



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

**Análisis de la regresión cuantílica para la distribución
del ingreso total mensual de la población
económicamente activa ocupada de Lima
Metropolitana**

TESINA

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Richard Henry HUIMAN MORALES

Lima, Perú

2016



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Huiman, R. (2016). *Análisis de la regresión cuantílica para la distribución del ingreso total mensual de la población económicamente activa ocupada de Lima Metropolitana*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

Dedicatoria

A Dios, porque en su infinita gracia me concede la vida.

Índice general

Índice general.....	3
Índice de Cuadros	4
Índice de Figuras.....	5
Resumen.....	6
Abstract.....	7
INTRODUCCIÓN	8
CAPÍTULO 1:.....	10
1.1 Situación Problemática	10
1.2 Formulación del Problema	11
1.3 Objetivo General.....	11
CAPÍTULO 2	12
MARCO TEÓRICO	12
2.1 Antecedentes del problema	12
2.2 Base Teórica.....	14
2.2.1 Definiciones Previas	14
2.2.2 Modelo de Regresión Cuantílica.....	16
2.2.3 Calculo de $\hat{\beta}_\tau$ por medio de programación lineal	17
2.3 Ventajas.....	20
CAPÍTULO 3	21
METODOLOGÍA	21
3.1 Cobertura.....	21
3.2 Fuentes de Información.....	21
3.3 Unidad de análisis	22
3.4 Población de estudio	22
3.5 Tipo de Muestreo	22
3.6 Base de Datos.....	22
3.7 Identificación de Variables	26
3.8 Procesamiento y análisis de datos	26
CAPÍTULO 4	28
RESULTADOS	28
4.1 Análisis Descriptivo de las Variables	28
4.2 Estimación por regresión cuantílica.....	32
CONCLUSIÓN	41

BIBLIOGRAFÍA.....	42
ANEXOS	43
Programación en R-Proyect	43
Programación en STATA	45

Índice de Cuadros

Cuadro 1:	
Estructura de la Población según condición de actividad	23
Cuadro 2:	
Lima Metropolitana: Población ocupada, según sexo Trimestre móvil:	
Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)	24
Cuadro 3:	
Lima Metropolitana: Población ocupada, según grupos de edad	
Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)	25
Cuadro 4:	
Lima Metropolitana: Población ocupada, según nivel de educación alcanzado	
Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)	25
Cuadro 5:	
Lima Metropolitana: Descripción Estadística del Ingreso total mensual	
de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles).....	28
Cuadro 6:	
Lima Metropolitana: Descripción Estadística del Ingreso total mensual según sexo	
de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles).....	28
Cuadro 7:	
Lima Metropolitana: Descripción Estadística del Ingreso total mensual de la población	
ocupada según nivel educativo Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)	29
Cuadro 8:	
Lima Metropolitana: Descripción Estadística del Promedio de horas trabajadas	
por semana de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)	
.....	29
Cuadro 9:	
Lima Metropolitana: Descripción Estadística del Ingreso total mensual según	
grupos de edad de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016	
(Soles)	30
Cuadro 10:Lima Metropolitana: Ingreso total mensual por cuantiles	

Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)	30
Cuadro 11:	
Lima Metropolitana: Ingreso total mensual por cuantiles según sexo, nivel educativo y grupos de edad Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)	31
Cuadro 12:	
Resultados de la regresión cuantílica(Coeficientes y Errores estándar)	
Variable dependiente: Log ingreso total mensual	32

Índice de Figuras

Figura1. Coeficientes de regresiones por cuantiles del intercepto. (Variable dependiente: Log ingreso total mensual)	34
Figura2. Coeficientes de regresiones por cuantiles del variable sexo. (Variable dependiente: Log ingreso total mensual).....	35
Figura3. Coeficientes de regresiones por cuantiles de la variable edad. (Variable dependiente: Log ingreso total mensual)	36
Figura4. Coeficientes de regresiones por cuantiles de la variable total de horas trabajadas a la semana. (Variable dependiente: Log ingreso total mensual)	36
Figura5. Coeficientes de regresiones por cuantiles de la variable nivel educativo secundaria. (Variable dependiente: Log ingreso total mensual)	37
Figura6. Coeficientes de regresiones por cuantiles de la variable nivel educativo superior no universitaria. (Variable dependiente: Log ingreso total mensual)	38
Figura7. Coeficientes de regresiones por cuantiles de la variable nivel educativo superior universitaria. (Variable dependiente: Log ingreso total mensual)	38
Figura8. Bondad de Ajuste del Modelo de la Regresión Cuantílica	40

Resumen

En el presente trabajo de investigación se aplica el método de la regresión cuantílica como un método alternativo de estimación de los parámetros en los modelos de regresión lineal para analizar la distribución del ingreso total mensual de la población ocupada de Lima Metropolitana.

Para estimar los parámetros de un modelo de regresión clásica es por el método de Mínimo Cuadrados Ordinarios (MCO), sin embargo esta provee poca información acerca del comportamiento de los extremos (colas) de la distribución del ingreso total mensual. En este caso, no es adecuado utilizar el modelo de regresión lineal estimado por mínimos cuadrados ordinarios, ya que proporciona estimaciones sesgadas. Se concluye la influencia positiva del nivel educativo y el sexo masculino en el ingreso total mensual de la población ocupada de Lima Metropolitana.

Palabra claves: Regresión, cuantil, regresión mediana, regresión cuantílica, ingreso total mensual.

Abstract

In this work the quantile regression method is applied as an alternative method of estimating the parameters in the linear regression models to analyze the distribution of total monthly income of the employed population of Metropolitan Lima.

To estimate the parameters of a classical regression model it is by the method of Minimum Ordinary Squares (OLS), however this provides little information about the behavior of the ends (tails) of the distribution of total monthly income. In this case, it is not appropriate to use the linear regression model estimated by ordinary least squares, since it provides biased estimates. We conclude the positive influence of the educational level and the male sex in the total monthly income of the employed population of Metropolitan Lima.

INTRODUCCIÓN

El Análisis de Regresión en General es una de las Técnicas estadísticas más utilizadas para la investigación debido a que nos permite modelar las relaciones entre las variables. Su campo de dominio o aplicaciones se concentra en la Ingeniería, las Ciencias Físicas, Biológicas, entre otras.

El término regresión fue utilizado por primera vez en un estudio realizado por Francis Galton sobre variables antropométricas en 1889, al comparar la estatura de padres e hijos, resultó que los hijos cuyos padres tenían una estatura superior al valor medio tendían a igualarse a este; mientras que aquellos cuyos padres eran muy bajos, tendían a reducir su diferencia respecto a la estatura media; es decir, “regresaban” al promedio. El término lineal es utilizado para distinguir de las demás técnicas de regresión, que emplean modelos basados en cualquier función matemática, como por ejemplo cuadráticas, cúbicas, exponenciales, etc.

Uno de los métodos más empleados para definir el modelo (ecuación lineal) de regresión es el método de mínimos cuadrados, el cual fue propuesto por Legendre en 1805 y Gauss en 1809. El término “mínimos cuadrados” proviene de la descripción dada por Legendre “moindres carrés”.

En este trabajo se presentará una alternativa de análisis de regresión: la regresión cuantílica que es una técnica de modelamiento de gran flexibilidad que permite la descripción del cuantil condicional de una variable respuesta Y dado un conjunto de variables explicativas X , siendo la regresión mediana un caso particular. Este tipo de regresión se ha desarrollado durante los últimos 30 años, logrando extenderse a varias áreas.

Mediante este método se analiza la distribución del ingreso total mensual usando variables independientes como: Edad, sexo, nivel educativo, total de horas

trabajadas. La fuente de datos es la Encuesta Permanente de Empleo En Lima Metropolitana – Trimestre Móvil (Julio-Agosto-Setiembre 2016).

La tesina está organizado de la siguiente forma: En el capítulo 2, está centrado en la presentación de un conjunto de definiciones que serán utilizadas en los siguientes capítulos de la presente tesina, los antecedentes previos y una revisión de las estadísticas de orden, el modelo de regresión cuantílica y su respectivo método de estimación.

En el capítulo 3, se muestra el desarrollo de la metodología y la aplicación de la regresión cuantílica con base de datos reales (ingreso total mensual de la población ocupada de lima metropolitana)

En el capítulo 4, se muestra los resultados del trabajo y sus principales conclusiones

Finalmente, en el anexo se presenta cuadros estadísticos y programas implementados en la aplicación al conjunto de datos reales.

CAPÍTULO 1:

1.1 Situación Problemática

En la actualidad, se muestra que la desigualdad del ingreso laboral ha venido reduciéndose en los últimos años. Por ello resulta de interés conocer los cambios operados en los factores que lo explican: el nivel educativo, el género, la edad y el total de horas trabajadas a la semana.

El problema de la distribución de los ingresos tiene origen en el modelo de crecimiento que se ha asentado en el país.

La desigualdad de los ingresos entre los varones y mujeres es un fenómeno mundial, persistente y de difícil reversión. En los últimos años las estadísticas acerca de las brechas de ingresos, así como estudios empíricos desarrollados en el país a través de los años muestran que las diferencias persisten.

