un residual grande en valor absoluto.

Definición 1.6. (Punto de Influencia)

Si un valor $x_i \in \mathbf{X}$, es un dato atípico, el par (x_i, y_i) se conoce como punto de influencia porque tiene un efecto importante sobre los estimadores de la regresión lineal por el método de mínimos cuadrados.

Cuando en la variable independiente existen valores que son grandes o pequeños en comparación con el resto de las observaciones y un punto de influencia se encuentra cercano a la recta de regresión, alrededor de la cual la mayoría de los pares de puntos están centrados, estos valores influyen mejorando la precisión de los estimadores.

En caso contrario, cuando un punto de influencia está situado lejos de la recta de regresión, alrededor de la cual la mayoría de los pares de puntos están centrados, los datos atípicos de la variable independiente influyen reduciendo la precisión de los coeficientes de regresión por el método de mínimos cuadrados.

En algunos casos los datos atípicos representa un error en la medida del fenómeno a ser estudiado o un error en la transcripción de los datos. Si éste es el caso, un dato atípico puede ser corregido o removido del conjunto de datos.

1.6 Regresión Cuantílica

Definición 1.7. (Regresión Cuantílica)

Suponga que $\{y_i : i = 1, \dots, n\}$ es una muestra aleatoria sobre un vector \mathbf{Y} de variables aleatorias dependientes, con una función de distribución condicional

$$F(y - \mathbf{x}'_i\boldsymbol{\beta}) = Pr(\mathbf{Y}_i \leq y/\mathbf{x}_i),$$

en la cual \mathbf{x}_i es un vector $n \times 1$ de variables independientes, $\boldsymbol{\beta}$ es un vector $n \times 1$ de parámetros desconocidos. Luego la función condicional cuantílica de \mathbf{Y} dado

\mathbf{x} puede expresarse como:

$$Q_Y(\tau/\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} \quad (1.18)$$

Koenker y Bassett (1982) determinaron que la función condicional cuantílica (1.18) puede definirse mediante la solución al problema de minimización

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}'_i \boldsymbol{\beta}), \quad (1.19)$$

en la cual ρ_τ está definida por

$$\rho_\tau(y_i - \mathbf{x}'_i \boldsymbol{\beta}) = \begin{cases} \tau |y_i - \mathbf{x}'_i \boldsymbol{\beta}| & \text{si } y_i \geq \mathbf{x}'_i \boldsymbol{\beta} \\ (1 - \tau) |y_i - \mathbf{x}'_i \boldsymbol{\beta}| & \text{si } y_i < \mathbf{x}'_i \boldsymbol{\beta} \end{cases} \quad (1.20)$$

y los estimadores de $\boldsymbol{\beta}$ de $\hat{\boldsymbol{\beta}}$ los definieron como regresión cuantílica.

Definición 1.8. (Robustez)

La robustez se refiere a la insensibilidad a los datos atípicos y al incumplimiento de las nociones distribucionales del modelo en cuanto a los datos de la variable dependiente. Bajo las nociones del modelo de regresión lineal, los estimadores pueden ser sensibles a los datos atípicos en comparación con el modelo de regresión cuantílica, en la cual los estimadores no son sensibles a los datos atípicos. Esta robustez se debe a la naturaleza de la función objetivo

$$\sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}'_i \boldsymbol{\beta}).$$

Además, para determinar la matriz de varianza-covarianza, en la regresión lineal se requiere la noción de normalidad para obtener los estadísticos inferenciales. La violación de la noción de normalidad puede causar inexactitud en los errores estándares. Por otro lado, el modelo de regresión cuantílica es robusto a las nociones distribucionales porque el estimador pondera el comportamiento local de la distribución cercana al cuantil específico más que el comportamiento asintótico de la distribución. Los estadísticos inferenciales del modelo de regresión

cuantílica pueden tener una distribución libre. Esta robustez es importante en el estudio de fenómenos de distribución muy insesgados.

1.7 Usos de la Regresión Cuantílica.

En esta sección se mencionan los casos en que resulta útil aplicar la regresión cuantílica en vez de la regresión por mínimos cuadrados.

De acuerdo con John y Nduka (2009), cuando en un conjunto de datos existe heteroscedasticidad el método de mínimos cuadrados resulta no ser muy seguro. No obstante, la regresión cuantílica es muy útil para visualizar los cambios en la distribución condicional del conjunto de datos.

Lee (2005), citado por John y Nduka (2009), establecieron que el beneficio de aplicar la regresión cuantílica es que, al usar los cuantiles, éstos tienen la propiedad de ser robustos en la estimación de los datos atípicos y por lo tanto la regresión cuantílica hereda su propiedad de robustez.

Kecojević (2011) afirmó que la regresión cuantílica se utiliza para medir el efecto de las covariables no sólo en el centro de la distribución sino en los extremos superiores e inferiores de ésta.

Surendra, et al. (2011), citando a Buchisky (1998) mencionaron que otra ventaja del modelo de regresión cuantílica, cuando la distribución de los errores no es normal, es que los estimadores de la regresión cuantílica son más eficientes que los estimadores de mínimos cuadrados.

Además, la regresión cuantílica permite modelar datos que presentan mucha variabilidad, describiendo su comportamiento para cada cuantil.

CAPITULO 2

ESTIMADORES DE LA REGRESIÓN CUANTÍLICA

La regresión cuantílica, introducida por Koenker y Bassett (1978), puede verse como una extensión del modelo de regresión lineal. Pero, mientras en este modelo el objetivo es minimizar la suma de los residuales al cuadrado y utilizar la media como estimador; en la regresión cuantílica se busca minimizar una suma de errores absolutos ponderados asimétricamente y utilizar los cuantiles como estimadores.

Una de las técnicas más usadas para solucionar un problema de regresión cuantílica es por medio de su representación como un problema de programación lineal.

Un problema de programación lineal es un problema matemático en el cual la función objetivo es lineal en las incógnitas y las restricciones constan de igualdades y desigualdades lineales.

En esencia, la programación lineal consiste en optimizar (maximizar o minimizar) una función objetivo, que es una función de varias variables, sujeta a una serie de restricciones expresadas por inecuaciones lineales. Una de las grandes ventajas de utilizar la programación lineal es que mediante un algoritmo de resolución se puede obtener la decisión más óptima aunque hayan miles de variables y relaciones entre ellas.

Desafortunadamente, para mediados del siglo XVIII, se conocía poco acerca de la teoría de programación lineal, por esta razón Boskovich (1760), en el problema de la estimación de la elasticidad de La Tierra aplicó un algoritmo geométrico para minimizar la suma de los valores absolutos de los residuales, sujeto a la restricción de que el promedio de los errores sea igual a cero. Sin embargo, en la aproximación geométrica utilizada por Edgeworth (1888), la minimización de la

suma de los valores absolutos de los residuales se aplicó sobre la intercepción y la pendiente de los parámetros del modelo. El uso de esta aproximación le permitió al autor anterior acercarse a lo que hoy se conoce como método Simplex.

Uno de los algoritmos utilizados para resolver problemas de programación lineal es el método Simplex, creado en 1947 por el matemático George Dantzig, que consiste en partir de una solución básica factible o punto extremo e ir a una solución básica factible adyacente con mejor valor de la función objetivo. Este proceso continúa hasta que ya no se pueden obtener mejoras, por lo que se habrá encontrado una solución óptima.

En este Capítulo se introducen las propiedades de los estimadores de la regresión cuantílica y las definiciones relacionadas con el algoritmo utilizado para determinar dichos estimadores.

2.1 Estimación de los Parámetros de la Regresión Cuantílica.

El punto de partida de este modelo son las funciones cuantílicas descritas en el Capítulo I, las cuales se utilizan para describir y comparar distribuciones de una variable aleatoria Y con función de distribución acumulada F_Y .

Ahora bien, cuando se necesita modelar la relación entre una variable respuesta y un conjunto de variables predictoras, se introducen modelos tipo regresión para explicar las variaciones de la variable respuesta en términos de una medida de tendencia central como la media θ a través de diversos cuantiles.

En el caso de la media, se tiene la regresión lineal clásica donde los estimadores se determinan mediante el método de mínimos cuadrados. Cuando la medida empleada son los cuantiles, se tiene la regresión cuantílica, y el método se cambia por el de las mínimas distancias absolutas.

A continuación, se esbozará el proceso a seguir en ambos casos: En la regresión lineal clásica, la variable respuesta \mathbf{Y} puede escribirse en forma matricial

como

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

donde \mathbf{Y} es el vector $n \times 1$ de las observaciones, $\boldsymbol{\beta}$ es el vector $(m + 1) \times 1$ de parámetros desconocidos, \mathbf{X} es la matriz $n \times (m + 1)$ de la regresión y $\boldsymbol{\varepsilon}$ es el vector $n \times 1$ de los errores.

Se supone que los errores $\boldsymbol{\varepsilon}_i$ son desviaciones que se comportan como variables aleatorias, tienen una distribución normal, son independientes y verifican las siguientes condiciones:

1. $E(\boldsymbol{\varepsilon}_i) = 0$, para $i = 1, \dots, n$.
2. $Var(\boldsymbol{\varepsilon}_i) = \sigma^2$, para $i = 1, \dots, n$, las varianzas son iguales pero desconocidas.
3. $E(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = 0$, para todo $i \neq j$, las n desviaciones no están relacionadas.

Para estimar los parámetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$ se utiliza el método de mínimos cuadrados, en el cual se determina el conjunto de valores $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_n)'$ que minimiza el cuadrado de las desviaciones

$$\begin{aligned} \|\boldsymbol{\varepsilon}\|^2 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned} \quad (2.1)$$

Diferenciando (2.1) con respecto a cada una de las componentes del vector $\boldsymbol{\beta}$, e igualando a cero se obtiene un conjunto de ecuaciones equivalente a la forma matricial

$$\nabla_B \|\boldsymbol{\varepsilon}\|^2 = -\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{X}$$

Usando la propiedad de la transposición de matrices

$$[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{X}]' = \mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Se tiene

$$\begin{aligned}\nabla_B \|\boldsymbol{\varepsilon}\|^2 &= -2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= 0\end{aligned}$$

$$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

y se resuelve para $\hat{\boldsymbol{\beta}}$.

Estas ecuaciones normales tienen una única solución $\hat{\boldsymbol{\beta}}$ cuando el rango de la matriz \mathbf{X} es máximo, es decir, cuando las columnas de esta matriz son linealmente independiente. Luego, $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, de manera que

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

El modelo de regresión ajustado es una media condicional, es decir,

$$E[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\hat{\boldsymbol{\beta}}.$$

En la regresión cuantílica, se procederá de forma similar a la regresión lineal; pero antes se da la definición de la derivada direccional en la dirección del vector $\boldsymbol{\omega}$.

Definición 2.1. (Derivada direccional en el Punto P)

Sea f una función escalar, la derivada direccional de f en el punto P , en la dirección del vector unitario $\boldsymbol{\omega}$ viene dada por

$$\frac{df}{ds} = \lim_{s \rightarrow 0} \frac{f(\mathbf{r} + s\boldsymbol{\omega}) - f(\mathbf{r})}{s}$$

la cual es equivalente a

$$\left. \frac{df}{ds}(\mathbf{r} + s\boldsymbol{\omega}) \right|_{s=0} = \boldsymbol{\omega} \nabla f(\mathbf{r}, \boldsymbol{\omega})$$

Para determinar los estimadores cuantílicos se requiere de la minimización de la función objetivo, proceso que se lleva a cabo usando el concepto de la derivada direccional.

La función objetivo

$$R(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_{\tau}(y_i - x'_i \boldsymbol{\beta}).$$

donde la función ρ_{τ} se define igual a la ecuación (1.20), es lineal por segmentos y continua. Además, es diferenciable excepto en los puntos donde uno o más residuales, $y_i - x'_i \boldsymbol{\beta}$, se anulan. En tales puntos, $R(\boldsymbol{\beta})$ tiene derivadas direccionales en todas las direcciones, dependiendo de la dirección elegida para evaluarla.

La derivada direccional de R en la dirección $\boldsymbol{\omega}$, viene dada por

$$\begin{aligned} \boldsymbol{\omega} \nabla R(\boldsymbol{\beta}, \boldsymbol{\omega}) &\equiv \left. \frac{d}{dt} R(\boldsymbol{\beta} + t \boldsymbol{\omega}) \right|_{t=0} \\ &= \left. \frac{d}{dt} \sum_{i=1}^n (y_i - x'_i \boldsymbol{\beta} - x'_i t \boldsymbol{\omega}) [\tau - I(y_i - x'_i \boldsymbol{\beta} - x'_i t \boldsymbol{\omega}) < 0] \right|_{t=0} \\ &= - \sum_{i=1}^n x'_i \boldsymbol{\omega} [\tau - I(y_i - x'_i \boldsymbol{\beta}) < 0] \end{aligned} \quad (2.2)$$

La suma (2.2) se parte en tres sumando: uno para los residuales nulos, otra para los residuales positivos y otro para los residuales negativos; pero todos en la dirección de $-x'_i \boldsymbol{\omega}$. Es decir,

$$\begin{aligned} \boldsymbol{\omega} \nabla R(\boldsymbol{\beta}, \boldsymbol{\omega}) &\equiv - \sum_{i \in \{i: u_i=0, -x'_i \boldsymbol{\omega}\}} \tau x'_i \boldsymbol{\omega} + \sum_{i \in \{i: u_i>0, -x'_i \boldsymbol{\omega}\}} \tau x'_i \boldsymbol{\omega} \\ &\quad - \sum_{i \in \{i: u_i<0, -x'_i \boldsymbol{\omega}\}} (1 - \tau) x'_i \boldsymbol{\omega} \\ &= - \sum \psi_{\tau}^*(y_i - x'_i \boldsymbol{\beta}, -x'_i \boldsymbol{\omega}) x'_i \boldsymbol{\omega} \end{aligned} \quad (2.3)$$

donde los $u_i = y_i - x'_i \boldsymbol{\beta}$ son los residuales y

$$\psi_{\tau}^*(u, w) = \begin{cases} \tau - I(u < 0) & \text{si } u \neq 0, \\ \tau - I(w < 0) & \text{si } u = 0. \end{cases} \quad (2.4)$$

Si en un punto $\hat{\boldsymbol{\beta}}$, las derivadas direccionales son todas no negativas, es decir, $\nabla R(\hat{\boldsymbol{\beta}}, \boldsymbol{\omega}) \geq 0$, para todo $\boldsymbol{\omega} \in \mathbb{R}^P$ con $\|\boldsymbol{\omega}\| = 1$, entonces $\hat{\boldsymbol{\beta}}$ minimiza $R(\boldsymbol{\beta})$. Esta es una generalización de $\nabla R(\boldsymbol{\beta}) = 0$ cuando se suaviza la función $R(\boldsymbol{\beta})$. Simplemente se requiere que la función crezca a medida que se aleja de $\hat{\boldsymbol{\beta}}$ independientemente de la dirección que se tome. Las soluciones que minimizan a $R(\boldsymbol{\beta})$ ocurren en aquellos vértices unidos por líneas donde está función no es decreciente.

2.2 Propiedades de los Estimadores de la Regresión Cuantílica

Suponga que $\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X})$ representa el τ -ésimo estimador de la regresión cuantílica, \mathbf{y} es un vector $1 \times n$ de variables aleatorias dependientes, $\boldsymbol{\gamma} \in \mathbb{R}^n$ y \mathbf{X}, \mathbf{A} son matrices de orden $n \times n$ con \mathbf{A} no singular. Luego, para cualquier $\tau \in [0, 1]$, se cumplen las siguientes propiedades:

- (i) Escala de la Equivarianza: Para $c > 0$,

$$c \hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) = \hat{\boldsymbol{\beta}}(\tau; c\mathbf{y}, \mathbf{X}) \quad (2.5)$$

Cuando el τ -ésimo estimador de la regresión cuantílica se multiplica por una constante $c > 0$, la variable dependiente cambia su escala a $c\mathbf{y}$ produciendo un nuevo estimador que cambia proporcionalmente.

- (ii) Para $c < 0$

$$\begin{aligned} -c \hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) &= \hat{\boldsymbol{\beta}}(\tau; -c\mathbf{y}, \mathbf{X}) \\ &= -c \hat{\boldsymbol{\beta}}(1 - \tau; \mathbf{y}, \mathbf{X}) \end{aligned} \quad (2.6)$$

Cuando el τ -ésimo estimador de la regresión cuantílica se multiplica por una constante $c < 0$, la variable dependiente cambia de signo y se obtiene un estimador que transforma a τ en $1 - \tau$.

(iii) Ajuste de la Equivarianza:

$$\hat{\beta}(\tau; \mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma} = \hat{\beta}(\tau; \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) \quad (2.7)$$

Cuando el τ -ésimo estimador de la regresión cuantílica se le suma un vector $\boldsymbol{\gamma}$ de orden $1 \times n$, se obtiene un nuevo estimador que aumenta en $\mathbf{X}\boldsymbol{\gamma}$ el valor de la variable dependiente.

(iv) Equivarianza para la Reparametrización del Diseño:

$$\mathbf{A}^{-1}\hat{\beta}(\tau; \mathbf{y}, \mathbf{X}) = \hat{\beta}(\tau; \mathbf{y}, \mathbf{X}\mathbf{A}) \quad (2.8)$$

Cuando el τ -ésimo estimador de la regresión cuantílica se multiplica por la inversa de la matriz \mathbf{A} , se obtiene un estimador que cambia la escala de \mathbf{X} a $\mathbf{X}\mathbf{A}$.

(v) Equivarianza para la Transformación Monótona: Para una función monótona h y cualquier variable aleatoria Y .

$$Q_{h(\mathbf{Y})}(\tau/\mathbf{x}) = h(Q_{\mathbf{Y}}(\tau/\mathbf{x})) \quad (2.9)$$

Así mismo, se puede aplicar una transformación logarítmica sustituyendo $h = \log$ en (2.9) para obtener:

$$Q_{\log(\mathbf{Y})}(\tau/\mathbf{x}) = \log(Q_{\mathbf{Y}}(\tau/\mathbf{x}))$$

(vi) Para cualquier $\tau \in (0, 1)$, los estimadores de la regresión cuantílica se encuentran resolviendo el problema de minimización

$$\text{mín } E [\rho_{\tau}(y - \mathbf{x}'\boldsymbol{\beta})]$$

Las propiedades (2.5), (2.6), (2.7) y (2.8) pueden resumirse en el Teorema 2.1:

Teorema 2.1. *Suponga que \mathbf{A} es cualquier matriz $n \times n$, no singular, $\boldsymbol{\gamma} \in \mathbb{R}^n$, y $c > 0$. Luego, para cualquier $\tau \in [0, 1]$,*

$$(i) \quad c\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) = \hat{\boldsymbol{\beta}}(\tau; c\mathbf{y}, \mathbf{X})$$

$$(ii) \quad -c\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) = -c\hat{\boldsymbol{\beta}}(1 - \tau; \mathbf{y}, \mathbf{X})$$

$$(iii) \quad \hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma} = \hat{\boldsymbol{\beta}}(\tau; \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X})$$

$$(iv) \quad \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) = \hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{XA})$$

Demostración. Ver Koenker y Bassett (1978) □

Definición 2.2. (Problemas de Programación Lineal)

Un problema de programación lineal consiste en maximizar o minimizar una función lineal en la presencia de restricciones lineales del tipo desigualdad, igualdad o ambas. Por ejemplo, el siguiente sistema representa un problema de programación lineal.

$$\begin{array}{ll} \text{Minimizar} & c_1x_1 + c_2x_2 + \cdots + c_nx_n. \\ & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq b_1. \\ \text{Sujeto a} & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \geq b_2. \\ & \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \geq b_m. \end{array}$$

En la cual $c_1x_1 + c_2x_2 + \cdots + c_nx_n$ es la función objetivo, que debe ser minimizada. Los coeficientes c_i, b_i y a_{ij} , ($i = 1, \dots, m$, $j = 1, \dots, n$) son constantes reales fijas, y las x_i son variables que deben determinarse. El vector (c_1, c_2, \dots, c_n) representa el vector de costo, y la desigualdad

$$\sum_{j=1}^n a_{ij}x_j \geq b_i \tag{2.10}$$

representa la i -ésima restricción. Los a_{ij} forman la matriz de restricciones

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

El vector columna cuya i -ésima componente es b_i , en la desigualdad (2.10) representa los requerimientos mínimos que deben satisfacerse.

El problema de programación lineal anterior puede escribirse en su forma canónica

$$\text{mín}\{c'x \mid Ax \geq b, x \geq 0\}. \quad (2.11)$$

o en su forma estándar

$$\text{mín}\{c'x \mid Ax = b, x \geq 0\}. \quad (2.12)$$

en la cual $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ y $A \in \mathbb{R}^{m \times n}$.

Definición 2.3. (Punto o Vector Factible)

Es un punto o vector para un problema de optimización que satisface todas las restricciones del problema.

Definición 2.4. (Región o Espacio Factible)

Es el conjunto de todos los puntos factibles que satisfacen las restricciones. Su representación viene dada por

$$S = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}. \quad (2.13)$$

Definición 2.5. (Conjuntos Poliédricos)

Es la intersección de un número finito de subespacios y se puede representar por $\{x : Ax \leq b\}$, en la cual A es una matriz de orden $m \times n$ cuyo i -ésimo renglón es a^i y b es un m vector.

Si el conjunto poliédrico es acotado, su representación viene dada por

$$\{x : Ax = b, x \geq 0\}. \quad (2.14)$$

Definición 2.6. (Conjuntos Convexos)

Un conjunto, $D \subset \mathbb{R}^2$ es convexo si, siempre que (t_1, y_1) y (t_2, y_2) pertenecen a D , $((1 - \lambda)t_1 + \lambda t_2, (1 - \lambda)y_1 + \lambda y_2)$ también pertenecen a D para cada número real λ , $0 \leq \lambda \leq 1$.

Definición 2.7. (Cono Convexo)

Un conjunto $K \subseteq \mathbb{R}^n$ es cono convexo si éste es cerrado bajo las siguientes operaciones:

1. La multiplicación por escalar no negativo, es decir, $\alpha \mathbf{x} \in K$, $\alpha \geq 0$, $\mathbf{x} \in K$
2. Las combinaciones convexas, $\alpha K + (1 - \alpha)K \subseteq K$, $\alpha \in (0, 1)$.

Definición 2.8. (Puntos Extremos)

Un punto $x \in X$, de un conjunto convexo X se llama punto extremo de X , si x no se puede representar como una combinación convexa de dos puntos distintos en X , es decir, $x = \lambda x_1 + (1 - \lambda)x_2$ con $\lambda \in [0, 1]$ $x_1, x_2 \in X$ entonces $x = x_1 = x_2$.

Definición 2.9. (Vértices)

Sea $S = \{x : Ax \geq b, x \geq 0\}$ un conjunto restringido, poliédrico y convexo, los vértices de S se definen como puntos extremos, esquinas, o puntos aislados, los cuales no permanecen sobre una recta conectando los puntos diferentes que están en S .

2.3 Soluciones Básicas

Definición 2.10. (Solución Básica)

Dada una matriz base B formada por m columnas de la matriz A se dice que X_B es una solución básica si verifica $BX_B = b$.

En la terminología de la programación lineal, a los subconjuntos de los p vértices donde $R(\beta)$ se minimiza, se conocen como soluciones básicas.

Suponga que h pertenece a un conjunto indexado \mathcal{H} , los subconjuntos de los p elementos para los primeros n enteros, $\mathcal{N} = \{1, \dots, n\}$, y que $X(h)$ representa una submatriz de X con filas $\{x_i : i \in h\}$. También suponga que $y(h)$ es un vector con coordenadas $\{y_i : i \in h\}$. El complemento de h con respecto a \mathcal{N} es \bar{h} ; asimismo, se definen los complementos de $X(h)$ y $y(h)$. Luego, cualquier solución básica que pase a través de estos puntos $\{(x_i, y_i), i \in h\}$ se define como:

$$b(h) = X(h)^{-1}y(h)$$

en la cual $X(h)$ es una matriz no singular. Además, existen muchas de estas soluciones básicas $\binom{n}{p}$.

2.4 Direcciones de la Pendiente

Suponga que x_0 es un punto inicial, no necesariamente factible, y considere la factibilidad de un paso de longitud σ en la dirección p con respecto a la i -ésima restricción, $c'_i x \geq b_i$. Por consiguiente

$$c'_i(x_0 + \sigma p) = c'_i x_0 + \sigma c'_i p,$$

luego se define la longitud del paso para la i -ésima restricción de x_0 en la dirección p como

$$\sigma_i = \frac{c'_i x_0 - b_i}{-c'_i p},$$

con $c'_i p \neq 0$. Para definir la factibilidad de la dirección de la pendiente considere

$$c'(x_0 + \sigma p) = c' x_0 + \sigma c' p,$$

una dirección p se dice que es una dirección de la pendiente en x_0 si y sólo si $c'p < 0$. Para determinar si existe una dirección factible de la pendiente, considere un punto x_0 y $S(A')$ un espacio generado por los vectores de la columna A' definido por

$$S(A') = \{y \mid y = A'v, \text{ para algún } v\},$$

luego si $y \in S(A')$ y $Ap = 0$, existe un v tal que

$$\begin{aligned} y &= A'v \\ (py)' &= (pA'v)' \\ y'p' &= (A'v)'p' \\ y'p &= v'A''p \\ y'p &= v'Ap \\ y'p &= v'0 \\ y'p &= 0 \end{aligned}$$

Luego, p no es una dirección de la pendiente.

Por otro lado, si $y \notin S(A')$, luego existe $p \in \mathbb{R}^n$ tal que $Ap = 0$ y $y'p < 0$, por lo tanto p es una dirección de la pendiente en x_0 .

Definición 2.11. (Restricciones Activas e Inactivas)

Una restricción de desigualdad se dice que es activa en un punto factible x si $r_i(x) = 0$. Una restricción de desigualdad es inactiva en x si $r_i(x) < 0$.

2.5 Condiciones para la Optimización

Considere el problema de programación lineal

$$\min\{c'x \mid Ax = b\}. \tag{2.15}$$

Luego, existen tres posibles circunstancias:

1. Las restricciones $Ax = b$ son inconsistentes, así, el conjunto factible es vacío y no existe punto óptimo.
2. Las restricciones $Ax = b$ son consistentes, así, el conjunto factible es no vacío y, además,
 - (a) $c \in S(A')$, luego existe v tal que $c = A'v$, y para cualquier punto factible x ,

$$\begin{aligned}
 c &= A'v \\
 xc &= xA'v \\
 (xc)' &= (xA'v)' \\
 c'x' &= v'(xA')' \\
 c'x &= v'A''x' \\
 c'x &= v'Ax \\
 c'x &= v'b
 \end{aligned}$$

así cualquier punto factible alcanza el mismo valor de la función objetivo, luego todos los puntos factibles son óptimos, o

- (b) $c \notin S(A')$, luego existe una dirección p tal que $Ap = 0$ y $c'p < 0$. Esta dirección es factible y así, partiendo de cualquier punto factible x , la función objetivo tiende a $-\infty$ haciendo un paso $x + \sigma p$ para un σ arbitrariamente grande.

3. Suponga que h representa el conjunto indexado de las restricciones activas y \bar{h} representa el conjunto indexado de las restricciones inactivas en un punto factible inicial x_0 . Una dirección factible de la pendiente p existe si

$$A(h)p \geq 0 \quad \text{y} \quad c'p < 0$$

El punto factible x_0 es óptimo si y sólo si no existe dirección de la pendiente.

Lema 2.1. *Suponga que A es una matriz de orden $m \times n$ y \mathbf{c} es un vector en \mathbb{R}^n ; luego*

$$\mathbf{c}'p \geq 0 \quad \text{para toda } p \text{ tal que } Ap \geq 0$$

si y sólo si

$$\mathbf{c} = A'v \quad \text{para algún } v \geq 0.$$

Además, \mathbf{c} puede ser expresada como una combinación lineal de las columnas de A con ponderaciones no negativas, o existe una dirección p tal que $\mathbf{c}'p \leq 0$ y $Ap \geq 0$. Aplicando el lema 2.1 se obtienen las condiciones de optimización para el problema de programación lineal y se genera el siguiente resultado.

Teorema 2.2. *(Gill, et al. 1991) Para el problema de la desigualdad restringida*

$$\text{mín}\{c'x \mid Ax \geq b\}$$

Se puede cumplir que:

(1) *Las restricciones $Ax \geq b$ son inconsistentes y por lo tanto el problema no tiene solución o*

(2) *Existe un punto factible \hat{x} y un vector \hat{v} tal que*

$$c = A(h)'\hat{v} \quad \text{para algún } \hat{v} \geq 0, \quad (2.16)$$

en el cual $A(h)$ representa la submatriz de A correspondiente a las restricciones activas en \hat{x} . En este caso, $c(\hat{x})$ es el único valor mínimo de $c'x$ para

$$\{x \mid Ax \geq b\}$$

y \hat{x} es un optimizador, o,

(3) *Las restricciones $Ax \geq b$ son consistentes, pero la condición (2.16) no se mantiene para cualquier punto factible x . En este caso, la solución no está acotada por abajo.*

Demostración. Ver Gill, et al. (1991)

□

2.6 Holgura Complementaria

En el teorema (2.2), las condiciones de optimización para (2.16) requieren que una suma ponderada de las m columnas de la matriz A' sean iguales al vector de costos \mathbf{c} . En el teorema (2.2) no se especifica la dimensión del vector $\hat{\mathbf{v}}$, pero se observa que si \hat{x} es un vértice, luego la matriz $A(h)'$ está compuesta de por lo menos n columnas linealmente independientes del conjunto completo de las m columnas de A' . Luego, se puede expresar la condición (2.16) como

$$\mathbf{c} = A'u \quad \text{con} \quad u \geq 0$$

con la condición de holgura complementaria definida por

$$r_i u_i = 0, \quad \text{para} \quad i = 1, \dots, m$$

en la cual $r_i = a'_i \hat{x} - b_i$. Para $i \in h$, lo que garantiza que $r_i u_i = 0$ es que r_i vale cero por la definición de las restricciones activas, mientras que $i \in \bar{h}$, $u_i = 0$. El vector $u \in \mathbb{R}^m$ puede considerarse como un vector de multiplicadores de Lagrange correspondientes a las restricciones $Ax \geq b$. Combinando la igualdad y desigualdad restringida se genera el siguiente teorema.

Teorema 2.3. *Considere el problema de programación lineal*

$$\text{mín}\{c'x \mid Ax = b, \quad x \geq 0\}. \quad (2.17)$$

y suponga que \hat{x} es un punto factible, luego \hat{x} es óptimo sí y sólo sí existe $\hat{u} \in \mathbb{R}^n$ y $\hat{v} \in \mathbb{R}^m$ tal que

$$\mathbf{c} = A'\hat{\mathbf{v}} + \hat{\mathbf{u}}, \quad \hat{\mathbf{u}} \geq 0$$

$$\text{con} \quad \hat{x}_i \hat{u}_i = 0 \quad \text{para} \quad i = 1, \dots, n.$$

2.7 Dualidad

La interpretación de los multiplicadores de Lagrange del vector \mathbf{u} en el teorema (2.3), representa un primer indicio de la teoría de la dualidad en la programación lineal. Por el teorema (2.3) se tiene que en una solución \hat{x} existe un vector de multiplicadores de Lagrange $(\hat{\mathbf{v}}', \hat{\mathbf{u}})'$, tal que

$$c = A'\hat{\mathbf{v}} + \hat{\mathbf{u}}'$$

y $(a_i\hat{x}_i - b_i)\hat{\mathbf{v}}_i = 0$ para $i = 1, \dots, m$ y $\hat{x}_i\hat{\mathbf{u}}_i = 0$ para $i = 1, \dots, n$. En tal punto se observa que

$$\begin{aligned} c &= A'\hat{\mathbf{v}} + \hat{\mathbf{u}}' \\ \hat{x}c &= \hat{x}[A'\hat{\mathbf{v}} + \hat{\mathbf{u}}] \\ (\hat{x}c)' &= (\hat{x}A'\hat{\mathbf{v}}' + \hat{x}\hat{\mathbf{u}})' \\ c'\hat{x}' &= (\hat{x}A'\hat{\mathbf{v}})' + (\hat{x}\hat{\mathbf{u}})' \\ c'\hat{x} &= (A'\hat{\mathbf{v}})' \hat{x}' + \hat{\mathbf{u}}' \hat{x}' \\ c'\hat{x} &= \hat{\mathbf{v}}'A''\hat{x} + \hat{\mathbf{u}}' \hat{x} \\ c'\hat{x} &= \hat{\mathbf{v}}'A \hat{x} + \hat{\mathbf{u}}' \hat{x} \\ c'\hat{x} &= \hat{\mathbf{v}}' b \end{aligned}$$

Esto sugiere que el problema de programación lineal

$$\text{mín}\{c'x \mid Ax = b, x \geq 0\}, \quad (2.18)$$

pueda ser reformulado como un problema que permita encontrar el vector de multiplicadores de Lagrange resolviendo

$$\text{máx}\{b'\mathbf{v} \mid A'\mathbf{v} + \mathbf{u} = c, (\mathbf{u}, \mathbf{v}) \geq 0\}.$$

El vector de holgura \mathbf{u} es un residual obtenido de la restricción

$$A'\mathbf{v} + \mathbf{u} = c.$$

Para examinar la dualidad de manera general Koenker (2005) citando a Berman (1973) formuló el problema de la ecuación (2.18) en forma canónica,

$$\min_x \{c'x \mid Ax - b \in T, x \in S\}, \quad (2.19)$$

en la cual S y T son conos convexos cerrados. Luego, el problema dual correspondiente a la ecuación (2.19) es

$$\max_y \{b'y \mid c - A'y \in S^*, y \in T^*\}, \quad (2.20)$$

donde S^* y T^* son los conos duales de T y S respectivamente. Luego, para cualquier punto factible (x, y) se define el espacio dualidad

$$b'y \leq y'Ax \leq c'x,$$

el cual es la diferencia entre la función objetivo y la función dual y se representa por $c'x - b'y \geq 0$.

Definición 2.12. (Dual de un Cono)

Para cualquier conjunto K , el dual del cono se define mediante

$$K^* = \{y \in \mathbb{R}^n \mid x'y \geq 0, \text{ para toda } x \in K\}$$

2.8 Método Simplex para la Regresión Cuantílica

Para describir este método, Koenker (2005) estableció la fórmula del problema de regresión cuantílica para $\tau \in (0, 1)$

$$\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i'b), \quad (2.21)$$

la cual puede ser escrita como una fórmula de optimización lineal introduciendo dos variables artificiales

$$\{u_i, v_i : i = 1, \dots, n\}$$

para representar las partes positivas y negativas del vector residual. Esta fórmula viene dada por

$$\min\{\tau e'_n u + (1 - \tau) e'_n v \mid y - Xb = u - v, b \in \mathbb{R}^p, (u, v) \in \mathbb{R}_+^{2n}\}, \quad (2.22)$$

en la cual \mathbf{e}_n representa un n vector de unos y X es una matriz de orden $n \times p$. La fórmula (2.22) puede ser representada en su forma canónica

$$\min\{c'x \mid Ax - b \in T, x \in S\} \quad (2.23)$$

haciendo las siguientes identificaciones:

$$c = \begin{pmatrix} \mathbf{0}_p \\ \tau \mathbf{e}_n \\ (1 - \tau) \mathbf{e}_n \end{pmatrix}, \quad x = \begin{pmatrix} b \\ u \\ v \end{pmatrix}, \quad A = [X, I, -I],$$

donde $\mathbf{0}_p$ es el p vector de ceros, $b = y$, los conjuntos $T = 0_n$ y $S = \mathbb{R}^p \times \mathbb{R}_+^{2n}$ son conos convexos cerrados.

La naturaleza poliédrica del conjunto restringido y la función objetivo de la ecuación (2.21) permiten que los vértices del conjunto restringido puedan ser indexados por los $\binom{n}{p}$ elementos de $h \in \mathcal{H}$ y toman la forma

$$b(h) = X(h)^{-1}y(h),$$

$$u(h) = v(h) = 0,$$

$$u(\bar{h}) = (y - Xb(h))^+,$$

$$v(\bar{h}) = (y - Xb(h))^-.$$

Luego, las condiciones de la holgura complementaria, $u_i v_i = 0$, se mantienen para cualquier vértice y hay por lo menos p índices, $i \in h$, con $u_i = v_i = 0$. Tales puntos proporcionan un ajuste exacto para las p observaciones en el subconjunto h

y el conjunto de los vectores u, v correspondientes de la solución igual a la parte positiva y negativa del vector residual resultante.

El problema de la regresión cuantílica en (2.21) tiene su representación dual correspondiente

$$\text{máx}\{y'd \mid X'd = 0, d \in [\tau - 1, \tau]^n\}. \quad (2.24)$$

Para esto, se observa que la restricción $(c - A'y) \in S^*$ se convierte en

$$\begin{pmatrix} \mathbf{0}_p \\ \tau \mathbf{e}_n \\ (1 - \tau) \mathbf{e}_n \end{pmatrix} - \begin{pmatrix} X' \\ I_n \\ -I_n \end{pmatrix} d \in \mathbf{0}_p \times \mathbb{R}_+^{2n},$$

con $S^* = \{y \in \mathbb{R}^{2n} \times \mathbb{R}^p \mid x'y \geq 0 \text{ si } x \in S\}$ mientras que la condición $y \in T^*$ se convierte en $d \in \mathbb{R}^n$. Equivalentemente, se puede reparametrizar el problema dual en (2.24) usando

$$a = d + (1 - \tau) e_n,$$

luego, sustituyendo $d = a - (1 - \tau) e_n$ en la función (2.24)

$$\begin{aligned} X'd &= 0 \\ X'[a - (1 - \tau) \mathbf{e}_n] &= 0 \\ X'a - (1 - \tau)X'\mathbf{e}_n &= 0 \\ X'a &= (1 - \tau)X'\mathbf{e}_n \end{aligned}$$

y se produce un nuevo problema

$$\text{máx}\{y'a \mid X'a = (1 - \tau)X'\mathbf{e}_n, a \in [0, 1]^n\}. \quad (2.25)$$

En forma dual, la función 2.25 puede ser vista como un problema de optimización con respecto a un vector que está situado en un cubo unitario. Por lo tanto, para cualquier solución, $\{\hat{b}, \hat{u}, \hat{v}, \hat{d}\}$, se tiene

$$\tau \mathbf{e}'_n \hat{u} + (1 - \tau) \mathbf{e}'_n \hat{v} = y' \hat{d},$$

luego se observa que, para $i \in \bar{h}$,

$$\hat{d} = \begin{cases} \tau & \text{si } \hat{u}_i > 0, \\ (\tau - 1) & \text{si } \hat{v}_i > 0. \end{cases}$$

Mientras que para las observaciones $i \in h$ con $\hat{u}_i = \hat{v}_i = 0$, $\hat{d}(h)$ se determina de la desigualdad restringida $X'd = 0$ como

$$\hat{d}(h) = -[X(h)']^{-1}X(\bar{h})'\hat{d}(\bar{h}).$$

El vector dual $\hat{d}(h)$ en una solución correspondiente a las observaciones básicas de h es el mismo vector $-\xi_h$ del Teorema (2.4). El hecho de que $\hat{d}(h) \in [\tau - 1, \tau]^p$, en una solución, es la optimización requerida que aparece en ese resultado.

Para resolver problemas de programación lineal, usando el método Simplex, resulta conveniente hablar de las dos fases del algoritmo. En la fase I se encuentra un vértice factible del problema y en la fase II procedemos de un vértice a otro hasta que se logra la optimización.

En algunos problemas de programación lineal, solamente determinar si existe un vértice factible puede ser un poco difícil, aunque en el problema de regresión cuantílica se puede elegir cualquier subconjunto h y definir un punto básico factible para el problema provisto de que la matriz $X(h)$ sea de rango completo y de ese modo completar la fase I del algoritmo. Luego, se puede enfocar la atención sobre la fase II del cálculo, describiendo una aproximación implementada en el algoritmo de Barrodale y Roberts (1974).