Esta tesina analiza la relación entre el ingreso total mensual y el sexo, edad, nivel educativo, total de horas trabajadas a la semana con data trimestral para el período 2016 por medio de un modelo. Sin embargo los ingresos totales mensuales de las personas de Lima Metropolitana resultan muy heterogéneos y con una presencia de observaciones atípicas y esto podría sobrevalorar la estimación MCO. Por lo que se plantea como alternativa una estimación a través de la regresión cuantílica.

1.2 Formulación del Problema

¿Qué relación existe entre el, sexo, edad, nivel educativo, número de horas trabajadas a la semana con el ingreso total mensual durante el período agosto-setiembre-octubre 2016?

1.3 Objetivo General

Determinar si el sexo, edad, nivel educativo y número de horas trabajadas a la semana tienen una relación importante con el ingreso total mensual mediante el análisis de la regresión cuantílica durante el período agosto-setiembre-octubre 2016.

CAPÍTULO 2

MARCO TEÓRICO

2.1 Antecedentes del problema

Los primeros intentos para la realizar el análisis de regresión están estrechamente relacionados con la regresión Cuantílica.

La regresión de mínimos cuadrados datan del año 1805, por los trabajos de Legendre, a continuación el trabajo inicial de Boscovich en regresión fue medio siglo antes. El problema interés de Boscovich era la elipticidad de la tierra.

El trabajo de Edgeworth sobre los métodos medianos para la regresión lineal nos lleva directamente a la regresión cuantil. (Edgeworth, 1888) descarta la restricción Boscovich-Laplace de que los residuos suman a cero y propone minimizar la suma de los residuos absolutos tanto en los parámetros de intercepción y pendiente, Llamándolo un método de "doble mediana" y señalando que podría ampliarse, en principio, a un método "mediana plural ". Proporcionó un algoritmo geométrico para el caso bivariado, y se estableció una discusión de condiciones bajo las cuales se preferiría minimizar el error absoluto en lugar del error cuadrático bien establecido por entonces. Desafortunadamente, el enfoque geométrico para calcular el nuevo estimador de regresión mediano de Edgeworth era bastante complejo, requiriendo, como admitió más tarde, "la atención de un matemático; Y en el caso de muchas incógnitas, algún poder de concepción hipergeométrica". Sólo con posterioridad considerablemente el advenimiento de la programación lineal proporcionó un enfoque computacional conceptualmente simple y eficiente.

El análisis de regresión cuantílica, desarrollado por Koenker y Bassett (1978), complementa y amplía el análisis clásico de la regresión mínimo Cuadrática.

Un trabajo sobre estos temas de la regresión cuantílica aplicados a los ingresos corresponde a los resultados reportados por (Budig & Hodges, 2010) que confirman la asociación entre diferentes covariables y los ingresos, ubicados en diferentes partes de la distribución de la renta relativa. Las regresiones cuantílicas permiten la medición de las asociaciones entre variables específicas y variables independientes y cuantiles específicos de la distribución de ingreso (variable dependiente). En su forma lineal, se puede expresar la regresión cuantílica de manera robusta aún frente a valores extremos y que no asumen ninguna distribución previa para el término de error, a diferencia del método de mínimo cuadrados ordinarios que asume que el error tiene una distribución normal

Otros temas de la regresión cuantílica aplicados a los ingresos corresponde a (Orlandoni Merli, Ramoni Perazzi, & Pérez Pulido, 2015), basado en la Gran Encuesta Integrada de Hogares del DANE, donde se analiza la distribución del ingreso laboral mediante regresión cuantílica usando covariables como edad, nivel educativo, experiencia y género y concluyen la influencia positiva de la experiencia y el nivel educativo en el ingreso laboral de los trabajadores, sobre todo en los formales masculinos.

Un trabajo a nivel nacional sobre estos temas se remonta en (Ministerio de Trabajo y Promoción del Empleo, 2006), basados de las Encuestas de Hogares Especializada en Niveles del Empleo, ejecutadas por el Ministerio de Trabajo y Promoción del Empleo en Lima Metropolitana en el periodo 1990 – 2004, donde tiene como objetivo analizar la evolución de la distribución de ingresos y sus indicadores de dispersión en Lima Metropolitana, durante los años 1990 – 2004.

2.2 Base Teórica

2.2.1 Definiciones Previas

2.2.1.1 Cuantil:

El cuantil de orden τ de una distribución (con $0 < \tau < 1$) es el valor de la variable Y que marca un corte de modo que una proporción τ de valores de la población es menor o igual que Y .

$$Q(\tau) = \inf \{Y : F(Y) \geq \tau\}$$

Donde F es una Distribución de Y

El cálculo de los cuantiles utilizando la definición anterior involucra la obtención de las estadísticas de orden Y que a continuación definiremos dicho concepto

2.2.1.2 Estadísticas de orden:

Sea $\{y_1, y_2, \dots, y_n\}$ una muestra aleatoria Proveniente de la distribución F . Los valores muestrales ordenados

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

Son llamados estadísticos de orden de la muestra

2.2.1.3 Distribución Empírica:

Sea Y_1, Y_2, \dots, Y_n una muestra con observaciones independientes entonces la distribución empírica se denota por:

$$\hat{F}(Y) = \frac{\#(Y_i \leq Y)}{n}$$

Definimos entonces una estimación para los cuantiles, por medio de la distribución empírica como:

$$\hat{Q}(\tau) = \inf \{Y : \hat{F}(Y) \geq \tau\}$$

Que es equivalente

$$\hat{Q}(\tau) = \arg \min_{\varepsilon_\tau \in \mathbb{R}} \left\{ \sum_{Y_i \geq \varepsilon_\tau} \tau \cdot (Y_i - \varepsilon_\tau) + \sum_{Y_i < \varepsilon_\tau} (1 - \tau) \cdot (\varepsilon_\tau - Y_i) \right\}$$

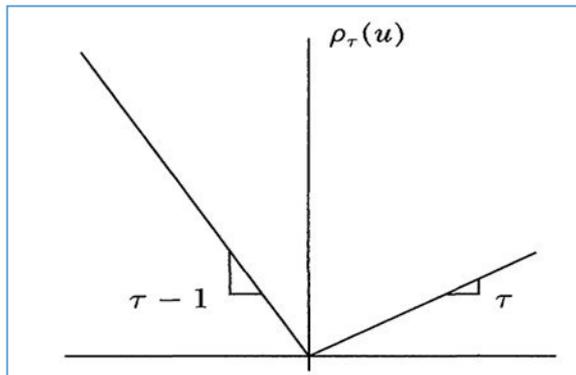
Y también podemos expresar como:

$$\hat{Q}(\tau) = \arg \min_{\varepsilon_\tau \in \mathbb{R}} \left\{ \sum_{Y_i \geq \varepsilon_\tau} \rho_\tau(Y_i - \varepsilon_\tau) \right\}$$

Donde la función de pérdida es descrita por $\rho_\tau(r) = r \cdot (\tau - I\{r < 0\})$ con $\tau \in (0, 1)$

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

Regresión cuantílica: Función de pérdida ρ



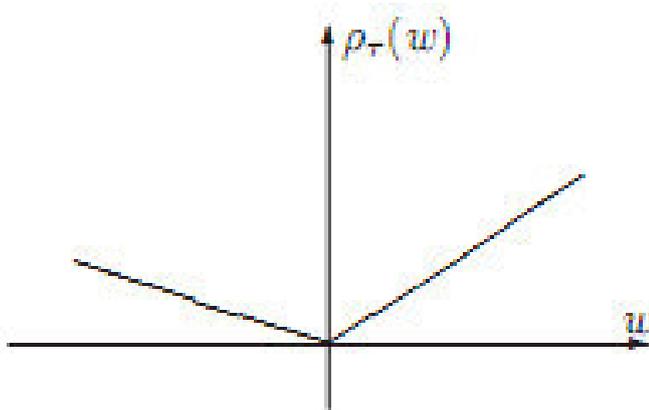
Como observamos la regresión cuantílica utiliza la función $\rho_\tau : \mathbf{R} \rightarrow \mathbf{R}$, con $0 < \tau < 1$ (Koenker y Bassett 1978). Las siguientes tres definiciones equivalentes.

$$\rho_{\tau}(w) = \begin{cases} (\tau - 1)w & \text{si } w < 0 \\ \tau w & \text{si } w \geq 0 \end{cases} \quad (2a)$$

$$\rho_{\tau}(w) = \max\{(\tau - 1)w, \tau w\} \quad (2b)$$

$$\rho_{\tau}(w) = \left(\tau - \frac{1}{2}\right)w + \frac{1}{2}|w| \quad (2c)$$

Ejemplo: Para $\tau = 2/3$, se tiene la siguiente gráfica:



2.2.2 Modelo de Regresión Cuantílica

Se define la regresión cuántilica lineal

$$Y_i = \beta_{0,\tau} + \beta_{1,\tau} \cdot X_i + \varepsilon_{i,\tau} \quad \forall i \in \{1, \dots, n\}$$

Con $0 < \tau < 1$ además el τ -ésimo cuantil del error con respecto a la variable regresora es cero ($Q_{\tau}(\varepsilon_{i,\tau} / X) = 0$), entonces el τ -ésimo cuantil de Y_i con respecto a X se expresa como

$$Q_{\tau}(Y_i / X) = \beta_{0,\tau} + \beta_{1,\tau} \cdot X_i$$

A continuación para obtener las estimaciones de los parámetros de $\beta_{0,\tau}$ y $\beta_{1,\tau}$ utilizaremos por medio de

$$\hat{\beta}_{\tau} = \arg \min_{\beta_{\tau} \in \mathbf{R}^2} \left\{ \sum_{Y_i \geq A} \tau \cdot |Y_i - \beta_{0,\tau} - \beta_{1,\tau} \cdot X_i| + \sum_{Y_i < A} (1 - \tau) \cdot |Y_i - \beta_{0,\tau} - \beta_{1,\tau} \cdot X_i| \right\}$$

Dónde:

$$\beta_{\tau} = (\beta_{0,\tau}, \beta_{1,\tau}) \quad \text{y} \quad A = \beta_{0,\tau} + \beta_{1,\tau} \cdot X_i$$

Como se observa para estimar los parámetros se tiene que minimizar la función $\hat{\beta}_{\tau}$ y para afrontar un problema de minimización se utiliza el método de programación lineal.