Suponga que $h_0 \in \mathcal{H}$ representa el conjunto indexado correspondiente a un vértice factible inicial del conjunto restringido. Considere la derivada direccional de la función objetivo

$$R(b) = \sum_{i=1}^n \rho_{\tau}(y_i - x_i'b), \quad (2.26)$$

para cualquier $\tau \in (0, 1)$ evaluada en $b(h_0) = X(h_0)^{-1}y(h_0)$ en la dirección de δ se obtiene de la ecuación (2.3)

$$\nabla R(b(h_0), \delta) = - \sum_{i=1}^n \psi_{\tau}^*(y_i - x'_i b(h_0), -x'_i \delta) x'_i \delta$$

en la cual $\psi_{\tau}^*(y_i - x'_i b(h_0), -x'_i \delta)$ se define igual a la ecuación (2.4)

Como se está en un vértice y el conjunto restringido es convexo, se puede restringir la atención a unas ciertas direcciones extremas de δ , que corresponden al movimiento fuera del vértice a lo largo de los bordes del conjunto restringido que forman una intersección en el vértice común. Estos bordes pueden ser representados algebraicamente como

$$\eta(\alpha, h_0, \delta_j) = b(h_0) + \alpha \delta_j. \quad (2.27)$$

en la cual δ_j es la j -ésima columna de la matriz $X(h)^{-1}$ y α es un escalar que indica las posiciones que se recorren a lo largo del borde. Esta representación del borde es precisamente la representación paramétrica usual de la recta a través del punto $b(h_0)$ en la dirección δ_j . Aquí, δ_j es el vector ortogonal para la restricción formada por las observaciones básicas restantes con la j -ésima columna eliminada. Se observa que el valor de α puede ser tanto positivo como negativo en (2.27) y su derivada direccional puede ser diferente para δ_j y $-\delta_j$

Suponga que

$$d(h) = -X(h)^{-1} \sum_{i \in h} \psi_{\tau}^*(y_i - x'_i b(h)) x'_i, \quad (2.28)$$

si δ_j es la j -ésima columna de $X(h)^{-1}$, luego para $i \in h$,

$$x'_i \delta_j = \begin{cases} 0, & \text{para } j \neq i, \\ 1 & \text{para } i = j. \end{cases}$$

así, para δ_j se obtienen los gradientes de la función (2.26) representados por

$$\nabla R(b(h_0), \delta_j) = d_j(h) + (1 - \tau) \quad y \quad \nabla R(b(h_0), -\delta_j) = -d_j(h) + \tau. \quad (2.29)$$

Si estas derivadas direccionales son todas positivas para $j = 1, \dots, p$, se tiene que las condiciones de optimización del Teorema (2.3)

$$d(h) \in [\tau - 1, \tau]^p \text{ se mantienen.} \quad (2.30)$$

De lo contrario, hay un borde que es una dirección de la pendiente guiando fuera del punto $b(h_0)$ y se puede reducir $R(b)$ por el movimiento en esta dirección fuera de $b(h_0)$.

Koenker (2005) mencionó que la elección convencional del borde adaptado por Barrodale y Roberts (1974), fue la primera de las pendientes más inclinadas y, en consecuencia, las derivadas direccionales $\nabla R(h(b), \pm\delta_j)$ son más negativas. Teniendo una dirección seleccionada $\delta^* = \sigma\delta_j$ para algún j y $\sigma \in \{-1, 1\}$, se plantea la siguiente pregunta ¿Hasta qué punto debería recorrerse en esta dirección?

Barrodale y Roberts (1974) respondieron esta interrogante, más que adoptar la estrategia usual del algoritmo Simplex de ir solamente hasta el próximo vértice, se propone continuar en la dirección original siempre que, haciendo esto, se continúe reduciendo el valor de la función objetivo. Luego, hay que moverse en la dirección δ^* hasta encontrar los puntos en los cuales las observaciones que no están en la base original tienen residuales que tienden a cero. Las estrategias convencionales del algoritmo Simplex, cuando se encuentra tal situación, se introduce la nueva observación en la base y se calcula una dirección de la pendiente. En cambio, Barrodale y Roberts (1974) eligieron continuar en la dirección determinada originalmente siempre que esto produzca una dirección factible de la dirección de la pendiente. En este caso, ellos redujeron dramáticamente el número de operaciones pivote que requiere el Simplex cuando la base cambia.

En los puntos intermedios todo lo que se requiere es un cambio de signo en la contribución al gradiente de la observación, cuyo residual pasa a través del cero. Esta estrategia constituye lo que algunas veces es llamado algoritmo de Cauchy en la teoría de optimización no lineal. Una de las pendientes más inclinadas se selecciona en cada iteración y luego se busca una recta unidimensional usada para minimizar la dirección elegida en cada paso y se resuelve el siguiente problema

$$\min_{\alpha \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' b(h) - \alpha x_i' \delta^*), \quad (2.31)$$

la cual es justamente la regresión a través del origen del problema de regresión cuantílica en α con la variable dependiente $y_i - x_i' b(h)$ y un diseño compuesto de una variable individual $x_i' \delta^*$. Este problema se resuelve encontrando el cuantil ponderado. Para esto, se generaliza la caracterización de la mediana ponderada.

Considere el problema de regresión a través del origen,

$$\min_{b \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - x_i b).$$

Las soluciones básicas de $b(h)$ tienen la forma simple $b_i = \frac{y_i}{x_i}$. Las derivadas direccionales son de la siguiente forma

$$\nabla R(b, \delta) = - \sum \psi_{\tau}^*(y_i - x_i b, -x_i \delta) x_i \delta,$$

y una solución debería ser no negativa en alguna solución para $\delta \in \{-1, 1\}$. Luego, eligiendo $sgn(0) = 1$, se escribe

$$\sum (\tau - I(y_i < x_i b) x_i$$

como

$$\tau - I(y_i < x_i b) = \begin{cases} \tau - 1 & \text{si } y_i < x_i b \\ \tau & \text{si } y_i \geq x_i b \end{cases}$$

Además,

$$\begin{aligned}
y_i < x_i b &\Leftrightarrow y_i - x_i b < 0 \\
&\Leftrightarrow \operatorname{sgn}(y_i - x_i b) = -1 \\
y_i > x_i b &\Leftrightarrow y_i - x_i b > 0 \\
&\Leftrightarrow \operatorname{sgn}(y_i - x_i b) = 1
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
\tau - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(y_i - x_i b) &= \begin{cases} \tau - \frac{1}{2} + \frac{1}{2}(-1) & \text{si } y_i < x_i b \\ \tau - \frac{1}{2} + \frac{1}{2}(1) & \text{si } y_i \geq x_i b \end{cases} \\
&= \begin{cases} \tau - 1 & \text{si } y_i < x_i b \\ \tau & \text{si } y_i \geq x_i b \end{cases}
\end{aligned}$$

Luego, haciendo uso de esta equivalencia se tiene

$$\begin{aligned}
\sum (\tau - I(y_i < x_i b)x_i) &= \sum \left[\tau - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(y_i - x_i b) \right] x_i \\
&= \sum \left[\tau - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(y_i/x_i - b) \operatorname{sgn}(x_i) \right] x_i \\
&= \left(\tau - \frac{1}{2} \right) \sum x_i - \frac{1}{2} \sum |x_i| + \frac{1}{2} \sum I(y_i/x_i \geq b) |x_i|.
\end{aligned}$$

Entonces, se pueden ordenar los candidatos de las pendientes $b_i = y_i/x_i$ como $b_{(i)}$ y buscar la indexación más pequeña para j tal que la sumatoria correspondiente de las $\frac{1}{2} |x_i|$ exceda la cantidad

$$-\left(\tau - \frac{1}{2} \right) \sum x_i + \frac{1}{2} \sum |x_i|.$$

El algoritmo continúa en esta forma hasta que no hay una dirección más grande de la pendiente en la cual el punto de optimización ha sido alcanzado, y el algoritmo termina. Este algoritmo provee una aproximación extremadamente eficiente para realizar cálculos de regresión cuantílica en problemas donde el tamaño de la muestra se considera pequeño.

De acuerdo a Koenker (2012), el algoritmo utilizado por defecto en el *software* R es el método de Barrodale y Robert (1974), y el comando utilizado es `method = "br"`.

Definición 2.13. (Envoltura Convexa)

La envoltura convexa de un conjunto C , denotado por $\mathbf{Conv}C$, es el conjunto de todas las combinaciones convexas de los puntos que están en C y su fórmula está dada por:

$$\mathbf{Conv}C = \{\theta_1 x_1, \dots, \theta_n x_n \mid x_i \in C, \theta_i \geq 0 \quad \text{para} \quad i = 1, \dots, k\}$$

Por otro lado, la envoltura convexa de un conjunto finito $\{u_1, \dots, u_k\}$ se define por

$$\mathbf{Conv}\{u_1, \dots, u_k\} = \{\theta_1 u_1, \dots, \theta_n u_k \mid \theta_i \geq 0, \mathbf{1}'\theta = \mathbf{1}\}$$

2.9 Condición del Subgradiente

La condición básica de optimización para el problema de regresión cuantílica

$$\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i' b), \tag{2.32}$$

y la solución básica

$$b(h) = X(h)^{-1} y(h),$$

con $h \in \mathcal{H}^* = \{h \in \mathcal{H} : |X(h)| \neq 0\}$. Se observa que la condición de optimización implica verificar que las derivadas direccionales son no negativas en todas las

direcciones. Para comprobar esto, en $b(h)$ se considera el gradiente

$$\nabla R(b(h), \omega) = - \sum_{i=1}^n \psi_{\tau}^*(y_i - x'_i b(h), -x'_i \omega) x'_i \omega.$$

Luego, reparametrizando las direcciones con $v = X(h)\omega$, se tiene la optimización si y sólo si

$$0 \leq - \sum_{i=1}^n \psi_{\tau}^*(y_i - x'_i b(h), -x'_i X(h)^{-1} v) x'_i X(h)^{-1} v$$

para todo $v \in \mathbb{R}^p$. Se observa que cuando $i \in h$, se tiene $e'_i = x'_i X(h)^{-1}$, la i -ésima base unitaria del vector de \mathbb{R}^p y se escribe

$$0 \leq - \sum_{i \in h} \psi_{\tau}^*(0, -v_i) v_i - \xi' v = - \sum_{i \in h} (\tau - I(-v_i < 0)) v_i - \xi' v,$$

en la cual

$$\xi' = \sum_{i \in \bar{h}} \psi_{\tau}^*(y_i - x'_i b(h), -x'_i X(h)^{-1} v) x'_i X(h)^{-1} v.$$

Luego, el espacio de las direcciones $v \in \mathbb{R}^p$ es generado por las direcciones

$$v = \pm e_k \quad \text{para} \quad k = 1, \dots, p.$$

Por lo tanto, la condición de las derivadas direccionales se mantiene para todo $v \in \mathbb{R}^p$ si y sólo si se mantiene para las $2p$ direcciones canónicas

$$\{\pm e_i : i = 1, \dots, p\}$$

Así, para $v = e_i$ se tienen p desigualdades

$$0 \leq -(\tau - 1) - \xi_i(h), \quad i = 1, \dots, p$$

mientras que para $v = -e_i$ se cumple

$$0 \leq \tau + \xi_i(h), \quad i = 1, \dots, p.$$

Combinando ambas desigualdades, se tiene la condición de optimización en su generalización completa. Si ninguno de los residuales de las observaciones no básicas, $i \in \bar{h}$, son iguales a cero, lo que podría darse en el caso de que la probabilidad sea 1 si las y variables tienen una densidad respecto a la medida de Lebesgue, entonces la dependencia de ξ sobre v desaparece y se combinan las dos desigualdades para producir

$$-\tau 1_p \leq \xi(h) \leq (1 - \tau) 1_p.$$

De acuerdo a lo mencionado se reformula el teorema 3.3 de Koenker y Bassett (1978) introduciendo la definición de posición general dada en Rousseeuw y Leroy (1987).

Definición 2.14. (Posición General)

Las observaciones de la regresión (y, X) están en posición general si cualquier p elemento de ellas produce un ajuste exacto único, para cualquier $h \in \mathcal{H}$,

$$y_i - x_i b(h) \neq 0 \quad \text{para cualquier } i \notin h.$$

Teorema 2.4. *Si (y, X) están en posición general, entonces existe una solución para el problema de regresión cuantílica (2.32) de la forma $b(h) = X(h)^{-1}y(h)$ si y sólo si para algún $h \in \mathcal{H}$,*

$$-\tau 1_p \leq \xi(h) \leq (1 - \tau) 1_p,$$

en la cual $\xi'(h) = \sum_{i \in \bar{h}} \psi_\tau(y_i - x_i' b(h)) x_i' X(h)^{-1} \quad y \quad \psi_\tau = \tau - I(u < 0)$. Además, $b(h)$ es la única solución si y sólo si las desigualdades son estrictas; en otro caso, el conjunto solución es la envoltura convexa de varias soluciones de la forma $b(h)$.

Demostración. Ver Koenker y Bassett (1978) □

CAPITULO 3

INFERENCIA SOBRE LOS ESTIMADORES DE LA REGRESIÓN CUANTÍLICA

De acuerdo a Kecojević (2011), los métodos de inferencia para la regresión cuantílica se clasifican en: Estimación directa de la matriz de varianza-covarianza asintótica, la Transformación Rango y el método del *Bootstrap*.

Los métodos basados en la Estimación directa de la matriz de varianza-covarianza no son muy precisos debido a que éstos dependen de la estimación del recíproco de la densidad del error, en el cuantil de interés, una cantidad conocida como la función esparcida. Esta cantidad refleja la densidad de las observaciones cercanas al cuantil dado. Si los datos están muy dispersos del cuantil de interés, éste sería difícil de estimar con exactitud. Sin embargo, en caso contrario, cuando hay muchas observaciones cercanas al cuantil dado hay menor dispersión, luego el cuantil puede ser estimado con mayor precisión. Otra aproximación para hacer inferencias sobre los estimadores de la regresión cuantílica es el método de la Transformación Rango. Este método evita la selección de los parámetros de suavizado de Hall & Sheather (1988) y Bofinger (1975).

Las pruebas del multiplicador de Lagrange para la regresión cuantílica son, en efecto, una clase de prueba de Rango generalizadas para el modelo de regresión cuantílica. Estas pruebas están basadas en la regresión Transformación Rango, son robustas a los datos atípicos en la variable dependiente y tienen una distribución asintóticamente libre en el sentido de que ningún parámetro que dependa de la distribución debe ser estimado para determinar el estadístico de prueba. En la construcción de los intervalos de confianza resulta necesario invertir la prueba de Rango y relacionar el dual de la regresión cuantílica con las funciones clásicas de Hájek y Šidák (1965). La principal ventaja de este método es que evita el problema

de la estimación de la función esparcida utilizada en la estimación directa.

El método de remuestreo introducido por Efron (1979), es un método de Monte Carlo utilizado para aproximar la distribución muestral de un estimador del parámetro. La ventaja de este método es que hace pocas nociones acerca de la forma de la distribución muestral. Existen varios métodos de remuestreo para determinar el error estándar en la regresión cuantílica. Entre éstos se tienen: el *bootstrap* del residual, el *bootstrap* del par (x, y) y el *bootstrap* sobre la condición del subgradiente.

De acuerdo a Kecojević (2011), el *bootstrap* del residual es una técnica no paramétrica que hace inferencias de la probabilidad basada en una característica de la población, Θ , con su estimador, $\hat{\Theta}$, usando una muestra etiquetada de la población. El método del *bootstrap* para el residual depende de una analogía entre la muestra y la población del cual se etiqueta una muestra y se considera ésta como si fuese la población. Los datos son remuestreados con reemplazamiento muchas veces para obtener un estimador empírico de la distribución muestral del estadístico de interés Θ . La técnica del *bootstrap* del residual para la regresión no lineal de la mediana fue adaptada para el ajuste de la regresión cuantílica pero, en este caso, la autora, citando a Efron (1979), afirmó que el método sólo resultó útil en los modelos de regresión cuantílica con errores i.i.d. Mientras que el método del *bootstrap* para el par (x, y) tiene una distribución independiente pero no idénticamente distribuida, además, provee una alternativa efectiva que se adapta a los modelos con heteroscedasticidad.

A diferencia del *bootstrap* para el residual, que etiqueta muestras de la distribución empírica de los residuales, el método del par (x, y) etiqueta muestras para la distribución conjunta de la muestra. Luego, el par (x_i^*, y_i^*) se etiqueta con reemplazamiento de los n pares de una muestra definida por

$$\{(x_i, y_i) : i = 1, \dots, n\}$$

con una probabilidad de $1/n$.

Cuando se implementa la técnica del *bootstrap* en la regresión cuantílica, hay que tomar en cuenta la elección del número de B repeticiones. Andrew y Buchinsky (2000), citado por Kecojević (2011), establecieron un procedimiento para determinar cuántas veces deben realizarse el número de B repeticiones para poder alcanzar un nivel de precisión basado en la distribución asintótica. El método *bootstrap* de Parzen, et al. (1994) sugiere el uso de un algoritmo poco común. En vez de remuestrear las observaciones o los residuales, se aplica el remuestreo sobre la condición del subgradiente. Como el estadístico $S_x(\boldsymbol{\beta})$ es una cantidad de pivote para el τ -ésimo parámetro de la regresión cuantílica, su distribución puede ser aproximada por un vector aleatorio \mathbf{U} de orden $p \times 1$ el cual es una suma ponderada de variables de Bernoulli. Para cualquier valor $\mathbf{u}_j \in \mathbf{U}$, el estadístico de pivote $S_x(\boldsymbol{\beta})$ es igual a \mathbf{u}_j para $j = 1, \dots, B$ y se repite este procedimiento B veces hasta obtener la solución del conjunto de vectores $\hat{\boldsymbol{\beta}}_{u_j}$, los cuales se usan para realizar la inferencia estadística.

En este Capítulo se presentan algunos de los métodos utilizados por el *software* R para obtener los errores estándares y los intervalos de confianza de la regresión cuantílica.

3.1 Asintóticos de la Regresión Cuantílica con Errores I.I.D.

Considere el modelo lineal para la τ -ésima función condicional cuantílica de una variable dependiente Y , dada las covariables $x_i \in \mathbb{R}^p$,

$$Q_Y(\tau/x) = \mathbf{x}'_i \boldsymbol{\beta}(\tau)$$

para $i = 1, \dots, n$. Luego, se pueden construir intervalos de confianza para la τ -ésima coordenada de $\boldsymbol{\beta}_\tau$. Para esto se obtiene un estimador puntual $\hat{\boldsymbol{\beta}}(\tau)$ del

parámetro $\boldsymbol{\beta}(\tau)$ mediante

$$\min_{\mathbf{b} \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}'_i \mathbf{b})$$

y la función ρ_{τ} se define como en (1.9).

En el caso más simple, para el modelo de la regresión lineal,

$$Y_i = \mathbf{x}'_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

donde $\boldsymbol{\varepsilon}_i \equiv Y_i - \mathbf{x}'_i \boldsymbol{\beta}$, deben cumplirse las siguientes condiciones:

- (1) Los errores $\{\boldsymbol{\varepsilon}_i\}$ son independientes e idénticamente distribuidos (*i.i.d.*).
- (2) Los errores $\{\boldsymbol{\varepsilon}_i\}$ tienen una función de distribución acumulada F continuamente diferenciable con su respectiva función de densidad de probabilidad f , la cual es continua y estrictamente positiva. Además, $f = F'$ y $f(F^{-1}(\tau)) > 0$, en una vecindad de τ .
- (3) $n^{-1} \sum \mathbf{x}_i \mathbf{x}'_i \equiv D_n$ converge a una matriz positiva D_0 y \mathbf{x}_i es la i -ésima fila de X .

Bajo las condiciones del diseño de Mild, Koenker (1994) formuló la siguiente distribución asintótica

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(0, \omega^2(\tau, F)D^{-1}). \quad (3.1)$$

con $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + F^{-1}(\tau)\mathbf{e}_1$, en la cual \mathbf{e}_1 es un vector $p \times 1$ con base unitaria $(1, 0, \dots, 0)'$ de \mathbb{R}^p , $D = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i$ y $\omega^2(\tau, F) = \tau(1 - \tau)/f^2(F^{-1}(\tau))$ es la matriz de varianza-covarianza.

3.2 Asintóticos de la Regresión Cuantílica con Errores no I.I.D.

De acuerdo a Koenker (2005), el modelo de regresión lineal con error *i.i.d.* produce una forma simple para la distribución limitante del estimador $\hat{\boldsymbol{\beta}}(\tau)$ de

la regresión cuantílica. Sin embargo, debe argumentarse que, en el cambio de la forma de las medidas de posición para modelos con error i.i.d., la regresión cuantílica es muy excesiva; en este caso, cualquier estimador de la tendencia central condicional de la variable dependiente, dadas las covariables, es perfectamente adecuado. En los ajustes del error no i.i.d., como el modelo condicional de las medidas de posición y la escala, la teoría asintótica de $\hat{\boldsymbol{\beta}}(\tau)$ es un poco más complicada, la matriz de varianza-covarianza de la distribución

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau))$$

toma la forma del Sandwich de Huber (1967),

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(0, \tau(1 - \tau)\mathbf{H}_n^{-1}\mathbf{J}_n\mathbf{H}_n^{-1}), \quad (3.2)$$

en la cual

$$\begin{aligned} \mathbf{J}_n(\tau) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i', \\ H_n(\tau) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' f_i(\xi_i(\tau)). \end{aligned} \quad (3.3)$$

La densidad condicional de la variable dependiente y_i , evaluada en el τ -ésimo cuantil condicional, está representada por $f_i(\xi_i(\tau))$. Bajo las condiciones de Mild sobre las f_i y x_i , se tiene la normalidad asintótica disjunta para los vectores $\xi(\tau_i), \dots, \xi(\tau_m)$. En el caso de tener una distribución con errores i.i.d., las f_i son idénticas y el Sandwich de Huber toma la forma de la ecuación (3.2).

3.3 Estimación del Escalar Esparcido (*Sparsity*)

El método del escalar esparcido es el más directo y rápido, involucra un estimador de la función esparcida, este estimador es sensible a las nociones de i.i.d. En este caso, Koenker y Machado (1999) utilizaron un estimador local de la función densidad expresado como la diferencia de los cocientes de funciones cuantílicas.

La dificultad de este método es la selección del parámetro de suavizado cuando se usa la diferencia del cociente. En una muestra pequeña, tanto el parámetro de Bofinger (1975) como el de Hall Sheather (1988) tienden a ser muy grandes para asegurar la linealidad local de la función condicional cuantílica.

Tomando en cuenta la distribución asintótica de la regresión cuantílica, se pueden construir los intervalos de confianza. Koenker (2005) afirmó que los estimadores de la regresión cuantílica dependen del recíproco de la función de densidad evaluada en un cuantil.

$$s(\tau) = [f(F^{-1}(\tau))]^{-1} \quad (3.4)$$

una cantidad que Turkey (1965), ha calificado como función esparcida y Parzen (1979) reconoció como función de densidad cuantílica. Para estimar $s(\tau)$, Siddiqui (1960) consideró la diferenciación de la identidad $F(F^{-1}(t)) = t$, luego se tiene que la función esparcida es la derivada de la función cuantílica.

$$\frac{d}{dt}F^{-1}(t) = s(t).$$

Tomando en cuenta las sugerencias de Siddiqui (1960), en la construcción de los intervalos de confianza para los cuantiles muestrales se puede estimar $s(t)$ por medio de la diferencia del cociente de la función cuantílica

$$\hat{s}_n(t) = [\hat{F}_n^{-1}(t + h_n) - \hat{F}_n^{-1}(t - h_n)]/2h_n, \quad (3.5)$$

en la cual h_n es un parámetro de suavizado que tiende a cero cuando $n \rightarrow \infty$. Este parámetro puede ser de Bofinger o de Hall Sheather y $\hat{F}^{-1}(\cdot)$ es el τ -ésimo cuantil de los residuales.

Koenker (2005) citando a Bofinger (1975) mencionó que este autor utilizó la estimación asintótica de la densidad y demostró que la secuencia del parámetro de suavizado esta dada por

$$h_n = n^{-1/5} \left[\frac{4,5 s^2(t)}{(s''(t))^2} \right]^{1/5}, \quad (3.6)$$

fue la elección óptima de h_n bajo las condiciones de Mild sobre F . Este resultado se obtiene de la densidad estándar de la estimación asintótica. Obviamente, si se conocen $s(t)$ y $s''(t)$ no se necesitaría usar h_n , pero afortunadamente $s(t)/s''(t)$ no son muy sensibles a F y se perdería poco si se sustituye alguna forma de distribución, por ejemplo, gaussiana.

Sheather y Maritz (1983) discutieron la estimación de s y s'' como un medio de establecer una conexión con h_n . En forma general, la proporción

$$\frac{s(t)}{s''(t)} = \frac{f^2}{2(f'/f)^2 + [(f'/f)^2 - f''/f]}. \quad (3.7)$$

Sí f es gaussiana, $(f'/f)(F^{-1}(t)) = \Phi^{-1}(t)$ y el término entre paréntesis elevado al cuadrado es igual a 1; luego, usando la ecuación (3.7), el parámetro de suavizado de Bofinger se convierte en

$$h_n = n^{-1/5} \left[\frac{4, 5\phi^4(\Phi^{-1}(t))}{(2\Phi^{-1}(t)^2 + 1)^2} \right]^{1/5}.$$

En la cual $\phi(\cdot) = f$ es la función de densidad de probabilidad y $\Phi(\cdot)$ es la función de densidad acumulada de una variable aleatoria normal.

Basados en las expansiones de Edgeworth, para cuantiles muestrales estudentizados, Hall y Sheather (1988) propusieron el parámetro de suavizado

$$h_n = n^{-1/3} Z_\alpha^{2/3} \left[\frac{1, 5 s(t)}{s''(t)} \right]^{1/3} \quad (3.8)$$

en el cual Z_α satisface $\phi(Z_\alpha) = 1 - \alpha/2$ para la construcción de intervalos de confianzas del tipo $1 - \alpha$, α representa el tamaño deseado para la prueba y n es el tamaño muestral. Asimismo, usando la fórmula (3.7), el parámetro de suavizado de Hall y Sheather (1988) se convierte en

$$h_n = n^{-1/3} Z_\alpha^{2/3} \left[\frac{1, 5 \phi^2(\Phi^{-1}(t))}{(2\Phi^{-1}(t)^2 + 1)^2} \right]^{1/3}. \quad (3.9)$$

En la figura (3.1) se grafica h_n versus el tamaño muestral, el parámetro de suavizado de Hall & Sheather (1988) se representa con líneas punteadas, y el

parámetro de suavizado de Bofinger (1975) con líneas sólidas, respectivamente, para los cuantiles $\tau = 0,50; 0,75$ y $0,95$.

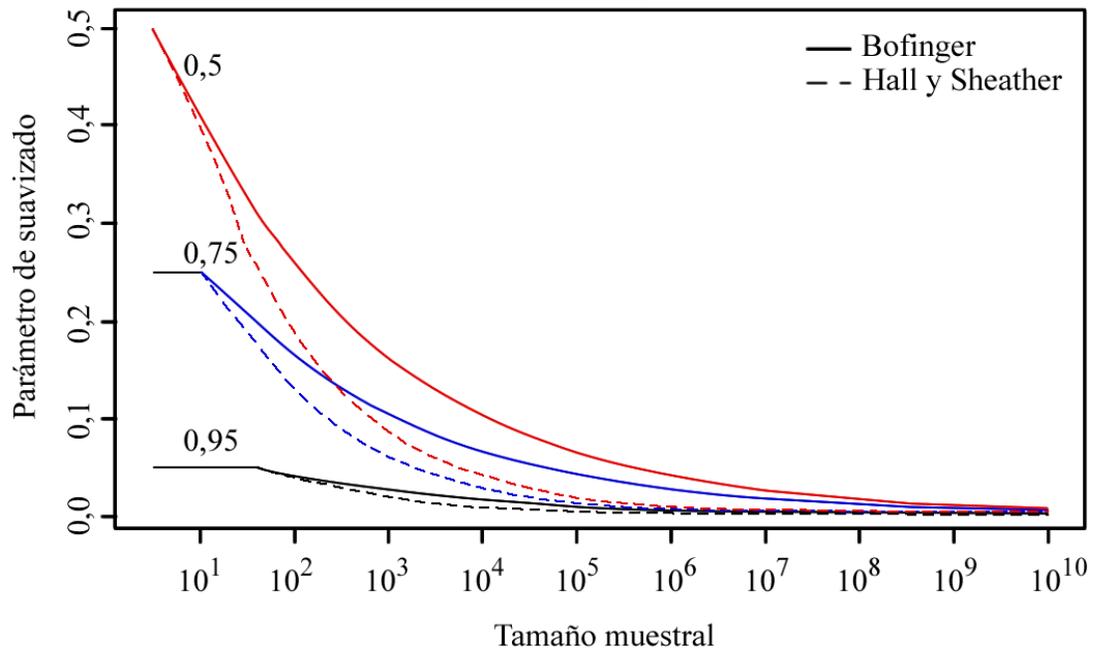


Figura 3.1: Parámetro de Suavizado h_n . Las curvas punteadas representan el parámetro de Hall & Sheather y las curvas sólidas el parámetro de Bofinger para los cuantiles 0,50; 0,75 y 0,95.

Además, se observa que h_n tiene un valor más grande cuando se usa el parámetro de suavizado de Hall & Sheather (1988) para muestras pequeñas, y cuando el tamaño de la muestra se incrementa de moderado a grande, el parámetro de suavizado de Hall & Sheather produce valores más pequeños para h_n que el parámetro de Bofinger (1975).

En la figura (3.2) se observa que las sucesiones de parámetros de suavizado h_n para los cuantiles τ y $1 - \tau$, cuando F tiene una distribución normal, son idénticamente iguales.

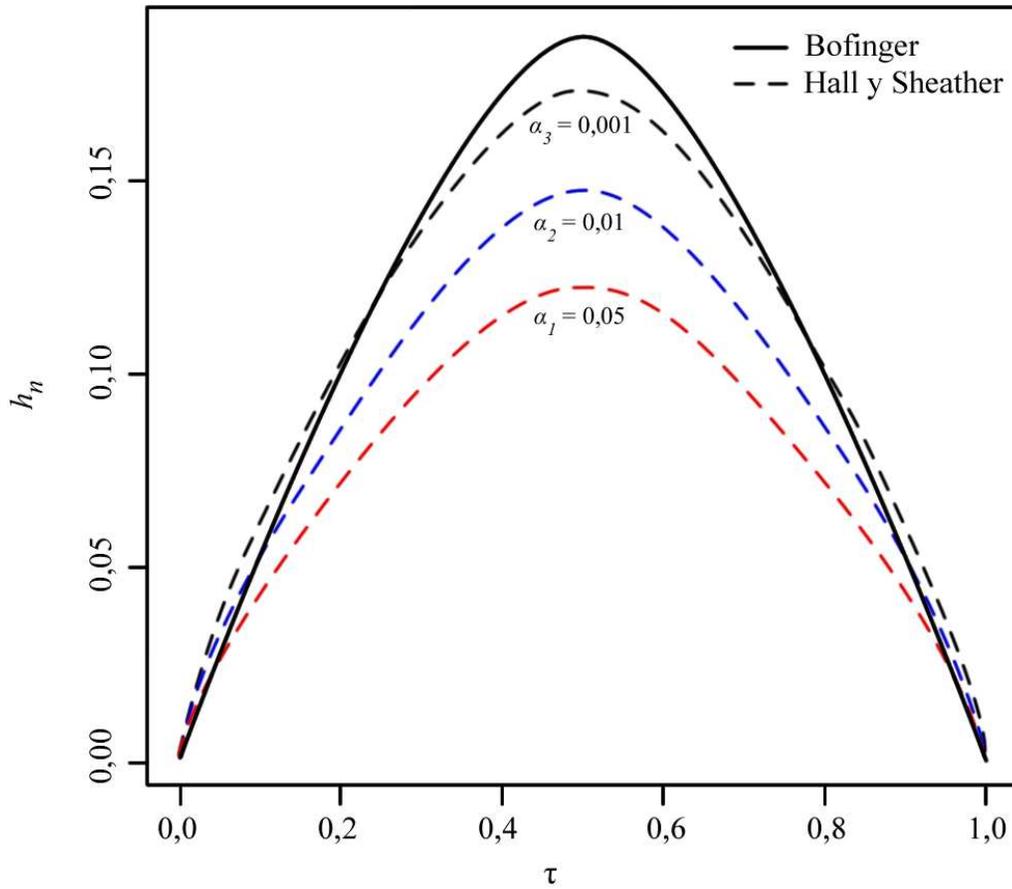


Figura 3.2: Diferentes elecciones del Parámetro de Suavizado h_n para una muestra de tamaño $n=500$, basada en la distribución gaussiana con diferentes niveles de significancia $\alpha_1 = 0,05$; $\alpha_2 = 0,01$ y $\alpha_3 = 0,001$.

Una vez elegido el parámetro de suavizado, se debe determinar el valor \hat{F}_n^{-1} de la ecuación (3.5), la aproximación más simple es usar los residuales de la regresión cuantílica. Suponga que $\{u_i : i = 1, \dots, n\}$ son esos residuales, y su estimador

$$\hat{u}_i = y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}(\tau),$$

para $i = 1, \dots, n$. Luego, la función cuantílica empírica está dada por

$$\hat{F}_n^{-1}(\tau) = \hat{u}_i$$

para $\tau \in \left[\frac{i-1}{n}, \frac{i}{n} \right)$

Otra alternativa para estimar \hat{F}^{-1} consiste en utilizar la función cuantílica sugerida en Bassett y Koenker (1982), definida por

$$F_n^{-1}(\tau) = \bar{\mathbf{x}}' \hat{\boldsymbol{\beta}}(\tau),$$

en el cual $\bar{\mathbf{x}} = n^{-1} \sum x_i$ y $\hat{\boldsymbol{\beta}}(\cdot)$ son los estimadores de la regresión cuantílica.

De acuerdo a Kocherginsky, et al. (2005), el paquete *quantreg* del *software* R para la estimación cuantílica, utiliza la opción `se="iid"` de la función *summary.rq* y el parámetro de suavizado utilizado por defecto es el de Hall y Sheather (1988). La función para calcular el parámetro de suavizado por defecto es `hs="TRUE"`, si se quiere elegir el parámetro de suavizado de Bofinger (1975), se coloca en R el comando `hs="FALSE"`.

Además, Kecojević (2011) afirmó que la aproximación de la Estimación esparcida incorporada en el paquete *quantreg* de R asume homocedasticidad, así la covariable no afecta cualquier otro aspecto de la distribución condicional de la variable dependiente.

3.4 Estimación de la Matriz de Varianza-Covarianza para Modelos con Error no I.I.D.

En esta sección se describirán dos aproximaciones para la estimación de la matriz $\mathbf{H}_n(\tau)$. La primera es una extensión de los métodos de la estimación del escalar esparcido, sugerido por Hendricks y Koenker (1991), la cual asume la linealidad de la τ -ésima función cuantílica condicional. Para un parámetro de suavizado $h_n \rightarrow 0$, se puede estimar consistentemente los parámetros de las

funciones cuantílicas condicionales $\tau \pm h$ mediante $\hat{\beta}(\tau \pm h)$. La densidad $f_i(\xi_i)$ pueden ser estimada usando la diferencia del cociente.

$$\hat{f}_i(\xi_i(\tau)) = \frac{2h_n}{\mathbf{x}'_i(\hat{\beta}(\tau + h_n) - \hat{\beta}(\tau - h_n))} \quad (3.10)$$

Sustituyendo el estimador de la ecuación (3.10) en la siguiente expresión

$$H_n(\tau) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i f_i(\xi_i(\tau)) \quad (3.11)$$

se obtiene un estimador de la matriz de varianza-covarianza de $\hat{\beta}(\tau)$ para los modelos con errores no i.i.d. El estimador $\hat{f}_i(\xi_i(\tau))$ se puede escribir

$$\hat{f}_i(\xi_i(\tau)) = \frac{2h_n}{d_i},$$

luego, la cantidad

$$d_i = \mathbf{x}'_i(\hat{\beta}(\tau + h_n) - \hat{\beta}(\tau - h_n)), \quad (3.12)$$

es necesariamente positiva sólo cuando

$$x = \bar{x} = n^{-1} \sum_{i=1}^n \mathbf{x}_i.$$

En caso contrario, cuando las observaciones de la muestra no son positivas, existen problemas con la estimación de $\hat{f}_i(\xi_i(\tau))$. Para solucionar este tipo de problemas Koenker (2005) reemplazó las \hat{f}_i por su parte positiva

$$\hat{f}_i^+ = \max \left\{ 0, \frac{2h_n}{d_i - \varepsilon} \right\}$$

con un parámetro de tolerancia $\varepsilon > 0$.

La otra aproximación está basada en la estimación de la densidad del Kernel propuesta por Powell (1991), para estimar la matriz $\mathbf{H}_n(\tau)$, el autor propuso el siguiente estimador

$$\hat{\mathbf{H}}_n(\tau) = (nh_n)^{-1} \sum_{i=1}^n K(\hat{u}_i(\tau)/h_n) \mathbf{x}_i \mathbf{x}'_i,$$

con $\hat{u}_i(\tau) = y_i - \mathbf{x}'_i \hat{\beta}(\tau)$ y h_n es un parámetro de suavizado que cumple con las condiciones $h_n \rightarrow 0$ y $\sqrt{nh_n} \rightarrow \infty$.

El Kernel K es una función de densidad de probabilidad definida por

$$K(x) : [0, 1] \rightarrow \mathbb{R}$$

y tiene las siguientes propiedades:

- i. $\int K(x)dx = 1$
- ii. $\int xK(x)dx = 0$
- iii. $\int x^2K(x)dx = \sigma_k^2$

Asimismo, Koenker (2005) citando a Powell (1991) estableció que la elección del Kernel y parámetro de suavizado viene dado por

$$\hat{H}_n(\tau) = \frac{1}{2n c_n} \sum I(|\hat{u}_i| < c_n) \mathbf{x}_i \mathbf{x}'_i,$$

en la cual $c_n \rightarrow 0$ y $\sqrt{nc_n} \rightarrow \infty$ con un parámetro de suavizado

$$c_n = k(\Phi^{-1}(\tau + h_n) - \Phi^{-1}(\tau - h_n)),$$

donde k es un estimador robusto de la escala y $h_n = \mathcal{O}(n^{-1/3})$

De acuerdo a Kocherginsky, et al. (2005), la aproximación de Hendricks y Koenker (1991), en el paquete *quantreg* de R está dada por `se="nid"` y la aproximación de Powell (1991), se encuentra implementada con la opción `se="ker"`.

3.5 Intervalos de Confianza para la Regresión Cuantílica Basados en la Transformación Rango (*Rank scores*)

De acuerdo a Koenker (1994), la teoría clásica de la Transformación Rango puede ser usada para realizar pruebas de hipótesis. Considere un modelo de regresión lineal

$$\mathbf{Y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{\Gamma} \mathbf{U}, \tag{3.13}$$

en el cual \mathbf{Y} es un vector $n \times 1$ de variables dependiente, \mathbf{X}_1, X_2 son vectores $n \times p$ de variables independientes, \mathbf{U} es un vector aleatorio de orden $n \times 1$ y $\mathbf{\Gamma}$ es una matriz $n \times n$. En este caso, $\boldsymbol{\beta}_1(\tau)$ y $\boldsymbol{\beta}_2(\tau)$ son vectores de parámetros desconocidos, el primero de orden $p \times 1$ y el segundo de orden $q \times 1$, especificados por la hipótesis nula

$$H_0 : \boldsymbol{\beta}_2(\tau) = \mathbf{0}$$

la cual se prueba en contra de la hipótesis alternativa

$$H_1 : \boldsymbol{\beta}_2(\tau) \neq \mathbf{0}.$$

Tomando en cuenta el vector $n \times 1$, de la Transformación Rango

$$\mathbf{R}(\tau) = \hat{\mathbf{a}} - (1 - \tau)\mathbf{1},$$

en el cual $\mathbf{1}$ representa un vector de unos de orden $n \times 1$, $\hat{\mathbf{a}}$ es la función Transformación Rango definida por Hájek y Šidák (1967), mediante

$$\hat{\mathbf{a}}_i(\tau) = \begin{cases} 1 & \text{si } t \leq (R_i - 1)/n \\ R_i - t_n & \text{si } (R_i - 1)/n < t \leq R_i/n \\ 0 & \text{si } R_i/n < t \end{cases}$$

y tiene las siguientes propiedades:

1. $\hat{\mathbf{a}}_i(\tau)$ es continua.
2. $\hat{\mathbf{a}}_i(\tau)$ es lineal por segmentos.
3. $\hat{\mathbf{a}}_i(\tau) = 1$ cuando (x_i, y_i) está por debajo del plano estimado $H(x) = x'\hat{\boldsymbol{\beta}}(\tau)$.
4. $\hat{\mathbf{a}}_i(\tau) = 0$ cuando $y_i > x'_i\hat{\boldsymbol{\beta}}(\tau)$.
5. $0 \leq \hat{\mathbf{a}}_i(\tau) \leq 1$.