Dados m puntos x^1, x^2, \dots, x^m en $\mathbf{R}^{n \times 1}$ y m valores reales y_1, y_2, \dots, y_m , en el problema de regresión se busca un vector.

2.2.3 Cálculo de $\hat{\beta}_{\tau}$ por medio de programación lineal

Para la resolución de la ecuación (1) según lo propuesto por (Koenker y Basset 1978), la función ρ_{τ} también se puede definir por medio de las funciones parte positiva y parte negativa.

$$\begin{aligned}
p^+(w) &= \begin{cases} 0 & \text{si } w < 0 \\ w & \text{si } w \geq 0 \end{cases} \\
p^-(w) &= \begin{cases} -w & \text{si } w < 0 \\ 0 & \text{si } w \geq 0 \end{cases}
\end{aligned} \tag{2}$$

Entonces

$$\rho_\tau(w) = \tau p^+(w) + (1 - \tau) p^-(w)$$

En particular,

$$\begin{aligned}
w &= p^+(w) - p^-(w) \\
|w| &= p^+(w) + p^-(w)
\end{aligned}$$

Tomando $u_i = p^+(y_i - \beta^T x^i)$, $v_i = p^-(y_i - \beta^T x^i)$, $u = (u_1, u_2, \dots, u_m)$, $v = (v_1, v_2, \dots, v_m)$ y $e_m = (1, 1, \dots, 1) \in \mathbf{R}^m$ (e es un vector de unos), entonces la ecuación (1) se escribe de la siguiente forma:

$$\begin{aligned}
\min f(\beta, u, v) &= \sum_{i=1}^m \tau u_i + \sum_{i=1}^m (1 - \tau) v_i \\
u_i - v_i &= y_i - \beta^T x^i \\
u_i, v_i &\geq 0
\end{aligned}$$

$$\min f(\beta, u, v) = \begin{bmatrix} 0_n^T & \tau e_m^T & (1 - \tau) e_m^T \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} X & I_m & -I_m \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} = y$$

$$u, v \geq 0$$

Donde X es la matriz $m \times n$ cuyas filas son los vectores x^1, x^2, \dots, x^m (tienen que ser transpuestos para que puedan ser filas de la matriz). Entonces lo anterior es un problema de *Optimización lineal* con $n+2m$ variables restricciones (igualdades) y $2m$ variables no negativas.

La formulación del problema dual para regresión cuantílica es equivalente a la usada en la formulación estándar de los métodos de punto interior para programación lineal con variables acotadas. Dicho algoritmo se encuentra descrito en Koenker (2005) e implementado en el paquete *quantreg* del software estadístico R. Este paquete es el más usado por los investigadores que trabajan la regresión cuantílica.

Hasta este momento, hemos conseguido expresar la búsqueda de un cuantil como la solución de un problema de programación lineal.

Barrodale y Roberts (1973) proponen una simplificación de la forma estándar del método del Simplex para el resolver el problema del cálculo de los estimadores en el caso de la regresión en mediana, donde la función de pérdida sería el valor absoluto. Posteriormente, Koenker y D'Orey (1987) extendieron este razonamiento a cualquier cuantil $0 < \tau < 1$.

Realmente, la simplificación del método del Simplex es consecuencia de que en un único paso del algoritmo propuesto por Barrodale y Roberts (1973) realizamos varios pasos del método del Simplex. Como consecuencia, este nuevo algoritmo es computacionalmente mucho más reciente que el método del Simplex clásico.

Por último, 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. En regresión cuantílica, para obtener la bondad de ajuste del modelo, se cuenta con el Pseudo- R^2 , como medida equivalente al R^2 del MCO:

$$Pseudo - R^2 = 1 - \frac{\sum_{i=1}^n |y_i - \hat{y}_1|}{\sum_{i=1}^n |y_i - \hat{y}_\tau|} \quad (8)$$

2.3 Ventajas

La regresión en la cual se modela el valor esperado condicional $E[y/x]$ constituye la manera más común de modelamiento, sin embargo Koenker (2004,cap.2) expone ejemplos en los cuales la regresión cuantílica provee ventajas como:

- Permitir modelar los extremos de la variable respuesta.
- Permitir identificar mejor el efecto de las covariables sobre la distribución condicional.
- Brinda mayor flexibilidad en el modelamiento de los datos con altos niveles de variabilidad, describiendo el comportamiento para cada cuantil deseado.

CAPÍTULO 3

METODOLOGÍA

El enfoque es de tipo cuantitativo, en el tipo del diseño es de tipo no experimental observacional. Se realiza esta investigación pues no se realiza una manipulación deliberada de la variable de interés, es decir, se trata de un estudio donde no hacemos variar en forma intencional la variable independiente. Lo que se realiza en la investigación no experimental es observar fenómenos tal como se da en su contexto natural.

3.1 Cobertura

Área Metropolitana de Lima y Callao, constituida por 43 distritos en la Provincia de Lima y 6 distritos de la Provincia Constitucional del Callao.

3.2 Fuentes de Información

La fuente de información lo constituye la Encuesta Permanente de empleo Lima Metropolitana - 2016. Especialmente el trimestre móvil julio – agosto – setiembre.

La Encuesta Permanente de Empleo (EPE), se viene efectuando desde marzo 2001 y sus objetivos son:

- Generar indicadores sobre empleo e ingreso en el Área Metropolitana de Lima y Callao, para el seguimiento y análisis del mercado laboral.
- Desarrollar indicadores anticipatorios de la evolución del empleo, para fines prospectivos.

- Servir de fuente de información a instituciones públicas y privadas; así como a investigadores.
- Permitir la comparabilidad con investigaciones similares, en relación con las variables investigadas.

3.3 Unidad de análisis

Un integrante de la población económicamente activa ocupada en el área Metropolitana de Lima y Callao

3.4 Población de estudio

La población bajo estudio está constituida por el conjunto de personas económicamente activa ocupada en el área Metropolitana de Lima y Callao.

3.5 Tipo de Muestreo

El tipo de muestro de la Encuesta Permanente de Empleo de Lima Metropolitana es Probabilístico, de áreas y bietápica (sistemático con PPT en la primera etapa y sistemático simple de una muestra compacta en la segunda etapa). La muestra es estratificada de manera implícita, porque previamente a la selección la población se ha dividido en estratos socioeconómicos, con el objeto de mejorar su representatividad.

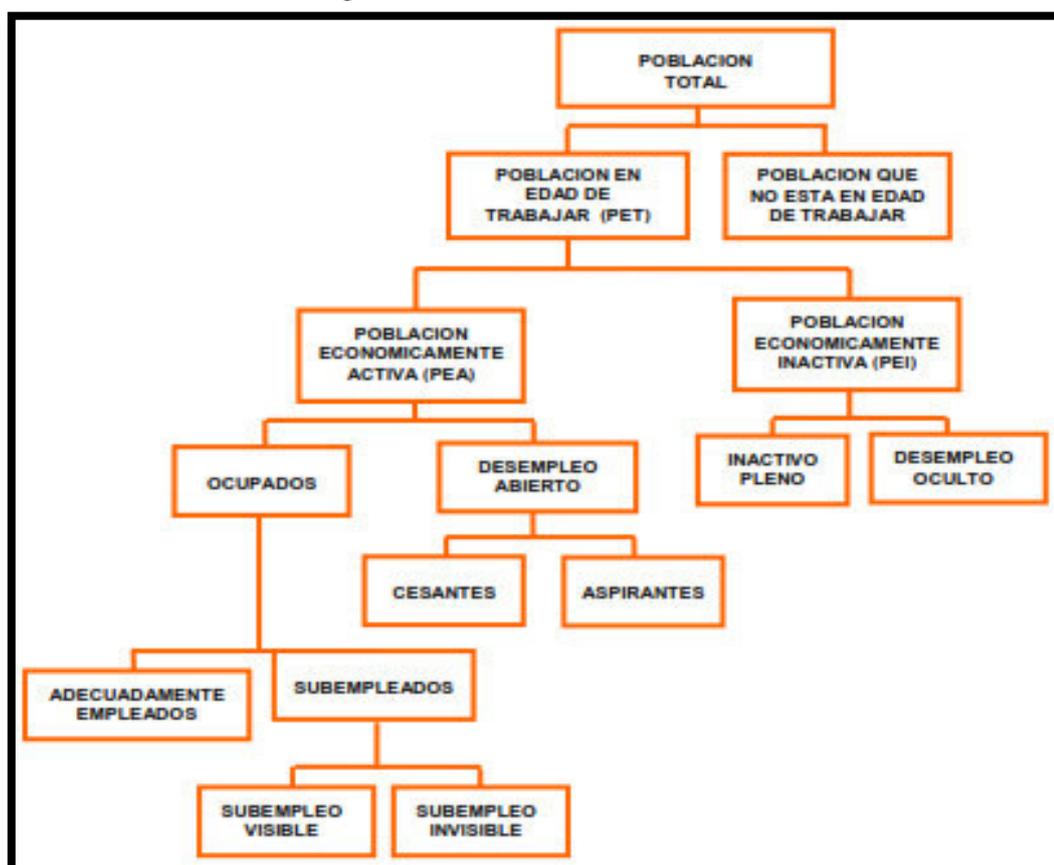
3.6 Base de Datos

Como se ha mencionado anteriormente la base de datos con lo que se trabaja en la presente tesina proviene de la Encuesta Permanente de Empleo en Lima Metropolitana 2016, Dicha base de datos fue obtenida desde la sección de Microdatos del INEI, este sistema, proporciona las bases de datos y la documentación derivada de las investigaciones y encuestas ejecutadas por el INEI durante los últimos años, resguardando el secreto estadístico de la información.

Se ha seleccionado solo la *población económicamente activa ocupada* que viene a ser parte de la Población económicamente Activa para la aplicación del modelo de la regresión cuantílica.

Cuadro 1:

Estructura de la Población según condición de actividad



Fuente: (Instituto Nacional de Estadística e Informática, 2000)

La “Población económicamente activa ocupada” es el conjunto de todas las personas que contando con la edad mínima especificada para la medición de la PEA durante el período de referencia, se encontraban realizando “algún trabajo” (13° CIET), ya sea como “Asalariado”, percibiendo un sueldo o salario, monetario o en especie o como “Empleado Independiente”, obteniendo un beneficio o ganancia familiar, monetario o en especie.

Para determinar que una persona se encuentra ocupada se utilizan cuatro criterios:

- i. Ocupados son las personas de 14 años y más de edad que estuvieron participando en alguna actividad económica, en el período de referencia.
- ii. Los trabajadores dependientes, que teniendo empleo fijo, no trabajaron, la semana anterior, por hallarse de vacaciones, huelga, licencia por enfermedad, licencia pre y post-natal, etc., todas ellas pagadas.
- iii. Los trabajadores independientes, que estuvieron temporalmente ausentes del trabajo durante el período de referencia; pero, la empresa o negocio siguió funcionando.
- iv. A las personas que no estuvieron en ninguna de las condiciones anteriores se les indaga si realizaron alguna actividad económica en el período de referencia, al menos una hora, por lo cual recibirá pago en dinero y/o especie. El objetivo es recuperar las actividades realizadas; pero, que no son consideradas como trabajo por las personas.

A continuación se presenta la información correspondiente a la Población económicamente activa ocupada según sexo, grupos de edad y nivel educativo alcanzado según el informe técnico elaborado por el Instituto Nacional de Estadística e Informática (INEI).

Según INEI en el trimestre julio-agosto-setiembre 2016, la población ocupada de Lima Metropolitana alcanzó los 4 millones 736 mil 300 personas.

Cuadro 2:

Lima Metropolitana: Población ocupada, según sexo Trimestre móvil:

Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)

	Valor Estimado	Intervalo Inferior	Intervalos Superior
Sexo			
Hombre	2600.6	2519	2682.1
Mujer	2135.7	2062.9	2208.5

Fuente: Instituto Nacional de Estadística e Informática- Encuesta Permanente de Empleo 2016

En el cuadro 2 se observa que del total de población ocupada, el 54,9% (2 millones 600 mil 600) son hombres y el 45,1% (2 millones 135 mil 700) son mujeres.

Cuadro 3:

Lima Metropolitana: Población ocupada, según grupos de edad
Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)

	Valor Estimado	Intervalo Inferior	Intervalos Superior
Grupos de edad			
De 14 a 24 años	953	888.7	1017.2
De 25 a 44 años	2576	2483.8	2668.2
De 45 y más años	1207.3	1167.7	1246.9

Fuente: Instituto Nacional de Estadística e Informática- Encuesta Permanente de Empleo 2016

En el cuadro 3 se observa que del total de población ocupada, el 20.1% (953 mil) tienen entre 14 a 24 años de edad; el 55.4% (2 millones 576 mil 600) de 25 a 44 años y el 25.5% (1 millón 207 mil 600) de 45 y más años de edad.

Cuadro 4:

Lima Metropolitana: Población ocupada, según nivel de educación alcanzado
Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles) (miles de personas)

	Valor Estimado	Intervalo Inferior	Intervalos Superior
Grupos de edad			
Primaria 1/	372.2	338.4	406
Secundaria	2203.9	2100.3	2307.6
Superior no universitaria	948.5	893	1004
Superior universitaria	1211.6	1135.7	1287.6

1/ incluye inicial y Sin nivel.

Fuente: Instituto Nacional de Estadística e Informática- Encuesta Permanente de Empleo 2016

En el cuadro 4 se observa que del total de población ocupada, el 7.9% tienen primaria o menor nivel educativo, el 46.5% educación secundaria, el 20% superior no universitaria y el 25.6% cuenta con educación universitaria.

3.7 Identificación de Variables

2.2.3.1 Variables Dependientes o Respuesta:

La Variable Respuesta o dependiente usada en el modelo de la regresión cuantílica es de tipo cuantitativa continua

Y_i = Ingreso Total Mensual (Soles) (Variable cuantitativa continua)

Es el ingreso total mensual por trabajo (ingreso monetario y en especie), proveniente de la actividad principal y secundaria (s). Para nuestro análisis se excluye al trabajador familiar no remunerado (que trabaja de 15 a más horas a la semana) y a los practicantes que no reciben ningún tipo de ingreso (ni monetario ni en especie).

2.2.3.2 Variables Independientes o Predictoras (Covariables):

X_1 = Sexo (Variable cualitativa nominal)

Categoría: Hombre; Mujer

X_2 = Edad (Variable cuantitativa discreta)

X_3 = Total de horas trabajadas (Variable cuantitativa)

X_4 = Nivel Educativo: (Variable cualitativa discreta)

Categoría: Primaria; Secundaria; superior no universitaria y superior universitaria

3.8 Procesamiento y análisis de datos

El análisis se desarrolló en 2 etapas, en la primera se realizó una descripción descriptiva de las variables involucradas al modelo.

Debido a que nuestro conjunto de datos no cumple con los supuestos del MCO es decir requiere unas hipótesis previas sobre la aleatoriedad de la relación [1] expresadas en términos de que los errores (residuos) siguen una distribución normal con media cero y varianza sigma al cuadrado (homocedasticidad), entonces 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. En definitiva se planteó un modelo de regresión cuantílica para evaluar el efecto de los factores que influyen en la distribución del ingreso total mensual de la población ocupada de Lima Metropolitana.

El modelo empírico a estimar tiene la siguiente forma general:

$$\begin{aligned} \text{Log}(Y_i) = & \beta_{0,\tau} + \beta_{1,\tau}\text{Sexo} + \beta_{2,\tau}\text{Edad} + \beta_{3,\tau}\text{Tot.Hrs.Trabajadas} \\ & + \beta_{4,\tau}\text{Niv.Educativo} + \varepsilon_{i,\tau} \end{aligned}$$

Donde Y_i es el ingreso total mensual, X_i para $i = 1,2,3,4$ son las covariables o variables independientes y finalmente $\varepsilon_{i,\tau}$ representa el error.

Para establecer la correcta transformación del ingreso total mensual se ha usado frecuentemente la función logaritmo, dado que esta se aproxima a una distribución normal, también algunos estudios han demostrado que esta transformación es la mejor en la clase de transformaciones de Box-Cox, sin embargo para nuestro análisis va más para su interpretación en el análisis como cambio porcentual es muy conveniente.

Para la realización de este trabajo se ha utilizado mayoritariamente el software libre R, SPSS y STATA

Dentro del software R, hemos utilizado la librería *quantreg* que ha sido desarrollada por R. Koenker y está dedicada íntegramente a la regresión cuantil. Dentro de dicha librería tenemos disponible la función *rq* que nos permite ajustar modelos lineales de regresión cuantil.