Además, R_i es el rango de la i -ésima observación Y_i , para la muestra $\{Y_1, \dots, Y_n\}$.

En este caso, el estadístico de prueba viene dado por

$$T_\tau = \frac{\mathbf{S}'(\tau)\mathbf{Q}^{-1}\mathbf{S}(\tau)}{\tau(1-\tau)}$$

con

$$\begin{aligned}\mathbf{S}'(\tau) &= n^{-1/2}(\mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X}_2)'\mathbf{R}(\tau), \\ \mathbf{Q} &= n^{-1}\mathbf{X}'_2(I - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1)\mathbf{X}_2\end{aligned}$$

y T_τ tiene una distribución asintótica χ^2 con q grados de libertad.

Mediante la inversión de la prueba de rango se provee un método de estimación para un intervalo de confianza de los parámetros del modelo de la regresión cuantílica. El método de la Transformación Rango evita calcular la función esparcida y es muy robusto a las nociones distribucionales del modelo.

Koenker (1994) definió la función *score* para el τ -ésimo cuantil como:

$$\varphi_\tau : (0, 1) \rightarrow \mathbb{R}$$

$$\varphi_\tau = \tau - I(t < \tau),$$

en la cual I es la función indicatriz y φ es la τ -ésima función creciente dos veces integrable en el intervalo $(0, 1)$. Luego, integrando $\hat{\mathbf{a}}_i$ con respecto a la función *score*, se obtiene el i -ésimo vector de rango

$$\hat{b}_i = - \int_0^1 \varphi_\tau(t) d\hat{\mathbf{a}}_i(t) = \hat{\mathbf{a}}_i(\tau) - (1 - \tau) \quad \text{para} \quad i = 1, \dots, n.$$

Considerando el modelo definido en (3.13) con una prueba de hipótesis

$$H_0 : \beta_2 = \xi$$

basada sobre $\hat{\mathbf{a}}_i$, se resuelve el problema de programación lineal

$$\text{máx}\{(y - x_2\xi)'\mathbf{a} \mid \mathbf{X}'_1\mathbf{a} = (1 - \tau)\mathbf{X}'_1\mathbf{1}, \mathbf{a} \in [0, 1]^n\} \quad (3.14)$$

y el hecho de que

$$\mathbf{S}_n(\xi) = n^{-1/2} \mathbf{X}_2 \hat{\mathbf{b}}_n(\xi) \rightsquigarrow \mathcal{N}(0, A^2(\varphi_\tau) q_n^2)$$

en la cual $\hat{\mathbf{b}}_n = (\hat{\mathbf{b}}_{n1}, \dots, \hat{\mathbf{b}}_{nn})'$, y se definen

$$A^2(\varphi_\tau) = \tau(1 - \tau)$$

$$\text{y } q_n^2 = n^{-1} \mathbf{X}'_2 (I - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1) \mathbf{X}_2.$$

Luego, se encuentra el valor del estadístico de prueba

$$T_n(\xi) = \frac{\mathbf{S}_n(\xi)}{(A(\varphi_\tau) q_n)}$$

con $A(\varphi_\tau) = \sqrt{\tau(1 - \tau)}$ y se rechaza H_0 si $|T_n(\xi)| > \Phi^{-1}(1 - \alpha/2)$.

Si se quiere un intervalo de confianza para ξ , se invierte la prueba de rango y luego se encuentran los valores de ξ tal que no se rechace la prueba. Por lo tanto, la ecuación (3.14) puede ser vista como un problema de programación lineal uniparamétrico en ξ . El vector dual $\hat{\mathbf{a}}_i(\xi)$ es constante por segmentos. Luego, el valor de ξ puede alterarse sin comprometer la optimización de $\hat{\mathbf{a}}_i(\xi)$ siempre que el signo de los residuales en el problema principal de la regresión cuantílica no cambie. Cuando ξ alcanza tal acotamiento, la solución cambia pero la optimización puede restablecerse haciendo un pivote del método Simplex. El proceso continúa en esta dirección hasta que $T_n(\xi)$ supera el valor crítico específico. Por lo tanto, $T_\tau(\xi)$ es constante por segmentos, luego se utiliza la interpolación lineal en ξ para obtener el nivel esperado para el intervalo de confianza, el cual es asimétrico, pero se encuentra centrado en el punto de estimación $\hat{\boldsymbol{\beta}}_2(\tau)$ en el sentido que $T_n(\hat{\boldsymbol{\beta}}_2(\tau)) = 0$. Esto se deduce inmediatamente de la restricción $\mathbf{X}'\hat{\mathbf{a}} = (1 - \tau)\mathbf{X}'_1\mathbf{1}$ en el problema completo.

La ventaja principal de esta aproximación es que ésta hereda la propiedad de la invarianza escalar del estadístico de prueba T_n y, por lo tanto, evita el problema de estimación de la función esparcida.

Kocherginsky, et al. (2005) afirmaron que el método de la Transformación Rango implementado por defecto en el paquete *quantreg* de R, utiliza la función `se="rank"` cuando la distribución del error es i.i.d. Además, cuando la muestra es de tamaño $n < 1001$ este método produce límites de confianza inferiores y superiores pero, para problemas de tamaño superior, el método por defecto es estimar la matriz de varianza-covarianza asintótica utilizando el comando `se="nid"` para producir los errores estándares en cada cuantil.

3.6 Remuestreo del Residual.

En el ajuste de la regresión cuantílica lineal, suponga que

$$\hat{\beta}(\tau) = \min \sum_{i=1}^n \rho_{\tau}(y_i - x_i' b)$$

y \hat{F}_n representa la distribución empírica de los residuales,

$$\hat{\mathbf{u}}_i = y_i - \mathbf{x}_i' \hat{\beta}(\tau),$$

para $i = 1, \dots, n$ con respecto a su ajuste:

$$\hat{F}_n(\mathbf{u}) = n^{-1} \sum_{i=1}^n I(\hat{\mathbf{u}}_i < u),$$

en la cual $I(\cdot)$ es la función indicatriz.

Ahora, se consideran las muestras del *bootstrap* etiquetadas $\{\mathbf{u}_i^*, \dots, \mathbf{u}_n^*\}$, correspondientes a la función $\hat{F}_n(u)$ con reemplazamientos. Luego, se genera un vector remuestreado de la variable dependiente, el remuestreo se genera añadiendo el vector remuestreado de los residuales al vector de los valores estimados de la variable dependiente,

$$\hat{\mathbf{y}}_i = \mathbf{x}_i' \hat{\beta}(\tau),$$

desde la muestra

$$\mathbf{y}_i^* = \mathbf{x}_i' \hat{\beta}(\tau) + \mathbf{u}_i^* \tag{3.15}$$

luego se usa la ecuación (3.15) para estimar los coeficientes de la regresión cuantílica determinando

$$\hat{\boldsymbol{\beta}}^* = \min_{\boldsymbol{\beta} \in \mathbb{R}^k} \sum \rho_\tau(\mathbf{y}_i^* - \mathbf{x}_i' \boldsymbol{\beta}),$$

en la cual ρ_τ es la función de chequeo definida en (1.9). Este procedimiento del remuestreo sobre el residual para la estimación de $\hat{\boldsymbol{\beta}}_n^*(\tau)$ se repite B veces, generando una matriz de orden $B \times p$ en la cual $\hat{\boldsymbol{\beta}}_n^*(\tau)$ es un vector de los coeficientes de la regresión cuantílica de orden $p \times 1$. Cada columna en esta matriz se convierte en un estimador de la distribución muestral de $\hat{\boldsymbol{\beta}}_j(\tau)$, para $j = 1, \dots, p$ colocando una probabilidad de $1/B$ en cada valor de $\hat{\boldsymbol{\beta}}_n^*(\tau)$ para cada parámetro dado por $\boldsymbol{\beta}_j$. Por el teorema (A.1), el τ -ésimo cuantil muestral ξ_τ converge débilmente a

$$\mathcal{N}(0, \tau(1 - \tau)/f(0)^2)$$

y aproximando esta distribución, se pueden construir intervalos de confianza para $\boldsymbol{\beta}(\tau)$.

Assumiendo la condición del error i.i.d. cuando se aplica el residual del *bootstrap*, De Angelis, et al. (1993) mostraron que si $\mathbf{z} = (z_1, \dots, z_p)' \in \mathbb{R}$ el estimador *bootstrap* de la distribución G de $\hat{\boldsymbol{\beta}}$

$$G(\mathbf{z}) = P\{(\sqrt{n} (\hat{\boldsymbol{\beta}}_j(\tau) - \boldsymbol{\beta}_{0j}(\tau))) \leq z_j, \quad 1 \leq j \leq p\},$$

condicional sobre la muestra inicial

$$\chi = \{(\mathbf{x}_i, \mathbf{Y}_i) : i = 1, \dots, n\},$$

viene dado por

$$\hat{G}(\mathbf{z}) = P\{\sqrt{n} (\hat{\boldsymbol{\beta}}_n^*(\tau) - \hat{\boldsymbol{\beta}}_j(\tau)) \leq z_j, \quad \text{para } j = 1, \dots, p \mid \chi\},$$

y converge a la distribución limitante de

$$\sqrt{n} (\hat{\boldsymbol{\beta}}_n^*(\tau) - \boldsymbol{\beta}(\tau))$$

luego, el error de esta aproximación es de orden $O(n^{-1/4})$ cuando $n \rightarrow \infty$.

Teniendo $\hat{\beta}_1^*, \dots, \hat{\beta}_B^*$ y asumiendo que la distribución del error es i.i.d., bajo la condición de que no exista normalidad dado por el teorema (A.2), se puede estimar la matriz de varianza-covarianza asintótica $\omega^2 \mathbf{D}_0^{-1}$ de $\hat{\beta}$ en la cual

$$\omega^2 = \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}$$

y $\mathbf{D}_0^{-1} = \lim n^{-1} \sum x_i x_i'$.

De Angelis, et al. (1993) compararon la estrategia convencional de la estimación directa de la matriz de covarianza con una aproximación no paramétrica basada en el estimador Kernel de $f(F^{-1}(\tau))$, luego

$$\hat{f}(F^{-1}(\tau)) = (nh)^{-1} \sum_{i=1}^n K(\hat{u}_i/h),$$

en la cual el Kernel K es una función de densidad simétrica. Se observa que $f(F^{-1}(\tau))$ puede ser estimada de una forma no paramétrica para la precisión de $\mathcal{O}(n^{-2/5})$. Consecuentemente, el error de la aproximación $\mathcal{N}(0, \hat{\omega}^2 \mathbf{D}_0^{-1})$ también es de orden $\mathcal{O}(n^{-2/5})$. El comando utilizado en R para este método viene dado por `se = "boot"`.

3.7 Remuestreo del par (x,y).

El Remuestreo del par (x,y) provee una alternativa más efectiva para los modelos con errores independientes pero no idénticamente distribuidos. En vez de considerar muestras etiquetadas del *Bootstrap* para la distribución empírica de los residuales, se consideran pares $(\mathbf{x}_i, \mathbf{y}_i)$ de las muestras etiquetadas para la distribución conjunta de la muestra. El par $(\mathbf{x}_i^*, \mathbf{y}_i^*)$ se etiqueta con reemplazamiento de los n pares $\{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, \dots, n\}$ de la muestra original, cada uno con probabilidad $1/n$. Dada las realizaciones del *Bootstrap* $\{\hat{\beta}_b^* : b = 1, \dots, B\}$, existen muchas opciones que se pueden considerar para construir los intervalos de

confianza. De acuerdo a Kocherginsky, et al. (2005), cuando el conjunto de datos es grande el intervalo de confianza toma la forma

$$\hat{\beta}_\tau \pm \mathbf{Z}_{\alpha/2} SD(\hat{\beta}_\tau)$$

Este método de *bootstrap* se encuentra implementado en el *software* R con el comando `se = "boot", bsmethod = "xy"`.

3.8 Remuestreo sobre la Condición del Subgradiente.

Koenker (1994) citando a Parzen, et al. (1994) sugirió un mecanismo de remuestreo sobre el subgradiente. Para describir este método, suponga que \mathbf{Y}_i es la i -ésima variable dependiente, \mathbf{x}_i es el vector de covariables para $i = 1, \dots, n$ y $\tau \in (0, 1)$. Luego, considere la función de estimación $S_n(\mathbf{b})$ para el τ -ésimo cuantil definida por

$$S_n(\mathbf{b}) = -n^{-1/2} \sum_{i=1}^n \mathbf{x}_i(\tau - I(y_i \leq \mathbf{x}_i' \mathbf{b})), \quad (3.16)$$

se observa que $S_n(\mathbf{b})$ es un estadístico de pivote cuando $\mathbf{b} = \beta(\tau)$ en el modelo de regresión cuantílica, lo cual hace que su distribución sea independiente de $\beta(\tau)$. Una solución para $\hat{\beta}$ en la ecuación $S_n(\beta) = 0$ es un estimador puntual consistente de β_0 . En la práctica, $\hat{\beta}$ se obtiene mediante

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i' \beta)$$

La distribución de S_n puede ser generada por un vector aleatorio \mathbf{U} , el cual es una suma ponderada de variables aleatorias de Bernoulli

$$\mathbf{U}_n = -n^{-1/2} \sum_{i=1}^n \mathbf{x}_i(\tau - \xi_i)$$

donde $\{\xi_i\}$ es una muestra aleatoria de una variable aleatoria de Bernoulli, las cuales toman el valor de 1 para una probabilidad τ y el valor de cero para una probabilidad $1 - \tau$.

Para cualquier realización, \mathbf{u}_n de \mathbf{U}_n , se encuentra

$$\boldsymbol{\beta}_U^* = \{\mathbf{b} : S_n(\mathbf{b}) = \mathbf{u}_n\} \quad (3.17)$$

y se resuelve el problema de regresión cuantílica argumentado

$$\boldsymbol{\beta}_U^* = \min_{\mathbf{b} \in \mathbb{R}} \sum \rho_\tau(y_i - \mathbf{x}_i \mathbf{b}) + \rho_\tau(\zeta - \sqrt{n} \mathbf{u}'_n \mathbf{b} / \tau),$$

en el cual la variable ζ se elige lo suficientemente grande para asegurar que ésta exceda a $\sqrt{n} \mathbf{u}'_n \mathbf{b} / \tau$. Esto garantiza que el último término para el subgradiente sea

$$\sqrt{n} \mathbf{u}_n = \sum \mathbf{x}_i (\tau - \xi_i).$$

y así, se satisface $S_n(\mathbf{b}) = \mathbf{u}_n$

Parzen, et al. (1994) afirmaron que para probar que la distribución $\boldsymbol{\beta}_n^* - \hat{\boldsymbol{\beta}}_n$ tiene la misma distribución asintótica que $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$, debe mostrarse que, para

$$\|\mathbf{b} - \boldsymbol{\beta}_0\| \leq n^{-1/3} \quad \text{y} \quad \|\mathbf{b}^* - \boldsymbol{\beta}_0\| < n^{-1/3},$$

$$\text{el } \sup \|S_n(\mathbf{b}^*) - S_n(\mathbf{b}) - n^{-1/2} \sum \mathbf{x}_i [(F_i(\mathbf{x}'_i \mathbf{b}^* - \boldsymbol{\beta}_0)) - F_i(\mathbf{x}'_i (\mathbf{b} - \boldsymbol{\beta}_0))]\| \rightarrow 0,$$

en la cual F_i es la función de distribución de la función de densidad continua f_i y n es el tamaño muestral, luego, por la continuidad de f_i se tiene

$$\begin{aligned} n^{-1/2} \sum \mathbf{x}_i [(F_i(\mathbf{x}'_i (\mathbf{b}^* - \boldsymbol{\beta}_0)) - F_i(\mathbf{x}'_i (\mathbf{b} - \boldsymbol{\beta}_0)))] &= n^{-1/2} \sum f_i(0) \mathbf{x}_i \mathbf{x}'_i (\mathbf{b}^* - \mathbf{b}) \\ &+ n^{-1/2} \sum o(1) \mathbf{x}_i \mathbf{x}'_i (\mathbf{b}^* - \mathbf{b}). \end{aligned}$$

Ahora, haciendo $\mathbf{b}^* = \boldsymbol{\beta}_n^*(\tau)$ y $\mathbf{b} = \hat{\boldsymbol{\beta}}_n(\tau)$ y tomando en cuenta que

$$S_n(\boldsymbol{\beta}_n^*(\tau)) - S_n(\hat{\boldsymbol{\beta}}_n(\tau)) = \mathbf{U}_n,$$

se tiene que $\sqrt{n}(\boldsymbol{\beta}_n^*(\tau) - \hat{\boldsymbol{\beta}}_n(\tau)) \rightsquigarrow D_1^{-1} \mathbf{U}_n$ para $D_1 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i$.

Este método de *bootstrap* se encuentra implementado en el *software* R con el comando `se = "boot", bsmethod = "pwy"`.

3.9 Coeficiente de Determinación.

En los modelos de regresión lineal, la bondad del ajuste se mide por el coeficiente de determinación R^2 y su fórmula viene dada por:

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

R^2 se interpreta como la razón de variación entre la variable dependiente y las variables independientes del modelo. Además, se expresa como un porcentaje multiplicado por 100, esta cantidad varía entre 0 y 1. Cuando R^2 tiene un valor grande indica un mejor ajuste del modelo. Para una muestra y_1, \dots, y_n , el estadístico R^2 puede ser desarrollado en los modelos de regresión cuantílica, dado que estos modelos se basan en la minimización de una suma de distancias ponderadas

$$\sum_{i=1}^n d_{\tau}(y_i, \hat{y}_i),$$

con diferentes ponderaciones que dependen de si $y_i \geq \hat{y}_i$ o $y_i < \hat{y}_i$, en el cual el valor estimado se define por

$$\hat{y}_i = \hat{\beta}_0(\tau) + x_i \hat{\beta}_1(\tau).$$

Hao y Naiman (2007) citando a Koenker y Machado (1999) sugirieron medir la bondad del ajuste comparando la suma de las distancias ponderadas para el modelo de interés con la suma en la cual sólo la intercepción del parámetro aparece. Suponga que $\hat{V}(\tau)$ es la suma de las distancias ponderadas para todos los τ -ésimos

modelos de regresión cuantílica definida por:

$$\begin{aligned}\hat{V}(\tau) &= \sum_{i=1}^n d_{\tau}(y_i, \hat{y}_i) \\ &= \tau \sum_{y_i \geq \beta_0 + \beta_1 x_i} |y_i - \beta_0 - \beta_1 x_i| + (1 - \tau) \sum_{y_i < \beta_0 + \beta_1 x_i} |y_i - \beta_0 - \beta_1 x_i|\end{aligned}$$

y $\tilde{V}(\tau)$ es la suma de las distancias ponderada para el modelo que incluye sólo un término constante y está definida por:

$$\begin{aligned}\tilde{V}(\tau) &= \sum_{i=1}^n d_{\tau}(y_i, \hat{Q}(\tau)) \\ &= \tau \sum_{y_i \geq \bar{y}} |y_i - \hat{Q}(\tau)| + (1 - \tau) \sum_{y_i < \bar{y}} |y_i - \hat{Q}(\tau)|\end{aligned}$$

en la cual $\hat{Q}(\tau)$ es el τ -ésimo cuantil muestral para una muestra $\{y_1, \dots, y_n\}$ definido mediante

$$\hat{Q}(\tau) = \hat{F}^{-1}(\tau) = \inf \left\{ y \mid \min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \xi) \right\}.$$

Luego, la bondad del ajuste para el modelo de regresión cuantílica está dada por:

$$R^1(\tau) = 1 - \frac{\hat{V}(\tau)}{\tilde{V}(\tau)} \quad (3.18)$$

Puesto que $\hat{V}(\tau)$ y $\tilde{V}(\tau)$ son no negativas, $R^1(\tau)$ está dentro del rango del intervalo $[0, 1]$. Un valor grande para $R^1(\tau)$ indica un mejor ajuste del modelo. La ecuación (3.18) es una medida local de la bondad del ajuste del modelo de regresión cuantílica en τ .