Este paquete enfoca la regresión cuantil desde su resolución como un problema de programación lineal. Dicho problema primal se transforma en su dual, para lograr así una disminución del número de variables. El algoritmo utilizado para su resolución se encuentra descrito en (Koenker, Quantile Regression, 2005) e implementado en el paquete *quantreg* (Koenker, *quantreg: Quantile Regression*, 2010).

CAPÍTULO 4

RESULTADOS

4.1 Análisis Descriptivo de las Variables

Desde el cuadro 1 al cuadro 5 se puede apreciar la el análisis descriptiva de las variables que involucran al modelo. En el cuadro 1 observamos que el ingreso promedio de la población ocupada de lima metropolitana es de 1655.4 nuevos soles sin embargo notamos que la variable tiene mucha variabilidad.

Cuadro 5:

Lima Metropolitana: Descripción Estadística del Ingreso total mensual de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

	Media	Desviación Est	Mediana	Asimetría
Total	1655.44	1521.9	1299	3.88

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 2 se observa una desigualdad en el ingreso total mensual según sexo de la población ocupada de lima metropolitana, los hombres tiene un ingreso promedio mensual aproximadamente de 1900 nuevos soles en cambio las mujeres solo tienen un ingreso promedio mensual de 1345 nuevos soles.

Cuadro 6:

Lima Metropolitana: Descripción Estadística del Ingreso total mensual según sexo de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

	Media	Desviación Est	Mediana	Asimetría
Sexo				
Hombre	1905	1633.962	1500	3.434
Mujer	1345	1304.694	1034	5.06

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 3 se aprecia que el ingreso promedio mensual de la población ocupada de lima metropolitana que tienen un nivel educativo superior universitaria es de 2675 nuevos soles que es evidentemente mayor a la población que no tienen el grado superior universitaria, por ejemplo observamos que la diferencia del ingreso promedio mensual de la población ocupada que tiene un nivel educativo primario y secundario no es mucha debido a que sus ingresos son de 1049 y 1236 nuevos soles respectivamente.

Cuadro 7:

Lima Metropolitana: Descripción Estadística del Ingreso total mensual de la población ocupada según nivel educativo Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

	Media	Desviación Est	Mediana	Asimetría
Nivel Educativo				
Primaria	1049.4	685.7	1000	1.62
Secundaria	1236	900.3	1100	4.157
Superior no universitaria	1554	1083.26	1400	4.019
Superior universitaria	2675.76	2240.13	2000	2.662

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 4 se observa que el promedio de horas trabajadas por semana de la población ocupada de lima metropolitana es de 44.75 horas es decir un poco menos de 8 horas diarias (no se consideran los domingos).

Cuadro 8:

Lima Metropolitana: Descripción Estadística del Promedio de horas trabajadas por semana de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

	Media	Desviación Est	Mediana	Asimetría
Total	44.75	17.426	47	-0.236

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 5 la variable edad se clasifica en tres grupos (de 14 a 24 años, 25 a 44 años y 45 a más años) y se aprecia que existe una diferencia muy marcada respecto a la población ocupada de lima metropolitana, las personas de 14 a 24 años tienen un ingreso promedio mensual de aproximadamente de 975 nuevos soles sin embargo las personas de 25 y más años tienen un ingreso promedio de casi el doble. Observamos que todas las variables tienen una distribución asimétrica positiva debido a que su coeficiente de asimetría es mayor que cero a excepción de la variable total de horas trabajadas que tiene una leve distribución asimétrica negativa.

Cuadro 9:

Lima Metropolitana: Descripción Estadística del Ingreso total mensual según grupos de edad de la población ocupada Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

	Media	Desviación Est	Mediana	Asimetría
Grupos de Edad				
De 14 a 24 años	975.71	675.263	900	1.445
De 25 a 44 años	1799.45	1504.071	1500	4.114
De 45 y más años	1864.32	1847.947	1400	3.158

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 6, se observa una gran desigualdad del ingreso total mensual de la población ocupada de lima metropolitana: el cuantil 0.5 (mediana) recoge a la población ocupada que reciben un ingreso mensual de aproximadamente de 1300 nuevos soles o menos, además se observa que el cuantil 0.75 (cuartil 3) recoge a la población ocupada que percibe un ingreso total mensual de 2000 nuevos soles o menos, también se aprecia que el cuantil 0.25 (cuartil 1) recoge a la población ocupada que recibe un ingreso total mensual de 850 nuevos soles (sueldo mínimo) o menos.

Cuadro 10:

Lima Metropolitana: Ingreso total mensual por cuantiles Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

Cuantil	0.05	0.1	0.25	0.5	0.75	0.9	0.95
Ingreso total mensual	200	410	850	1299	2000	3000	4000

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

En el cuadro 7 muestra la distribución del ingreso total mensual por cuantiles, clasificada por sexo, nivel educativo y grupos de edad. Esto permite examinar los cambios que se han dado en la distribución condicional del ingreso total mensual, con el propósito de estimar su efecto sobre las variables incluidas en el modelo.

El 50% de la población ocupada masculina de lima metropolitana percibe un ingreso total mensual de 1500 nuevos soles o menos en cambio en el lado femenino solo percibe un ingreso total mensual de 1034 nuevos soles. Respecto al nivel educativo, observamos que la distribución de los ingresos totales mensuales son más altos los que tienen un nivel educativo superior respecto a los que tienen un nivel educativo bajo (primaria y secundaria) a los largo de los cuantiles, asimismo la población ocupada de lima metropolitana cuyas edades oscilan entre 14 a 24 años perciben ingreso mensuales bajos respecto a la población ocupada cuyas edades son de 25 años a más.

Cuadro 11:

Lima Metropolitana: Ingreso total mensual por cuantiles según sexo, nivel educativo y grupos de edad Trimestre móvil: Julio-Agosto-Setiembre 2016 (Soles)

Cuantil	0.05	0.1	0.25	0.5	0.75	0.9	0.95
Sexo							
Hombre	390	702	1013	1500	2159	3500	5000
Mujer	130	260	650	1034	1600	2600	3500
Nivel Educativo							
Primaria	130	260	563	1000	1425	1840	2165
Secundaria	126	300	779	1100	1559	2122	2598
Superior no universitaria	300	600	950	1400	2000	2577	3260
Superior universitaria	425	800	1200	2000	3347	5000	7000
Grupos de Edad							
De 14 a 24 años	0	178	504	900	1260	1732	2200
De 25 a 44 años	320	637	1000	1500	2096	3100	4200
De 45 y más años	203	400	850	1400	2122	3893	5000

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

4.2 Estimación por regresión cuantílica

En la tabla 1 se aprecia los resultados de las distintas estimaciones planteadas. Entre las columnas se aprecia los resultados para los cuantiles 0.25, 0.5, 0.75. En todas las estimaciones planteadas se determinaron que todas las variables que están involucradas al modelo obtuvieron parámetros estadísticamente significativos. La edad, el total de horas trabajadas a la semana y el nivel educativo influyen positivamente en el ingreso total mensual. Centrándonos en los resultados obtenidos, la variable sexo femenino influye negativamente en el ingreso, es decir su ingreso se reduce aproximadamente de 12% a 13% respecto al ingreso total mensual de los hombres. También la variable nivel educativo resulta relevante, debido a que dicha variable con el ingreso total mensual incrementa con el grupo de personas que perciben sus ingresos mensuales altos (Q75), la población ocupada que tiene un nivel educativo superior universitaria llegan a aumentar un ingreso total mensual en 47% más, que la población ocupada que tiene un nivel educativo primario (referencia), este valor es inferior al grupo de personas que tienen un ingreso total mensual bajo (Q25), quienes llegan a aumentar en un 39.5% más respecto a las personas que tiene un nivel educativo primario. La relación entre el total de horas trabajadas a la semana y el ingreso total mensual tienden a reducirse a medida que nos situamos en segmentos de la población ocupada con ingresos totales mensuales altos pasando de 0.89% a 0.54%, lo contrario resulta con relación a la edad. El incremento de la edad tiene un impacto positivo sobre el ingreso total mensual.

Cuadro 12:

Resultados de la regresión cuantílica(Coeficientes y Errores estándar)

Variable dependiente: Log ingreso total mensual

Cuantil	0.25		0.5		0.75		MCO	
	Coef	Error Estándar						
Intercepto	2.3624	0.0210	2.5842	0.0189	2.7594	0.0180	2.4922	0.0178
Sexo(Femenino)	-0.1259	0.0080	-0.1232	0.0062	-0.1294	0.0065	-0.1369	0.0062
Edad	0.0010	0.0003	0.0020	0.0002	0.0031	0.0002	0.0022	0.0002
T. Horas Trabajadas	0.0089	0.0002	0.0071	0.0002	0.0054	0.0002	0.0081	0.0002
Nivel Educativo(Sec)	0.1233	0.0113	0.0878	0.0118	0.0795	0.0102	0.1127	0.0109

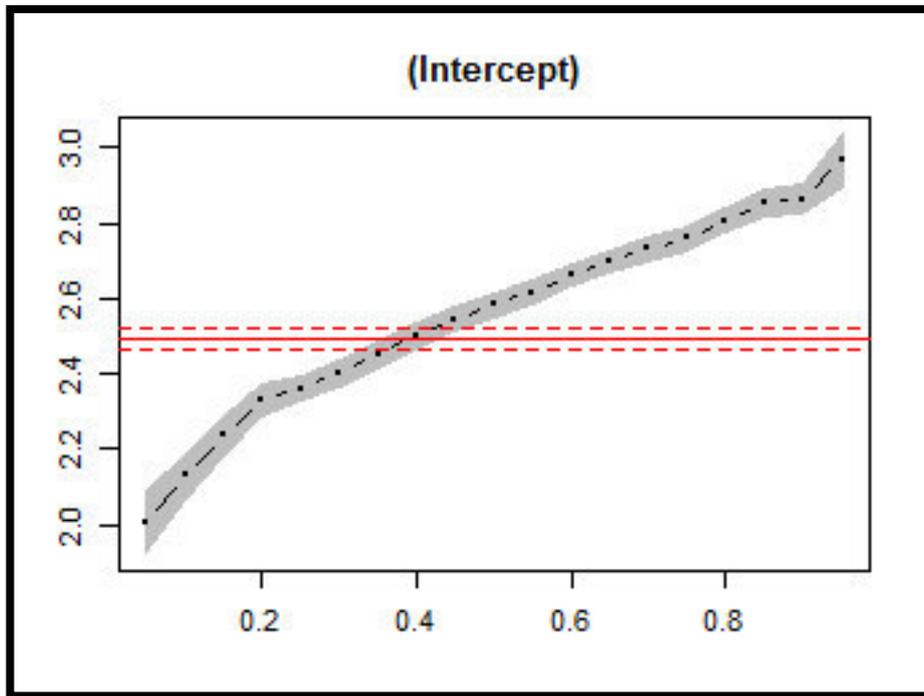
Nivel Educativo (Sup. No Univ.)	0.2257	0.0143	0.1923	0.0131	0.2069	0.0120	0.2376	0.0122
Nivel Educativo (Sup. Univ.)	0.3958	0.0149	0.4188	0.0145	0.4769	0.0130	0.4660	0.0118

Fuente: Elaboración Propia a partir de la Encuesta Permanente de Empleo 2016

Nota: todas las estimaciones son estadísticamente significativas ($p < 0.01$)

En la figura 1 se presenta un 4 covariables más el intercepto. Por ese motivo, siguiendo a Koenker y Hallock (2001) graficamos para cada coeficiente de regresiones por cuantiles para $\lambda = \{0.05, 0.10, 0.15, \dots, 0.95\}$ que en la figura 1 están representados por las líneas de puntos negros. Entonces, para cada variable explicativa, dichos estimadores pueden interpretarse como el impacto que tiene un cambio unitario de dicha variables sobre el log (Ingreso total mensual), manteniendo constante el resto de las variables explicativas. De esta manera, la figura tiene la escala en cuantiles en su eje horizontal y la escala en log (Ingreso Total Mensual) en su eje vertical, que señala el efecto de la correspondiente variable explicativa sobre dicha variable. Los contornos de la nube de puntos sombreada corresponden a los valores inferior y superior de las bandas de confianza del correspondiente estimador de la regresión cuantílica. La línea roja horizontal corresponde al valor del estimador de la media condicional estimado por medio de Mínimos Cuadrados Ordinarios (MCO). Por último, las líneas rojas horizontales con rayas pequeñas corresponden a los límites inferior y superior del intervalo de confianza de dicho estimador de MCO. El nivel de significatividad usado para los intervalos de confianza es de 95%.

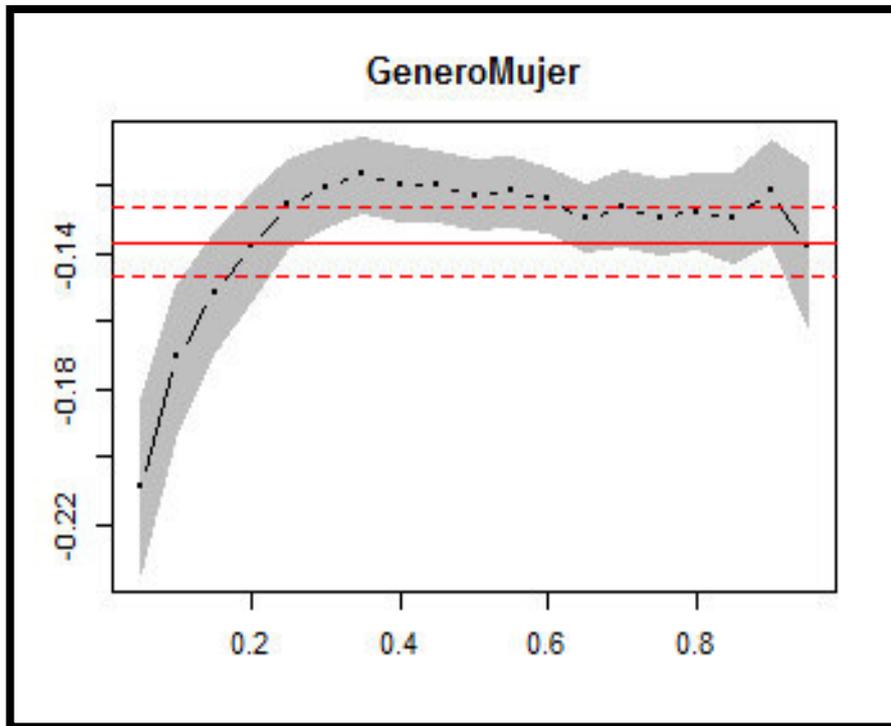
Figura1. Coeficientes de regresiones por cuantiles del intercepto. (Variable dependiente: Log ingreso total mensual)



Fuente: Elaboracion propia realizada en R- Project

En la figura 1 se muestra la evolución del intercepto para los distintos niveles del ingreso total mensual de la población ocupada de Lima Metropolitana de la Figura 1. El intercepto es la constante de regresión, que en nuestro análisis corresponde a una persona cuya edad, total de horas trabajadas a la semana son promedio, y además cuenta con la persona ocupada es de sexo masculino y el nivel educativo es primario. Como era de esperarse, el intercepto evoluciona de manera creciente de izquierda a derecha del gráfico.

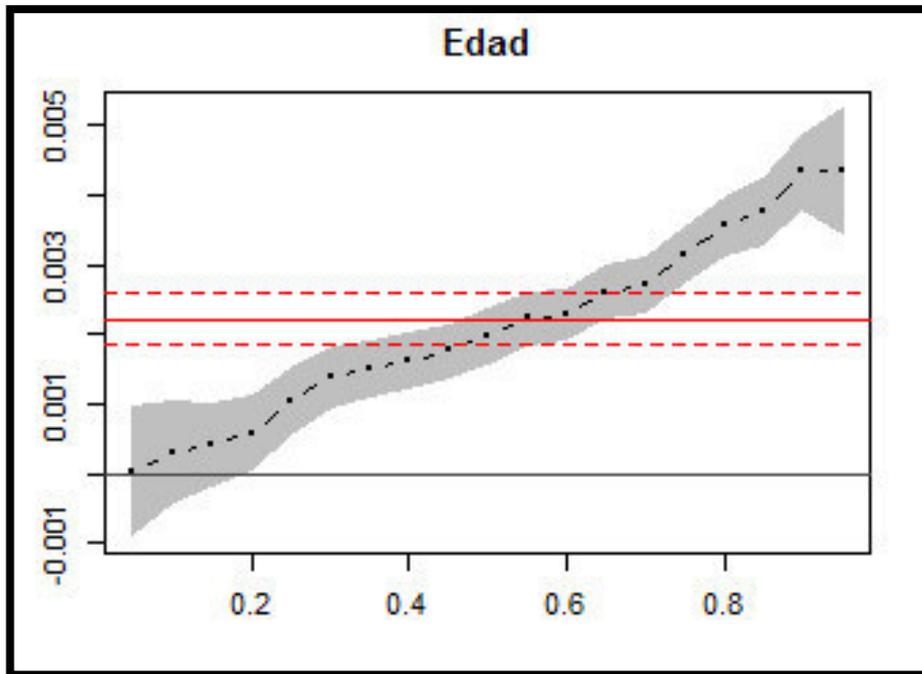
Figura2. Coeficientes de regresiones por cuantiles del variable sexo. (Variable dependiente: Log ingreso total mensual)



Fuente: Elaboracion propia realizada en R- Proyect

En la figura 2 del variable sexo, el sexo femenino, su efecto influye negativamente en el ingreso total mensual aunque con diferente intensidad a partil del cuartil 0.35 tiende a estabilizarse el ingreso total mensual.