El *software* Stata provee la medida del ajuste para el modelo de regresión cuantílica usando la ecuación (3.18) y se refiere a este coeficiente como pseudo R^2 para diferenciarlo del coeficiente de determinación para el modelo de regresión lineal.

CAPITULO 4

APLICACIÓN DE LA REGRESIÓN CUANTÍLICA

En este capítulo se presentarán cuatro ejemplos donde se estima, mediante el paquete *quantreg* del *software* R, rectas de regresión cuantílica para diversos valores de τ . El primero es un ejemplo tomado de Koenker(2005), diseñado para explorar algunas características básicas del paquete antes mencionado.

Los demás ejemplos se construyen con datos suministrados, por diversas agencias sismológicas, referentes a magnitud y latitud de los sismos ocurridos en la zona nororiental del país. A continuación se dan los ejemplos.

4.1 Ejemplo 1: Datos Engel (1857).

Considere el estudio de Engel (1857) sobre los gastos familiares de comida versus los ingresos anuales de las 235 clases trabajadoras Europeas. Los datos están disponibles en el *software* R.

1. Instalar los paquetes *quantreg*.

```
instalar.packages("quantreg")
```

2. Ordenar a R que dé acceso a los paquetes *quantreg*.

```
library(quantreg)
```

3. Leer los datos.

```
data(engel)
```

4. Hacer que las columnas del conjunto de datos sean visible como variables.

```
attach(engel)
```

5. Almacenar los datos en una variable cualquiera para el cuantil $\tau = 0,5$

```
fit.5 <- rq(foodexp~income, tau = 0.5, data = engel)
```

- Mostrar en pantalla los estimadores de los coeficientes, los grados de libertad y los residuales para el cuantil $\tau = 0,5$. Ver tabla (4.1)

```
fit.5
```

Tabla 4.1: Estimadores de la regresión cuantílica para las variables ingresos familiar y gastos de comida en el cuantil 0,5.

```
Estimadores:  
(Intercepción)  Ingresos  
                81,482    0,560
```

```
Grados de Libertad: 235 total; 233 residual
```

- Graficar un diagrama de dispersión para los datos Engel (1857). Ver figura (4.1).

```
plot(income, foodexp, xlab = "Ingreso Familiar", ylab = "Gastos  
de Comida")
```

- Trazar la recta de regresión lineal sobre el diagrama de dispersión.

```
abline(lm(foodexp ~ income), col = "red")
```

- Trazar la recta de regresión cuantílica con $\tau = 0,5$ sobre el diagrama de dispersión.

```
abline(rq(foodexp ~ income, tau = 0.5), col = "blue")
```

- Crear un vector para $\tau \in [0, 1]$.

```
taus <- c(0.05, 0.25, 0.5, 0.75, 0.9, 0.95)
```

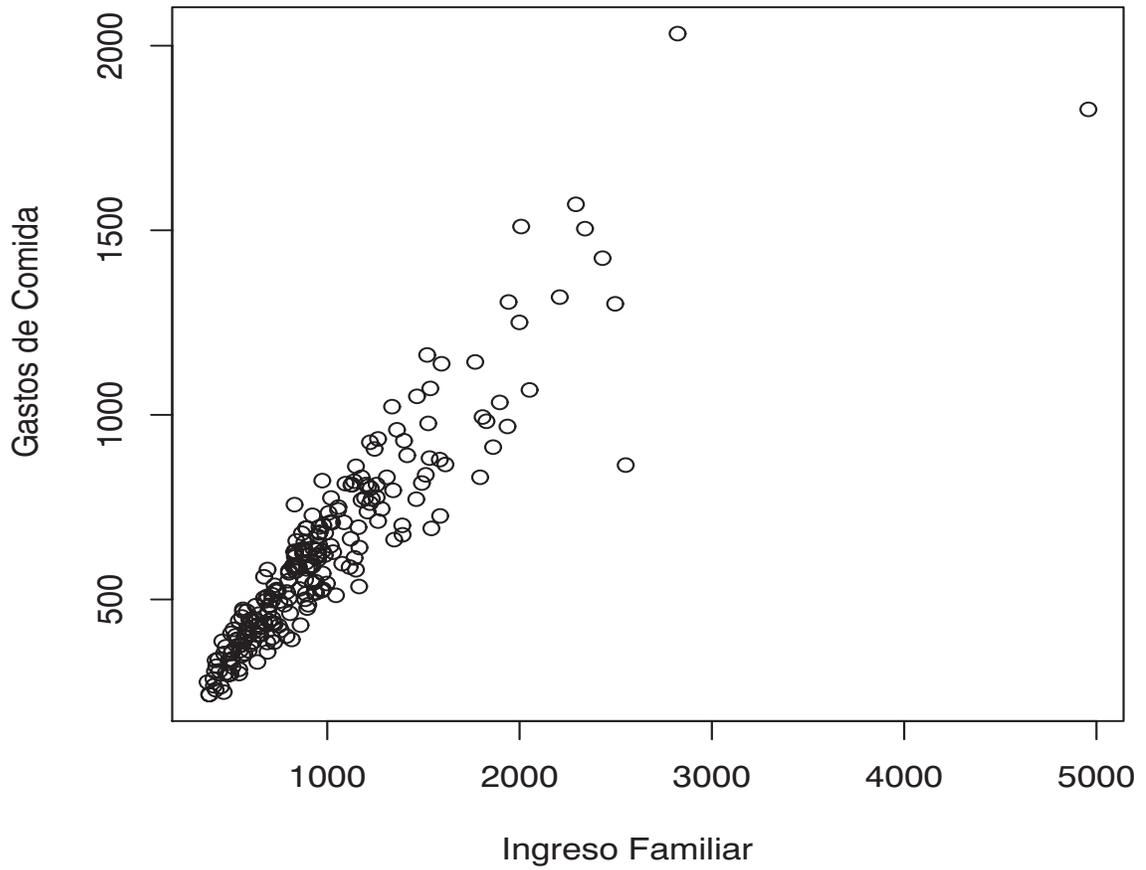


Figura 4.1: Datos Engel (1857). La figura muestra el diagrama de dispersión de los gastos de comida vs. el ingreso familiar para una muestras de 235 observaciones.

11. Para cada uno de los valores contenidos en el vector τ se traza una recta de regresión cuantílica. Ver figura (4.2).

```
for (i in 1:length(taus)){abline(rq(foodexp ~ income,
tau = taus[i]))}
```

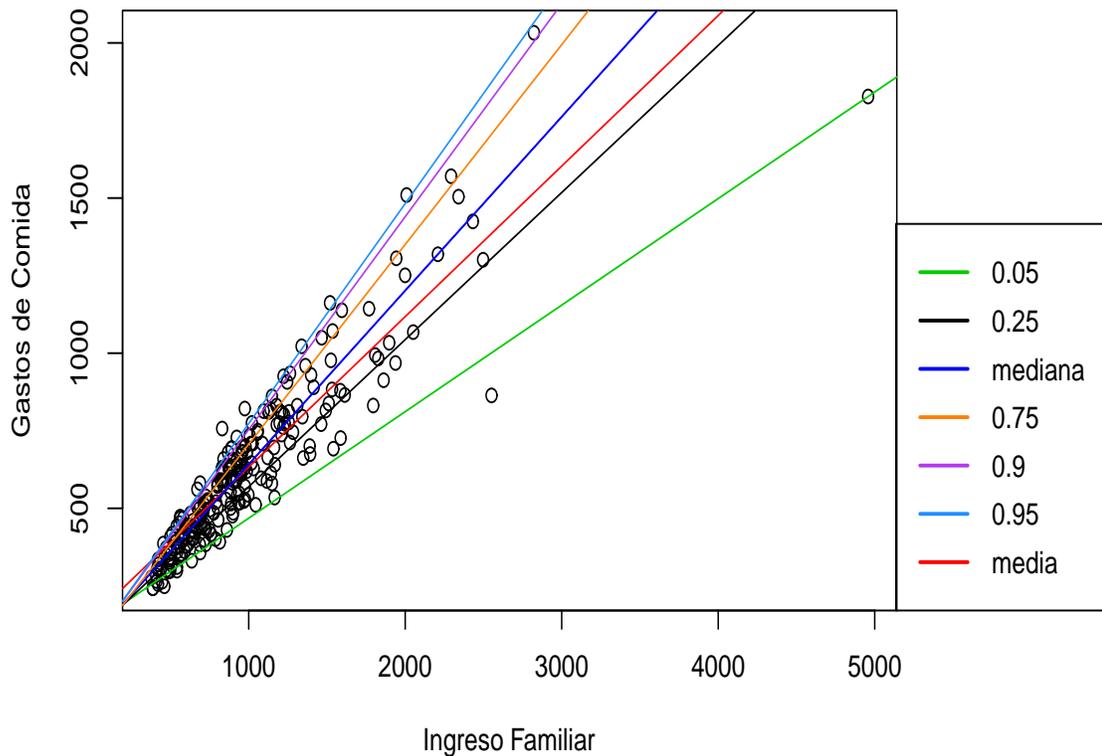


Figura 4.2: Datos Engel (1857). La figura muestra la recta de regresión lineal y las rectas de regresión cuantílicas de los gastos de comida vs. el ingreso familiar.

12. Asignar a una variable cualquiera a la función *Summary* para determinar los estimadores e intervalos de confianza de los coeficientes para cada cuantil usando el método por defecto del *software* R (Transformación Rango).

```
regtaus <-summary(rq(foodexp ~ income, tau = taus))
```

13. Mostrar en pantalla los estimadores, cotas superiores y cotas inferiores para cada cuantil. Ver tabla (4.2).

```
regtaus
```

Tabla 4.2: Estimadores e intervalos de confianza de la regresión cuantílica obtenidos usando el método Transformación Rango para las variables ingreso familiar y gastos de comida en los cuantiles 0,05 al 0,95.

tau	Variable	Coefficientes	Cota Inferior	Cota Superior
0.05	Intercep	124,880	98,302	130,516
	Ingresos	0,343	0,343	0,389
0.25	Intercep	95,483	73,786	120,098
	Ingresos	0,474	0,4203	0,494
0.5	Intercep	81,482	53,259	114,011
	Ingresos	0,5601	0,487	0,601
0.75	Intercep	62,396	32,744	107,313
	Ingresos	0,644	0,580	0,690
0.90	Intercep	67,350	37,118	103,173
	Ingresos	0,686	0,649	0,7422
0.95	Intercep	64,103	46,264	83,578
	Ingresos	0,709	0,673	0,734

14. Asignar a una variable cualquiera los estimadores de los parámetros para cada cuantil.

```
inftaus <-rq(foodexp ~ income, tau = taus, data = engel)
```

15. Mostrar en pantalla los estimadores de los coeficientes para cada cuantil. Ver tabla (4.3).

```
inftaus
```

```
Call:rq(formula = foodexp ~ income,tau = taus, data = engel)
```

Tabla 4.3: Resumen de los estimadores de la regresión cuantílica para las variables ingresos familiar y gastos de comida en cada uno de los cuantiles presentados anteriormente.

taus	0,05	0,25	0,5	0,75	0,9	0,95
Intercepción	124,880	95,483	81,482	62,396	67,350	64,103
Ingresos	0,343	0,474	0,560	0,644	0,686	0,709

Los modelos de regresión cuantílica para los ingresos y gastos de comida son:

$$Q(0,05 | Ingreso) = 124,88 + 0,343 Ingreso$$

$$Q(0,25 | Ingreso) = 95,48 + 0,47 Ingreso$$

$$Q(0,50 | Ingreso) = 81,48 + 0,56 Ingreso$$

$$Q(0,75 | Ingreso) = 62,39 + 0,64 Ingreso$$

$$Q(0,90 | Ingreso) = 67,35 + 0,68 Ingreso$$

$$Q(0,95 | Ingreso) = 64,10 + 0,709 Ingreso$$

16. Realizar y mostrar en pantalla inferencias de la regresión cuantílica por el método del escalar esparcido con el parámetro de suavizado de Hall y Sheather (1988). Ver tabla 4.4.

```
summary(inftaus, se = "iid")
```

Tabla 4.4: Inferencias de la regresión cuantílica por el método del escalar esparcido con el parámetro de suavizado de Hall y Sheather (1988) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	124,880	17,654	7,073	0,000
	Ingresos	0,343	0,015	21,602	0,000

Tabla 4.4: Inferencias de la regresión cuantílica por el método del escalar esparcido con el parámetro de suavizado de Hall y Sheather (1988) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95 (Continuación)

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.25	Intercep	95,483	15,861	6,019	0,000
	Ingresos	0,474	0,014	33,198	0,000
0.5	Intercep	81,482	13,239	6,154	0,000
	Ingresos	0,560	0,0119	46,997	0,000
0.75	Intercep	62,396	10,671	5,847	0,000
	Ingresos	0,6440	0,009	67,033	0,000
0.90	Intercep	67,3508	20,567	3,274	0,001
	Ingresos	0,686	0,018	37,062	0,000
0.95	Intercep	64,103	9,846	6,510	0,000
	Ingresos	0,709	0,008	79,981	0,000

17. Efectuar inferencias de la regresión cuantílica por el método Remuestreo del residual. Ver tabla (4.5).

```
summary(inftaus, se = "boot")
```

Tabla 4.5: Inferencias de la regresión cuantílica por el método Remuestreo del residual para las variables ingresos familiar y gastos de comida en los taus 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	124,880	27,149	4,59969	0,00001
	Ingresos	0,343	0,0413	8,297	0,000
0.25	Intercep	95,483	25,222	3,785	0,0002

Tabla 4.5: Inferencias de la regresión cuantílica por el método Remuestreo del residual para las variables ingresos familiar y gastos de comida en los taus 0,05 al 0,95.(Continuación)

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.25	Ingresos	0,474	0,034	13,882	0,000
0.5	Intercep	81,482	27,994	2,910	0,0039
	Ingresos	0,560	0,0359	15,578	0,000
0.75	Intercep	62,396	26,752	2,332	0,0205
	Ingresos	0,6440	0,033	19,171	0,000
0.90	Intercep	67,3508	22,102	3,047	0,002
	Ingresos	0,686	0,027	25,272	0,000
0.95	Intercep	64,103	23,761	2,697	0,007
	Ingresos	0,709	0,027	25,871	0,000

18. Hacer inferencias de la regresión cuantílica por el método Remuestreo sobre el par (x, y) . Ver tabla (4.6).

```
summary(inftaus, se = "boot",bsmethod ="xy")
```

Tabla 4.6: Inferencias de la regresión cuantílica por el método Remuestreo del par (x, y) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	124,880	23,045	5,418	0,000
	Ingresos	0,343	0,0353	9,704	0,000
0.25	Intercep	95,483	28,221	3,383	0,0008
	Ingresos	0,474	0,038	12,316	0,000
0.5	Intercep	81,482	29,506	2,761	0,006

Tabla 4.6: Inferencias de la regresión cuantílica por el método Remuestreo del par (x, y) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95. (Continuación)

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.5	Ingresos	0,560	0,038	14,672	0,000
0.75	Intercep	62,396	23,429	2,663	0,008
	Ingresos	0,6440	0,030	21,398	0,000
0.90	Intercep	67,3508	22,104	3,046	0,002
	Ingresos	0,686	0,027	24,648	0,000
0.95	Intercep	64,103	23,404	2,738	0,006
	Ingresos	0,709	0,028	24,499	0,000

19. Efectuar inferencias de la regresión cuantílica por el método Remuestreo de Parzen, et al. (1994). Ver tabla (4.7).

```
summary(inftaus, se = "boot", bsmethod = "pwy")
```

Tabla 4.7: Inferencias de la regresión cuantílica por el método Remuestreo de Parzen, et al. (1994) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	124,880	14,981	8,335	0,000
	Ingresos	0,343	0,023	14,912	0,000
0.25	Intercep	95,483	18,656	5,118	0,000
	Ingresos	0,474	0,026	17,942	0,000
0.5	Intercep	81,482	23,793	3,4246	0,0007
	Ingresos	0,560	0,0302	18,538	0,000
0.75	Intercep	62,396	27,301	2,285	0,023

Tabla 4.7: Inferencias de la regresión cuantílica por el método Remuestreo de Parzen, et al. (1994) para las variables ingresos familiar y gastos de comida en los cuantiles 0,05 al 0,95. (Continuación)

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.75	Ingresos	0,6440	0,033	19,2474	0,000
0.90	Intercep	67,3508	25,116	2,681	0,007
	Ingresos	0,686	0,030	22,680	0,000
0.95	Intercep	64,103	29,504	2,172	0,030
	Ingresos	0,709	0,033	21,110	0,000

Como el *software* R no tiene disponible alguna función que permita medir la bondad del ajuste para el modelo de la regresión cuantílica, se usará el *software* Stata para determinar el coeficiente de determinación R^2 . Ver tabla (4.8).

Tabla 4.8: Bondad del ajuste para las variables ingreso familiar y gasto familiar en la regresión cuantílica.

	$\tau = 0,05$	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	$\tau = 0,90$	$\tau = 0,95$
<i>Pseudo R</i> ²	0,474	0,554	0,620	0,696	0,764	0,804

En la figura (4.2) se muestran las seis rectas de regresión cuantílica, correspondientes a los cuantiles 0,05; 0,25; 0,5; 0,75; 0,9; 0,95 y la recta de regresión para la media condicional.

El hecho de que la pendiente de la regresión cuantílica en el cuantil 0,05 sea menos inclinada que la pendiente de la recta de la regresión cuantílica en el cuantil 0,95, indica que las rectas convergen cuando se tienen ingresos inferiores, es decir, hay menor dispersión en la distribución de los gastos de comida cuando se tienen ingresos inferiores. Además, la separación en las rectas de regresión indica que la

distribución de los gastos de comida es sesgada hacia la izquierda.

Las rectas de regresión de la mediana y la media condicional son muy diferentes en este ejemplo, debido a la asimetría de la densidad condicional y a que los estimadores de mínimos cuadrados están fuertemente afectados por dos observaciones extremas una con gastos de comida muy bajos y otra con ingresos familiares altos. Se observa que una consecuencia de que no exista robustez es que el ajuste de mínimos cuadrados provee un estimador ligeramente deficiente de la media condicional para los ingresos más bajos en la muestra. Además, la recta de regresión por mínimos cuadrados pasa por encima de todos los ingresos muy bajos de las observaciones.

Se observa que los estimadores para la variable gastos de comida (intercepción) van disminuyendo de 124,88 en el cuantil $\tau = 0,05$ a 64,103 en el cuantil $\tau = 0,95$. Además, los estimadores para la variable ingreso familiar (pendiente) van aumentando de 0,34 en el cuantil $\tau = 0,05$ a 0,70 en el cuantil $\tau = 0,95$.

Los errores estándares por los métodos *bootstrap* para las variables gastos de comida e ingreso familiar, en cada cuantil, son mayores que los errores estándares del método del escalar esparcido, esto se debe a que la distribución de la variable dependiente es sesgada hacia la izquierda, lo cual produce heteroscedásticidad y el método del escalar esparcido resulta ser ineficaz debido a que este asume homoscedásticidad en la distribución de la variable dependiente. Los valores de R^2 indican que las rectas que explican mejor la variación son los $\tau = 0,5$; 0,75; 0,90 y 0,95.

4.2 Estimación de los Datos Emitidos por las Agencias Sismológicas.

Uno de los estados de mayor potencial sísmico, es el estado Sucre, representado por la zona 7, donde se recibirán las mayores aceleraciones del suelo, 0,40g. El estado Sucre forma parte del borde de interacción de las placas del Caribe y

Sur America, generando el sistema de fallas de El Pilar, entre Cabo Costera y Trinidad.

La mejor manera de cuantificar el tamaño de un sismo es determinar su momento sísmico. Las amplitudes de los movimientos de las ondas sísmicas son directamente proporcional a los momentos sísmicos. Las escalas basadas en las medidas de las amplitudes sísmicas se conocen como magnitudes sísmicas y fueron desarrolladas por Wadati (1933) y Richter (1935).

A continuación se utilizarán las técnicas de regresión cuantílica para determinar las diferencias existentes a lo largo de toda la distribución, entre las magnitudes de las ondas sísmicas (m_b), comprendidas entre 0° y 10° en la escala de Richter , las latitudes sísmicas (LAT), comprendidas entre 8° y 12° y las longitudes sísmicas (LONG), comprendidas entre -66° y -60° , suministradas por el *International Seismological Centre* (ISC) y las demás agencias sismológicas desde el año 1964 hasta el año 2011. Entre éstas se tienen: *International Seismological Centre*(ISCJB), *University of the West Indies* (TRN), *United States Coast and Geodetic Survey* (USCGS) y *Stiftelsen NOR SAR* (NAO).

4.3 Ejemplo 2: Magnitudes Sísmicas de ISC vs. USCGS.

Se trata de una muestra de 39 sismos ocurridos cerca de la costa de Venezuela entre -66° y -60° de longitud, 8° y 12° de latitud. Para cada sismo se tomo la magnitud dada por ISC y USCGS.

Los estimadores de los coeficientes, las cotas superiores y las cotas inferiores de las magnitudes de las ondas sísmicas de ISC y USCGS para cada cuantil se muestran en la tabla (4.9).

Tabla 4.9: Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las magnitudes de las ondas sísmicas de ISC y USCGS en los cuantiles 0,05 al 0,95.

tau	Variable	Coeficientes	Cota Inferior	Cota Superior
0.05	Intercep	0,720	-0,551	6,154
	ISC	0,800	0,185	0,994
0.25	Intercep	0,000	-0,160	0,792
	ISC	1,000	0,762	1,006
0.5	Intercep	0,000	0,000	0,000
	ISC	1,000	1,000	1,000
0.75	Intercep	0,100	0,100	0,339
	ISC	1,000	0,945	1,000
0.90	Intercep	0,775	-1,930	0,813
	ISC	0,875	0,867	1,056
0.95	Intercep	0,808	0,746	0,808
	ISC	0,869	0,869	0,885

Los modelos de regresión cuantílica para las magnitudes de las ondas sísmicas de ISC y USCGS son:

$$Q(0,05 | ISC) = 0,720 + 0,800 ISC$$

$$Q(0,25 | ISC) = 1,000 ISC$$

$$Q(0,50 | ISC) = 1,000 ISC$$

$$Q(0,75 | ISC) = 0,100 + 1,000 ISC$$

$$Q(0,90 | ISC) = 0,775 + 0,875 ISC$$

$$Q(0,95 | ISC) = 0,808 + 0,869 ISC$$

Los grados de libertad y los residuales para las magnitudes de las ondas sísmicas de ISC y USCGS son:

Grados de Libertad: 39 total; 37 residual

Efectuar inferencias sobre los estimadores de la regresión cuantílica por el método Remuestreo del residual para las magnitudes de las ondas sísmicas de USCGS y ISC. Ver tabla (4.10)

Tabla 4.10: Inferencias de la regresión cuantílica obtenidas usando el método Remuestreo del residual para las magnitudes de las ondas sísmicas de ISC y USCGS en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	0,720	1,318	0,545	0,588
	ISC	0,800	0,311	2,567	0,014
0.25	Intercep	0,000	0,280	0,000	1,000
	ISC	1,000	0,063	15,641	0,000
0.5	Intercep	0,000	0,113	0,000	1,000
	ISC	1,000	0,024	40,415	0,000
0.75	Intercep	0,100	0,255	0,391	0,697
	ISC	1,000	0,053	18,640	0,000
0.90	Intercep	0,775	0,394	1,965	0,056
	ISC	0,875	0,078	11,151	0,000
0.95	Intercep	0,808	0,338	2,391	0,021
	ISC	0,869	0,070	12,303	0,000

Tabla 4.11: Bondad del ajuste para las magnitudes de las ondas sísmicas de ISC y USCGS en la regresión cuantílica.

	$\tau = 0,05$	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	$\tau = 0,90$	$\tau = 0,95$
pseudo R ²	0,433	0,666	0,766	0,780	0,808	0,838

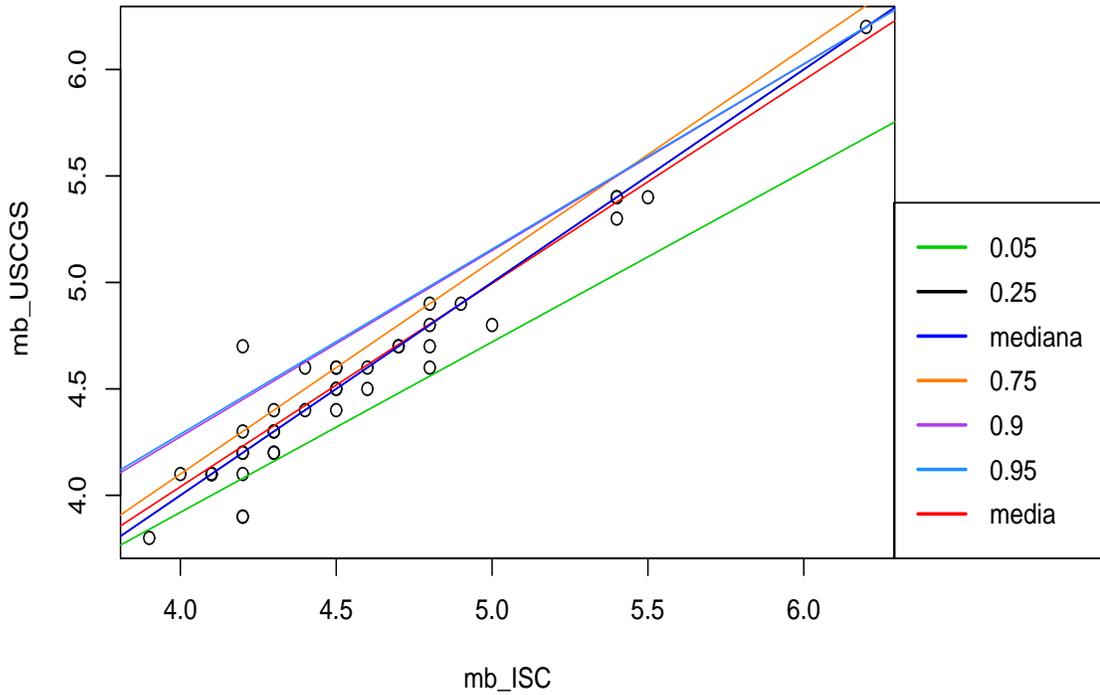


Figura 4.3: Rectas de Regresión Cuantílica y Regresión Lineal de las magnitudes de las ondas sísmicas (mb) de USC GS y ISC.

En la figura (4.3) se observa que el diagrama de dispersión sigue una regresión lineal entre las magnitudes sísmicas de ISC y USC GS. En la tabla (4.9) se muestran los coeficientes de las diferentes rectas de regresión cuantílica para los $\tau = 0,05$ al $\tau = 0,95$; junto con sus respectivos intervalos de confianza del 95% para cada uno de los coeficientes. Comparando diversas rectas puede verse que la de los cuantiles centrales tienen la misma pendiente; es decir, sus rectas son paralelas lo cual indica que las varianzas para ese conjunto de datos son iguales.

Según los valores de la tabla (4.10), hay una relación significativa entre mb de ISC y mb de USC GS, pues todas las probabilidades asociadas a las pendientes son menores que 0,05. Esto también se concluye observando los valores de R^2 en la tabla (4.11) para los cuantiles 0,25; 0,5; 0,75; 0,90 y 0,95. En la figura (4.3)

se observa que las rectas de regresión correspondiente a la media y a la mediana prácticamente se confunden; lo cual significa que la distribución es simétrica, en consecuencia en este caso es preferible usar la regresión lineal clásica en vez de la cuantílica.

4.4 Ejemplo 3: Magnitudes Sísmicas de ISC vs. NAO.

Para este ejemplo se seleccionó una muestra de 65 sismos ocurridos en la misma región, pero ahora se tiene la magnitud de cada sismo dada por ISC y NAO.

Los estimadores de los coeficientes, las cotas superiores y las cotas inferiores para las magnitudes de las ondas sísmicas de ISC y NAO para cada cuantil se muestran en la tabla (4.12)

Los modelos de regresión cuantílica para las magnitudes de las ondas sísmicas de ISC y NAO son:

$$Q(0,05 | ISC) = 2,271 + 0,428 ISC$$

$$Q(0,25 | ISC) = 0,800 + 0,800 ISC$$

$$Q(0,50 | ISC) = 0,533 + 0,888 ISC$$

$$Q(0,75 | ISC) = 0,200 + 1,000 ISC$$

$$Q(0,90 | ISC) = -0,136 + 1,090 ISC$$

$$Q(0,95 | ISC) = -0,285 + 1,142 ISC$$

Los grados de libertad y los residuales para las magnitudes de las ondas sísmicas de ISC y NAO son:

Grados de Libertad: 65 total; 63 residual

Tabla 4.12: Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95.

tau	Variable	Coeficientes	Cota Inferior	Cota Superior
0.05	Intercep	2,271	0,314	2,653
	ISC	0,428	0,341	0,846
0.25	Intercep	0,800	-0,176	1,817
	ISC	0,800	0,515	0,861
0.5	Intercep	0,533	-0,505	0,939
	ISC	0,888	0,786	1,016
0.75	Intercep	0,200	0,195	0,688
	ISC	1,000	0,868	1,020
0.90	Intercep	-0,136	-0,136	0,376
	ISC	1,090	0,965	1,090
0.95	Intercep	-0,285	-0,404	0,322
	ISC	1,142	1,020	1,183

Efectuar inferencias sobre los estimadores de la regresión cuantílica por el método de Remuestreo del residual para las magnitudes de las ondas sísmicas de ISC y NAO.

Tabla 4.13: Inferencias de la regresión cuantílica obtenidas usando el método *Bootstrap* para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	2,271	0,765	2,965	0,004
	ISC	0,428	0,173	2,467	0,016
0.25	Intercep	0,800	0,535	1,493	0,140

Tabla 4.13: Inferencias de la regresión cuantílica obtenidas usando el método Remuestreo del residual para las magnitudes de las ondas sísmicas de ISC y NAO en los cuantiles 0,05 al 0,95. (Continuación)

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.25	ISC	0,800	0,125	6,366	0,000
0.5	Intercep	0,533	0,311	1,712	0,091
	ISC	0,888	0,069	12,860	0,000
0.75	Intercep	0,200	0,300	0,665	0,508
	ISC	1,000	0,072	13,275	0,000
0.90	Intercep	-0,136	0,234	-0,582	0,562
	ISC	1,090	0,055	19,816	0,000
0.95	Intercep	-0,285	0,220	-1,296	0,199
	ISC	1,142	0,053	21,512	0,000

Tabla 4.14: Bondad del ajuste para las magnitudes de las ondas sísmicas de ISC y NAO en la regresión cuantílica.

	$\tau = 0,05$	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	$\tau = 0,90$	$\tau = 0,95$
pseudo R^2	0,420	0,398	0,505	0,581	0,624	0,664

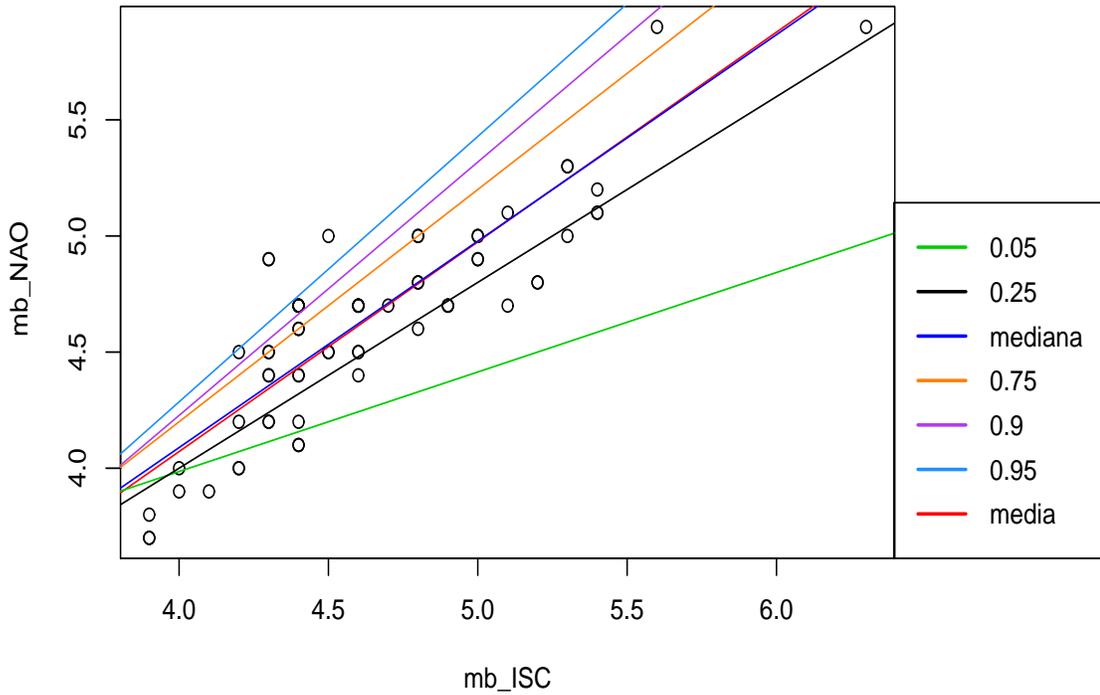


Figura 4.4: Rectas de Regresión Cuantílica y Regresión Lineal de las magnitudes de las ondas sísmicas (mb) de NAO y ISC.

En la tabla (4.12), se observa que las pendientes de las rectas aumentan a medida que aumentan los valores de τ , los cuales indican que hay heteroscedásticidad. En la figura (4.4) puede apreciarse como los ajustes mediante la regresión cuantílica expresan de forma adecuada la relación existente entre los datos de ISC y NAO para los diferentes cuantiles, pues en la tabla (4.13) se observa que existe una relación significativa entre estas dos variables. Además, el hecho de que la pendiente de la recta en el cuantil 0,05 sea menos inclinada que la pendiente de la recta en el cuantil 0,95 indica que las rectas convergen cuando se tienen valores pequeños de mb (ISC), es decir, hay menor dispersión en las magnitudes de NAO para valores pequeños de las magnitudes de ISC y la separación de las rectas de regresión en la parte superior indica que la distribución de mb dado por NAO es

sesgada hacia la izquierda.

4.5 Ejemplo 4: Latitudes Sísmicas de ISCJB vs. TRN.

Se tomó una muestra de 410 datos de las latitudes de los sismos ocurridas en la región antes descrita, suministrada por las agencias ISCJB y TRN .

Los estimadores de los coeficientes, las cotas superiores y las cotas inferiores para las latitudes sísmicas de ISCJB y TRN en cada cuantil se muestran en la tabla (4.15)

Tabla 4.15: Estimadores e intervalos de confianza de la regresión cuantílica obtenidos por el método Transformación Rango para las latitudes sísmicas de ISCJB y TRN en los cuantiles 0,05 al 0,95.

tau	Variable	Coeficientes	Cota Inferior	Cota Superior
0.05	Intercep	-1,266	-2,432	-0,559
	ISCJB	1,111	1,050	1,218
0.25	Intercep	-0,076	-0,397	0,222
	ISCJB	1,008	0,980	1,038
0.5	Intercep	0,289	-0,080	0,913
	ISCJB	0,978	0,919	1,012
0.75	Intercep	1,144	0,458	1,648
	ISCJB	0,904	0,857	0,968
0.90	Intercep	2,437	0,031	5,152
	ISCJB	0,793	0,539	0,977
0.95	Intercep	5,364	1,325	7,762
	ISCJB	0,528	0,307	0,812

Los modelos de regresión cuantílica para las latitudes sísmicas de ISCJB Y TRN son:

$$Q(0,05|ISCJB) = -1,266 + 1,111 ISCJB$$

$$Q(0,25|ISCJB) = -0,076 + 1,008 ISCJB$$

$$Q(0,50|ISCJB) = -0,032 + 1,002 ISCJB$$

$$Q(0,75|ISCJB) = 0,491 + 0,956 ISCJB$$

$$Q(0,90|ISCJB) = 1,134 + 0,898 ISCJB$$

$$Q(0,95|ISCJB) = 1,328 + 0,881 ISCJB$$

Los grados de libertad y los residuales para las latitudes sísmicas de ISCJB y TRN son:

Grados de Libertad: 410 total; 408 residual

Efectuar inferencias sobre los estimadores de la regresión cuantílica por el método de Remuestreo del residual para las latitudes sísmicas de ISCJB y TRN.

Tabla 4.16: Inferencias de la regresión cuantílica obtenidas usando el método Remuestreo del residual para las latitudes sísmicas de ISCJB y TRN en los cuantiles 0,05 al 0,95.

tau	Variable	Value	Error Estándar	valor t	Pr(> t)
0.05	Intercep	-1,266	0,700	-1,807	0,071
	ISCJB	1,111	0,064	17,229	0,000
0.25	Intercep	-0,076	0,260	-0,293	0,769
	ISCJB	1,008	0,024	41,555	0,000
0.5	Intercep	0,289	0,258	1,118	0,263
	ISCJB	0,978	0,024	40,502	0,000
0.75	Intercep	1,144	0,445	2,570	0,010
	ISCJB	0,904	0,041	21,785	0,000
0.90	Intercep	2,437	1,394	1,747	0,081
	ISCJB	0,793	0,129	6,130	0,000
0.95	Intercep	1,328	0,733	1,810	0,073
	ISCJB	0,881	0,068	12,965	0,000

Tabla 4.17: Bondad del ajuste para las latitudes sísmicas de ISCJB y TRN en la regresión cuantílica.

	$\tau = 0,05$	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	$\tau = 0,90$	$\tau = 0,95$
pseudo R^2	0,675	0,671	0,581	0,411	0,201	0,105

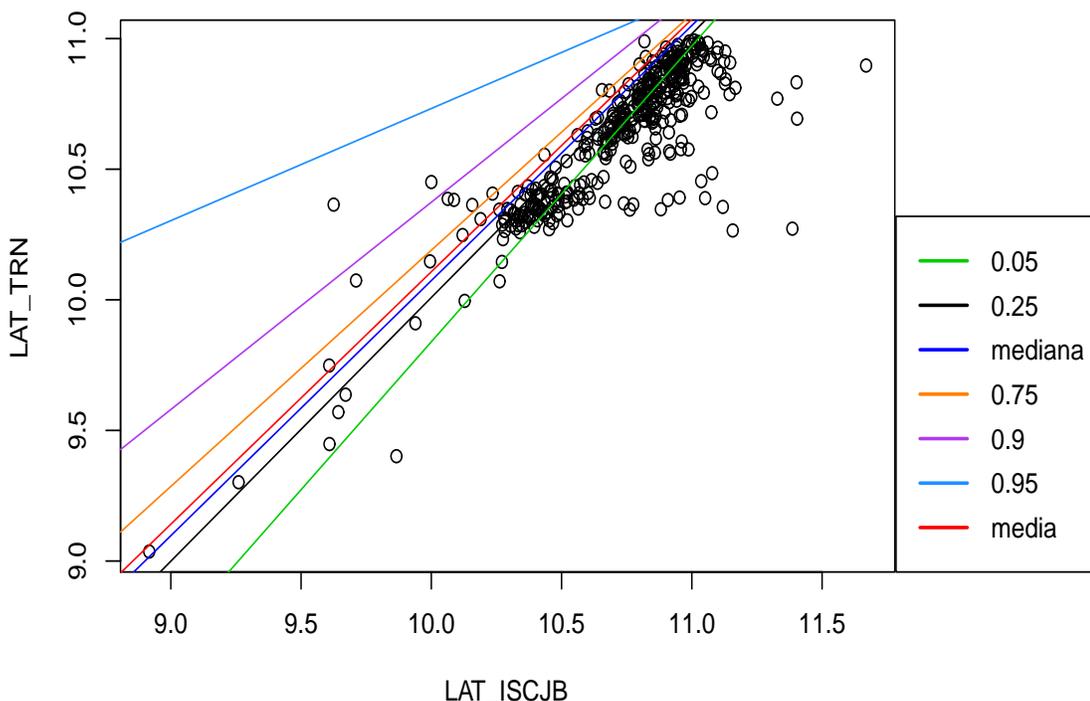


Figura 4.5: Rectas de Regresión Cuantílica y Regresión Lineal de las latitudes sísmicas de ISCJB y TRN

En la tabla (4.15) se observan que las pendientes de las rectas disminuyen a medida que aumentan los valores de τ , lo cual indica que hay heteroscedasticidad en los datos. Además, indica que las rectas convergen cuando se tienen valores altos de ISCJB señalando que hay mayor dispersión para este tipo de valores. La separación de las rectas en la parte inferior señala que la distribución de las latitudes dadas por TRN es sesgada hacia la derecha. De la tabla (4.16) se concluye que para los cuantiles utilizados existe una relación significativa entre la latitudes dadas por ISCJB Y TRN . Los valores de R^2 indican que las rectas que explican

mejor la variación son los $\tau = 0,05; 0,25$ y $0,5$.