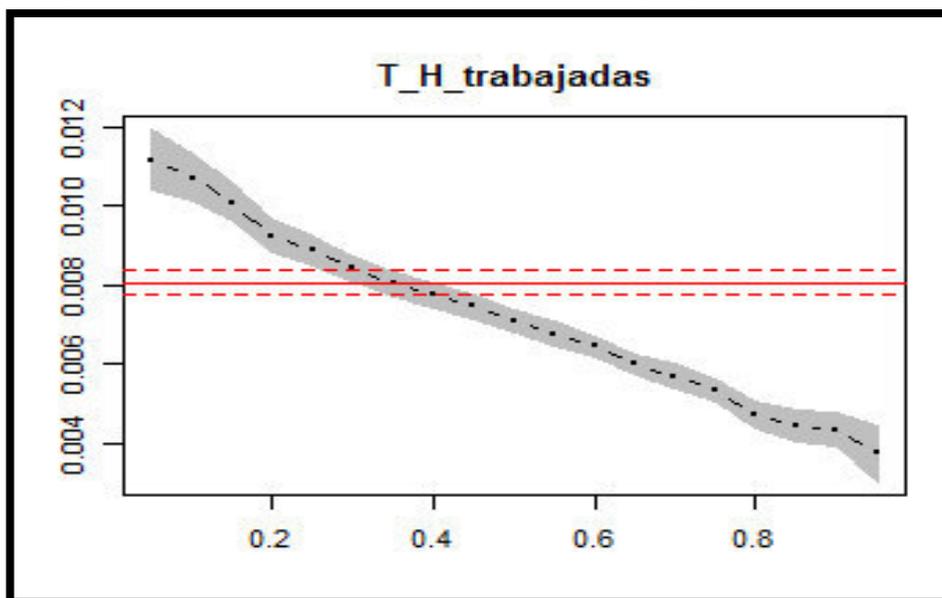
Figura3. Coeficientes de regresiones por cuantiles de la variable edad. (Variable dependiente: Log ingreso total mensual)



Fuente: Elaboracion propia realizada en R- Project

En la figura 3 el panel de la variable edad, su efecto influye positivamente en el ingreso total mensual, se puede visualizar que los segmentos de los ingresos totales mensuales altos es más.

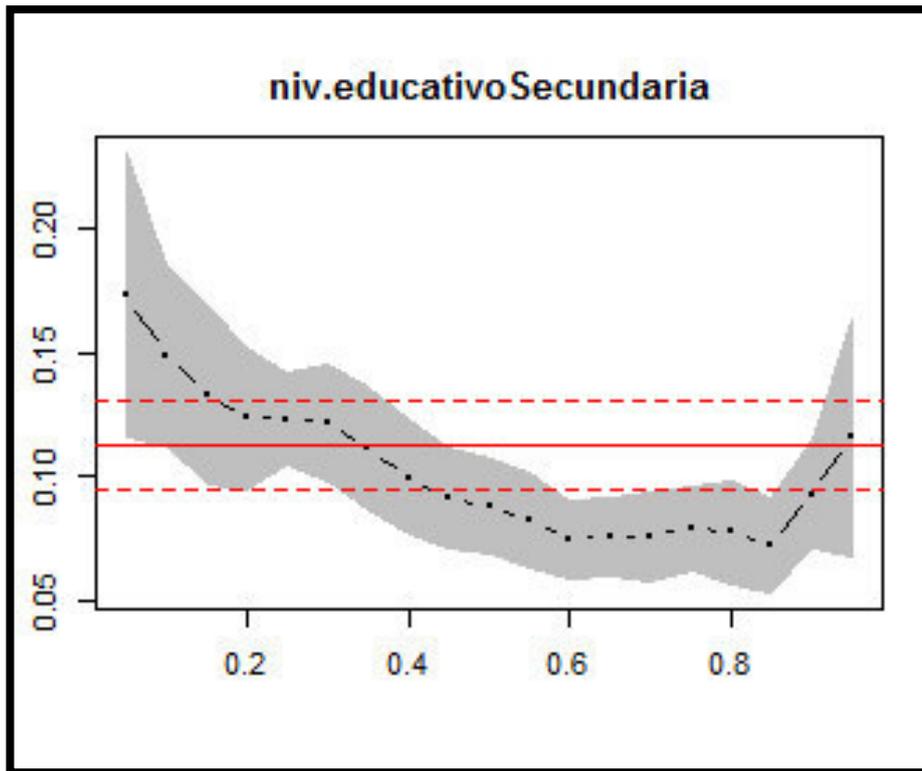
Figura4. Coeficientes de regresiones por cuantiles de la variable total de horas trabajadas a la semana. (Variable dependiente: Log ingreso total mensual)



Fuente: Elaboracion propia realizada en R- Project

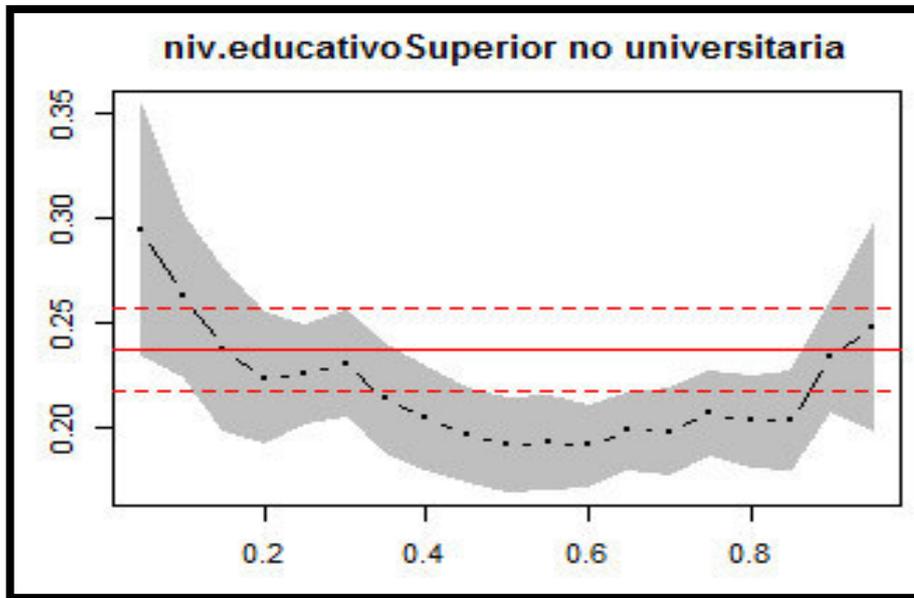
En el figura 4 de la variable total de horas trabajadas, su efecto influye positivamente en el ingreso total mensual, se puede visualizar que los segmentos de los ingresos totales mensuales altos su efecto de influencia va decreciendo esto es debido a que la población ocupada que tiene ingresos altos no necesariamente tiene que trabajar altas horas a la semana.

Figura5. Coeficientes de regresiones por cuantiles de la variable nivel educativo secundaria. (Variable dependiente: Log ingreso total mensual)



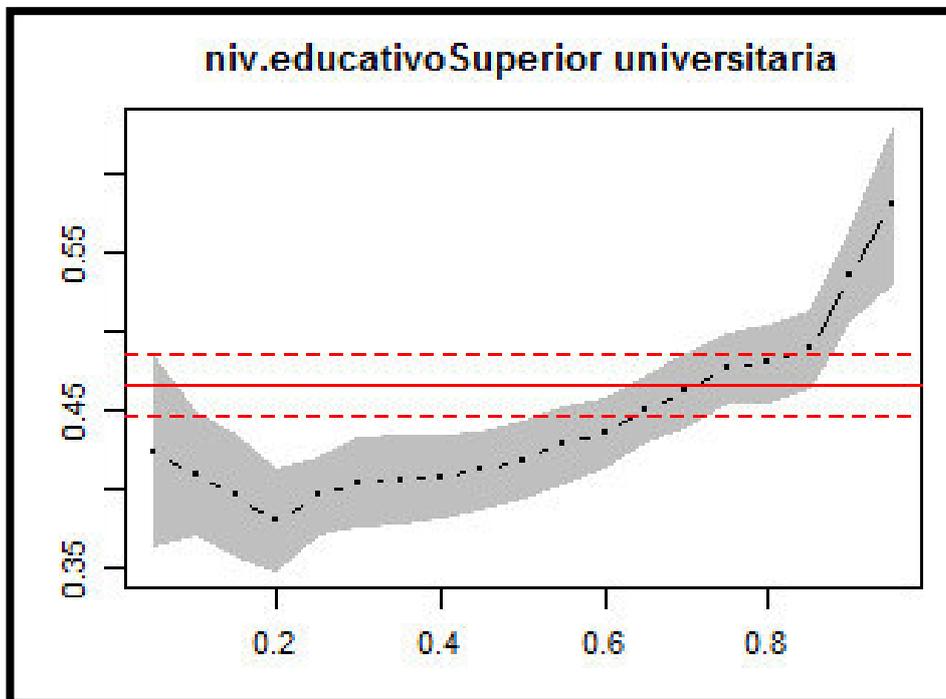
Fuente: Elaboracion propia realizada en R- Proyect

Figura6. Coeficientes de regresiones por cuantiles de la variable nivel educativo superior no universitaria. (Variable dependiente: Log ingreso total mensual)



Fuente: Elaboracion propia realizada en R- Project

Figura7. Coeficientes de regresiones por cuantiles de la variable nivel educativo superior universitaria. (Variable dependiente: Log ingreso total mensual)