CONCLUSIONES

Se concluyo que la regresión cuantílica permitió estudiar el efecto de la variable independiente sobre la variable de respuesta, para diferentes cuantiles; sin tomar en consideración la distribución de los residuales. La regresión cuantílica se utilizo para determinar la relación entre las magnitudes sísmicas dadas y la relación entre las latitudes para diversas agencias, encontrándose que en la mayoría de los casos es conveniente usar este tipo de metodología para explicar la relación existente entre las medidas dadas por las agencias.

BIBLIOGRAFÍA

- [1] Albright, B. (2010). *Mathematical Modeling with Excel*. Sudbury MA: Jones and Bartlett Publishers.
- [2] Andrews, D.W. K., & Buchinsky, M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, **68**(1), 23-51.
- [3] Bassett, Jr., G. & Koenker, R. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*, **77**(378), 407-415.
- [4] Barrodale, I. & Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the l_1 norm. *Communications ACM*, **17**, 319-320.
- [5] Bazarra, M. S., Jarvis, J.J. & Sherali, H. D. (1999). *Programación Lineal y Flujo en Redes*. México: Limusa.
- [6] Becker, R. A., Chambers, J.M. & Wilks, A. R. (1988). *The New S Language*. Wadsworth: Pacific Grove. CA.
- [7] Berman, A. (1973). *Cones, Matrices and Mathematical Programming*. New York: Springer.
- [8] Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian J of Statistics*, **17**(1), 1-7.
- [9] Boscovich, R. J. (1760). De recentissimis giadiiim dimensionibus. et iigura, ac magnitudine terrae inde derivanda. *Philosophiae Receniioris*, **2**, 406-426.
- [10] Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources*, **33**(1), 88-126.

- [11] Buhai, I. S. (2004). *Quantile Regression: Overview and Selected Applications*.
<http://www.ad-astra.ro/journal/7/buhai.pdf>
- [12] Burden, R. L. & Faires, D. J. (1985). *Análisis Numérico*. Grupo Editorial Iberoamérica.
- [13] Cameron, A. C. & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- [14] Dantzing, G. B. (1947). Programming in a linear structure, Comptroller, *Fuerza Aérea de los Estados Unidos*, Washington, D. C.
- [15] Draper, N. R. & Smith, H. (1998). *Applied Regression Analysis*. New York: John Wiley and Sons.
- [16] De Angelis, D., Hall, P. & Young, G. A. (1993). Analytical and Bootstrap Approximations to Estimator Distributions in L_1 Regression. *Journal of the American Statistical Association*, **88**(424), 1310-1316.
- [17] Edgeworth, F. Y. (1888). On a new method of reducing observations relating to several quantities, *Philosophical Magazine*, **25**, 184-191.
- [18] Edgeworth, F. Y. (1883). The method of least square. *Philosophical Magazine*, **16**, 360-75.
- [19] Efron, B. (1979). Bootstrap method: Another look at jackknife. *Annals of Statistic*, **7**(1), 1-26.
- [20] Engel, E. (1857). Die Production-und Consumptions Verhältnisse des Königreichs Sachsen. Reprinted in Die Lebenskosten Belgischer Arbeiter- Familien Fruher und Jetzt. *International Statistical Institute Bulletin*, **9**(1), 1-125.

- [21] Fowler, A. C. (1997). *Mathematical Models in the Applied Sciences*. 3rd edition. Cambridge: Cambridge University Press.
- [22] Gill, P. E., Murray, W. & Wright, M. H. (1991). *Numerical Linear Algebra and Optimization - Volume one*. Addison Wesley.
- [23] Girdano, F. R. (2003). *First Course in Mathematical Modeling*. Brooks Cole Publishing.
- [24] Glineur, F. (2001). Topics Convex Optimization: Interior Point Methods, Conic Duality and Approximations. *Politechmons*, **1**, 143-156.
- [25] Gourieroux, C. & Monfort, A. (1995). *Statistics and Econometric Models - Volume one*. Cambridge: Cambridge University Press.
- [26] Hájek, J. & Šidák, Z. (1967). *Theory of Rank Test*. Academia, Prague.
- [27] Hall, P. & Shearther, S. (1988). On the distribution of a Studentized quantile. *JRSS-B*, **50**(3), 381-391.
- [28] Hao, L. & Naiman, D. Q. (2007). *Quantile Regression*. Sage Publications.
- [29] Hendricks, W. & Koenker, R. (1991). Hierarchical Spline Models for Conditional Quantiles and Demand for Electricity. *Journal of the American Statistical Association*, **87**, 58-68.
- [30] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, **35**, 73-101.
- [31] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium*, **1**, 221-233.

- [32] International Seismological Centre (01/05/2012). *ISC Bulletin: arrivals search*.
<http://isc.ac.uk/iscbulletin/search/arrivals/>
- [33] International Statistical Institute (02/04/2012). *The ISI Glossary of Statistical Terms*.
<http://isi.cbs.nbl/glossary/>
- [34] John, O. O. & Nduka E. C. (2009). Quantile regression analysis as robust alternative to ordinary least square. *Scientia Africana*, **8**(2), 61-65.
- [35] Kecojević, T. (2011). Bootstrap Inference for Parametric Quantile Regression. *Univerity of Manchester*, 1-313.
- [36] Kocherginsky, M., He, X. & Mu, Y. (2005). Practical confidence intervals for regression quantile. *Journal of Computational and Graphical Statistics*, **14**(1), 41-55.
- [37] Koenker, R. (1994). Confidence intervals for regression quantiles. Pages 349-359 in Mandl, P. & Huškova, M., editores. *Asymptotics Statistics: Proceedings of the Fifth Prague Symposium*, Heidelberg: Physica-Verlag.
- [38] Koenker, R. (2000). Galton, Edgeworth, Frisch, and prospects for quantiles regression in econometrics. *Journal of Econometrics*, **95**(1), 347-374.
- [39] Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.
- [40] Koenker, R. (2012). *Package quantreg*.
<http://cran.r-project.org/web/packages/quantreg/quantreg.pdf>

- [41] Koenker, R. & Bassett, Jr., G. (1978). Regression quantiles. *Econometrica*, **46**(1), 33-50.
- [42] Koenker, R. & Basset, Jr.,G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50**(1), 43-62.
- [43] Koenker, R. W. & d'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36**(3), 383-393.
- [44] Koenker, R. & D'Orey, V. (1994). A Remark on Computing Regression Quantile. *Applied Statistics*, **43**(2), 410-414.
- [45] Koenker, R. W. & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, **15**(4), 143-156.
- [46] Koenker, R. & Machado, J. A. F. (1999). Goodness of Fit and Related Inference Process for Quantile Regression. *Journal of The American Statistical Association*, **94**(448), 1296-1310.
- [47] Koenker, R. W. & Portnoy, S. (21/10/2011). *Quantile Regression*.
<http://ysidro.econ.uiuc.edu/~roger/courses/472/1998/rq.pdf>
- [48] Lee, S. (2005). Lecture Note for MECT1 Quantile Regression.
- [49] Montezuma, D. E. (2012). Determinación de Areas de Riesgo Sísmico, Estado Sucre. Venezuela. *Revista Terra*, **26**(42), 13-45.
- [50] Parzen, E. (1979). Nonparametric Statistical Data modeling. *Statistical Science*, **74**(365), 105-121.
- [51] Parzen, M., Wei, L. & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**(2), 314-350.

- [52] Portnoy, S. & Koenker, R. (1997). The Gaussian Hare and Laplacian Tortoise: Computation of Squared Error vs. Absolute Estimators (with discussion), *Statistical Science*, **12**(4), 279-300.
- [53] Powell, J. L. (1991). *Estimation of the Monotonic Regression Models under Quantile Restrictions*. Cambridge: Cambridge University Press.
- [54] Remenyi, D., Onefrei, G. & English, J. (1998). *An Introduction to Statistics using Microsoft Excel*. Academic Publishing Ltd.
- [55] Richter, C. F. (1935). An instrumental earthquake magnitude scale. *Bull. Seismol. Soc. Am*, **25**(1), 1-32.
- [56] Rousseuw, P. & Leroy, A. (1987). *Robust Regression and Outlier Detection*. Jhon Wiley & Sons: New York.
- [57] Sheather, S. J. & Mariz, J. S. (1983). An Estimate of the Asymptotic Standard Error of the Sample Median. *Australian Journal of Statistics*, **25**(1), 109-122.
- [58] Siddiqui, M. M. (1960). Distribution of Quantiles from a Bivariate Population. *Journal of Research of the National Bureau of Standards*, **64**(B), 145-150.
- [59] Surendra, S., Ramoni, J., Torre, E. & Orlandoni, G. (2011). Modelado del Índice de Capacidad Profesional de profesores universitarios por regresión cuantil: El caso de la Universidad de Los Andes. 1-17.
- [60] Turkey, J. (1965). Which part of the Sample contains the information?. *Proceedings of the National Academy of Science of the United States of America*, **53**(1), 127-134.
- [61] Wadati, K. (1933). On the travel time of earthquake waves. *Geophys. Mg*, **7**(2), 101-111.

- [62] Warren, G. G. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall.

APÉNDICE A

Teorema A.1. *Suponga que $\{\hat{\xi}_{n\tau_1}, \dots, \hat{\xi}_{n\tau_m}\}$ representa una secuencia de cuantiles muestrales con $0 < \tau_1 < \tau_2 < \dots < \tau_m < 1$ y están basados en una muestra aleatoria de una población de tamaño n con una función de distribución inversa $\xi_\tau = F^{-1}(\tau)$. Si F es continua y tiene densidad positiva y continua f , en ξ_{τ_i} , para $i = 1, \dots, m$ entonces,*

$$\sqrt{n} [\hat{\xi}_{n\tau_1} - \xi_{\tau_1}, \dots, \hat{\xi}_{n\tau_m} - \xi_{\tau_m}]$$

converge en distribución a una m variable del vector aleatorio gaussiano con media igual a 0, y matriz de covarianza

$$\Omega(\tau_1, \dots, \tau_m; F)$$

con

$$\omega_{ij} = \frac{\tau_i(1 - \tau_j)}{f(\xi_{\tau_i})f(\xi_{\tau_j})}, \quad i \leq j$$

Demostración. Ver Koenker y Bassett (1978). □

Teorema A.2. *Suponga que $\{\hat{\beta}_n(\tau_1), \hat{\beta}_n(\tau_2), \dots, \hat{\beta}_n(\tau_m)\}$ representa una secuencia de cuantiles muestrales únicos con $0 < \tau_1 < \tau_2 < \dots < \tau_m < 1$ y están basados en una muestra aleatoria de una población de tamaño n con función de distribución inversa $\xi_\tau = F^{-1}(\tau)$, $\boldsymbol{\xi}_\tau = (\xi_\tau, 0, \dots, 0) \in \mathbb{R}^k$ y $\hat{\xi}_{n\tau} = \hat{\beta}_n(\tau) - \beta$. Se asume:*

1. *F es una función continua y tiene una densidad positiva y continua f , en ξ_{τ_j} , $j = 1, \dots, m$ y*
2. *$x_{ij} = 1$ para $i = 1, \dots, n$ y $\lim_{n \rightarrow \infty} n^{-1} \mathbf{X}'\mathbf{X} = Q_0$, una matriz definida positiva.*

Entonces,

$$\sqrt{n} [\hat{\xi}_{n\tau_1} - \xi_{\tau_1}, \dots, \hat{\xi}_{n\tau_m} - \xi_{\tau_m}]$$

converge en distribución a una mk variable del vector aleatorio gaussiano con media igual a 0 y matriz de covarianza

$$\Omega(\tau_1, \dots, \tau_m; F) \otimes Q_0^{-1}$$

donde Ω es la matriz de covarianza de los m cuantiles muestrales obtenidos de una muestra aleatoria de tamaño n de la distribución F .

Demostración. Ver Koenker y Bassett(1978). □

Nota: 1. y 2. se conocen como las condiciones de Mild.

A.1 Cálculo de la derivada direccional de f en forma vectorial

Sea

$$f(y_i, m) = \frac{1}{n} \sum_{i=1}^n |y_i - m|,$$

\hat{i} el vector unitario a lo largo del eje m ; \hat{j} el vector unitario a lo largo de las y_i .

Así que,

$$\vec{\nabla} = \hat{i} \frac{\partial}{\partial m} + \hat{j} \frac{\partial}{\partial y_i}$$

y la derivada direccional a lo largo de m , aproximándonos a un m_0 de izquierda a derecha es

$$\hat{i} \vec{\nabla} = \hat{i} \left(\hat{i} \frac{\partial}{\partial m} + \hat{j} \frac{\partial}{\partial y_i} \right)$$

Luego, se determina $\hat{i} \frac{\partial}{\partial m}$

$$\hat{i} \frac{\partial f(y_i, m)}{\partial m} = \frac{1}{n} \sum_I \hat{i} - \frac{1}{n} \sum_D \hat{i}$$

Por lo tanto,

$$\begin{aligned} \hat{u} \frac{\partial f(y_i, m)}{\partial m} &= \frac{1}{n} \sum_I \hat{u} - \frac{1}{n} \sum_D \hat{u} \\ &= \frac{1}{n} (r - s) \end{aligned}$$

HOJA DE METADATOS

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 1/6

Título	ESTUDIO Y APLICACIÓN DEL MODELO DE REGRESIÓN CUANTÍLICA
Subtítulo	

Autor(es)

Apellidos y Nombres	Código CVLAC / e-mail	
Rojas, Richard J.	CVLAC	16.256.902
	e-mail	richardjose Rojas@hotmail.com
	e-mail	
	CVLAC	
	e-mail	
	e-mail	
	CVLAC	
	e-mail	
	e-mail	
	CVLAC	
	e-mail	
	e-mail	

Palabras o frases claves:

Regresión cuantílica

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 2/6

Líneas y sublíneas de investigación:

Área	Subárea
Ciencias Básicas	Matemáticas

Resumen (abstract):

En este trabajo se introducen los conceptos y las propiedades básicas de la regresión cuantílica propuestos por Koenker y Bassett (1978). La regresión cuantílica, a diferencia de la regresión de la media, produce modelos que no sólo se basan en la estimación de la tendencia central de los datos sino que estiman toda la distribución de la variable dependiente. Para determinar los estimadores y los errores estándares de los modelos de regresión cuantílica, se requiere el uso de cualquier software con paquetes estadísticos que utilicen los métodos de regresión cuantílica. Se considera el software R para determinar las tablas de los estimadores y los errores estándares de una muestra de los datos del *International Seismological Centre* suministrados por el Centro de Sismología de la Universidad de Oriente.

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 3/6

Contribuidores:

Apellidos y Nombres	ROL / Código CVLAC / e-mail	
Ruiz, Nancy	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input checked="" type="checkbox"/> JU <input type="checkbox"/>
	CVLAC	4.187.024
	e-mail	najeru@yahoo.com.ar
	e-mail	
Avenidaño L., Jaime E.	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input type="checkbox"/> JU <input checked="" type="checkbox"/>
	CVLAC	8.438.392
	e-mail	javendano@udo.edu.ve
	e-mail	
Rodríguez, Milagros E.	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input type="checkbox"/> JU <input checked="" type="checkbox"/>
	CVLAC	6.611.165
	e-mail	melenamate@hotmail.com
	e-mail	
	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input type="checkbox"/> JU <input type="checkbox"/>
	CVLAC	
	e-mail	
	e-mail	

Fecha de discusión y aprobación:

Año Mes Día

2013	11	04
------	----	----

Lenguaje: spa

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 4/6

Archivo(s):

Nombre de archivo	Tipo MIME
Tesis-Richard.pdf	Aplication/pdf

Alcance:

Espacial: _____ Nacional _____ (Opcional)

Temporal: _____ Temporal _____ (Opcional)

Título o Grado asociado con el trabajo: Licenciado en Matemáticas

Nivel Asociado con el Trabajo: _____ Licenciatura _____

Área de Estudio:

Matemáticas _____

Institución(es) que garantiza(n) el Título o grado:

Universidad de Oriente Núcleo de Sucre _____

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 5/6



UNIVERSIDAD DE ORIENTE
CONSEJO UNIVERSITARIO
RECTORADO

CUN°0975

Cumaná, 04 AGO 2009

Ciudadano
Prof. JESÚS MARTÍNEZ YÉPEZ
Vicerrector Académico
Universidad de Oriente
Su Despacho

Estimado Profesor Martínez:

Cumplo en notificarle que el Consejo Universitario, en Reunión Ordinaria celebrada en Centro de Convenciones de Cantaura, los días 28 y 29 de julio de 2009, conoció el punto de agenda **"SOLICITUD DE AUTORIZACIÓN PARA PUBLICAR TODA LA PRODUCCIÓN INTELECTUAL DE LA UNIVERSIDAD DE ORIENTE EN EL REPOSITORIO INSTITUCIONAL DE LA UDO, SEGÚN VRAC N° 696/2009"**.

Leído el oficio SIBI - 139/2009 de fecha 09-07-2009, suscrita por el Dr. Abul K. Bashirullah, Director de Bibliotecas, este Cuerpo Colegiado decidió, por unanimidad, autorizar la publicación de toda la producción intelectual de la Universidad de Oriente en el Repositorio en cuestión.

UNIVERSIDAD DE ORIENTE
SISTEMA DE BIBLIOTECA
RECIBIDO POR *Martínez*
FECHA 5/8/09 HORA 5:20

Comunicación que hago a usted a los fines consiguientes.

Cordialmente,

Juan A. Bolaños Curyelo
JUAN A. BOLAÑOS CURYELO
Secretario



C.C: Rectora, Vicerrectora Administrativa, Decanos de los Núcleos, Coordinador General de Administración, Director de Personal, Dirección de Finanzas, Dirección de Presupuesto, Contraloría Interna, Consultoría Jurídica, Director de Bibliotecas, Dirección de Publicaciones, Dirección de Computación, Coordinación de Teleinformática, Coordinación General de Postgrado.

JABC/YGC/maruja

Hoja de Metadatos para Tesis y Trabajos de Ascenso - 6/6

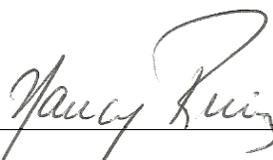
Artículo 41 del REGLAMENTO DE TRABAJO DE PREGRADO (vigente a partir del II Semestre 2009, según comunicación CU-034-2009):

“Los Trabajos de Grado son de la exclusiva propiedad de la Universidad de Oriente, y sólo podrán ser utilizados para otros fines con el consentimiento del Consejo del Núcleo respectivo, quién deberá participarlo previamente al Consejo Universitario, para su autorización.”



Br. Richard J. Rojas.

Autor



Profa. Nancy Ruiz

Tutor