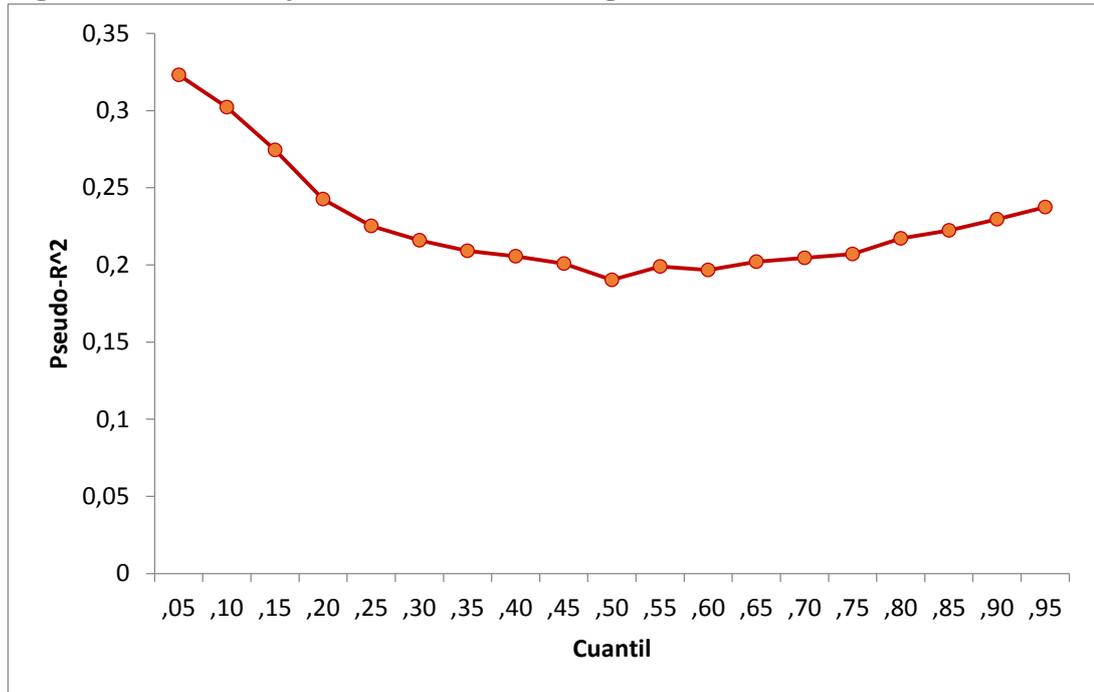
Fuente: Elaboracion propia realizada en R- Project

Desde la figura 5 hasta la figura 7 corresponde al variable nivel educativo. Por ejemplo en la figura 7, su efecto influye positivamente en el ingreso total mensual, se puede visualizar que los segmentos de los ingresos totales mensuales altos son mayores en las personas con nivel educativo superior.

En el anexo, el cuadro 8 muestra las medidas de la bondad de ajuste para el modelo propuesto. Stata provee la medida de bondad de ajuste usando la formula (8) y se refiere a ello “Pseudo-R²” para diferenciar del coeficiente de determinación R² del modelo de regresión lineal.

Cuando se analizan datos con una regresión cuantílica, no existe una estadística equivalente a R-cuadrado. Las estimaciones del modelo a partir de una regresión cuantílica son estimaciones realizadas a través de un proceso iterativo. No se calculan para minimizar la varianza, por lo que el enfoque MCO para la bondad de ajuste no se aplica en este modelo. Sin embargo, para evaluar la bondad de ajuste de los modelos de regresión cuantílica, se han desarrollado el Pseudo-R-cuadrado. Estos son "pseudo" R-cuadrados porque parecen R-cuadrado en el sentido de que están en una escala similar, que van de 0 a 1 (aunque existen algunos pseudo R-cuadrados nunca alcanzan 0 o 1) con valores más altos que indican mejor modelo Pero no pueden ser interpretados como uno interpretaría un R-cuadrado de MCO.

Figura8. Bondad de Ajuste del Modelo de la Regresión Cuantílica



Fuente: Elaboracion propia realizada en STATA

En la Figura 7 muestra la bondad de ajuste para el modelo propuesto, observamos que el Pseudo-R² es bajo en el cuantil 50 (mediana), sin embargo notamos que la bondad de ajuste en el cuantil 0,05 y 0,95 de la variable respuesta son altos en comparación a los otros cuantiles, debido a esto se evidencia como la regresión cuantílica es más flexible para explicar en los extremos de la variable respuesta.

CONCLUSIÓN

Las estimaciones por regresión cuantílica permiten analizar el comportamiento del ingreso total mensual ante cambios de covariables. Los resultados permiten observar como varía el efecto de las covariables para los diferentes cuantiles. Se concluye la importancia que tiene el nivel educativo en el ingreso total mensual, es decir la población económicamente activa ocupada de Lima Metropolitana y Callao, que tienen un nivel educativo superior en general perciben mejores ingresos; y las mujeres reciben en promedio ingresos totales mensuales inferiores a los varones. También se concluye que la población ocupada de lima metropolitana que perciben ingresos altos no necesariamente necesita tener altas horas trabajadas a la semana para que sus ingresos sean mayores.

BIBLIOGRAFÍA

- Barrodale, I., & Roberts, F. (1973). *An improved algorithm for discrete l_1 linear approximation*. SIAM Journal of Numerical Analysis.
- Budig, M., & Hodges, M. (2010). *Diferences and Disavantage: Variation in the Motherhood Penalty across White Women's Earnings Distribution*. American Sociological Review.
- Edgeworth, F. Y. (1888). *The Mathematical Theory of Banking*. Journal of the Royal Statistical Society.
- Hao, L., & Naiman, D. (2007). *Quantile Regression* (Vol. 149). Sage.
- Instituto Nacional de Estadística e Informática. (Febrero de 2000). *METODOLOGIA PARA EL CÁLCULO DE LOS NIVELES DE EMPLEO*. Lima: Centro de Documentación del INEI.
- Instituto Nacional de Estadística e Informática. (2016). *Encuesta Permanente de Empleo en Lima Metropolitana*. INEI, Lima. Lima: Centro Documentación del INEI.
- Instituto Nacional de Estadística e Informática. (2016). *Situación del Mercado Laboral en Lima Metropolitana*. Lima: Centro de Documentación del INEI.
- John, O., & Nduka, E. C. (2009). *Quantile regression analysis as a robust alternative to ordinary least squares* (Vol. 8). Sci. Afr.
- Koenker, R. (2005). *Quantile Regression* (Vol. 38). Cambridge university press.
- Koenker, R. (2010). *quantreg: Quantile Regression*. R package version 4.50.
- Koenker, R., & Bassett, G. (1978). *Regression Quantiles* (Vol. 46). Econometrica.
- Koenker, R., & D'Orey, V. (1994). *Computing Regression Quantiles*. USA.
- Koenker, R., & Hallock, K. (2001). *Quantile Regression* (Vol. 15). Journal of Economic Perspectives.
- Koenker, R., & Portnoy. (1997). *The Gaussian Hare and the Laplacian Tortoise: Computability of Squared Versus Absolute-Error Estimators* (Vol. 12). Statistical Science.
- Ministerio de Trabajo y Promoción del Empleo. (2006). *Análisis de la Distribución del Ingreso Laboral en Lima Metropolitana, 1990-2004*. Boletín de Economía Laboral, Lima.
- Orlandoni Merli, G., Ramoni Perazzi, J., & Pérez Pulidol, M. (2015). *La Distribución del Ingreso Laboral de los Trabajadores de Colombia*. Armenia - Colombia.

ANEXOS

Cuadro 8:

Bondad de Ajuste del modelo de la regresión cuantílica del ingreso total mensual de la población ocupada de Lima Metropolitana 2016

Cuantil	Pseudo-R ²
0.05	0.3231
0.1	0.3023
0.15	0.2745
0.2	0.2427
0.25	0.2252
0.3	0.216
0.35	0.2092
0.4	0.2057
0.45	0.2009
0.5	0.1903
0.55	0.1989
0.6	0.1967
0.65	0.2021
0.7	0.2046
0.75	0.2071
0.8	0.2172
0.85	0.2223
0.9	0.2296
0.95	0.2375

Elaboración Propia realizado con el Software Stata

Programación en R-Proyect

```
library(foreign, pos=14)
```

```
install.packages("quantreg")
```

```
library(quantreg)
```

```
dataempleo= read.spss("D:/Curso de Actualización/TESINA/Base de  
Datos/EPE2016/IngMenExcluyprac2500noremun5variables.sav",
```

```
use.value.labels=TRUE, max.value.labels=Inf, to.data.frame=TRUE)
```

```

dataempleo<-read.spss("D:/Estadística/Curso de Actualización/TESINA/Base de
Datos/EPE2016/IngMenExcluyprac2500noremun5variables.sav",
use.value.labels=TRUE, max.value.labels=Inf, to.data.frame=TRUE)

names(dataempleo)

names(dataempleo)=c("Sexo","Edad","T_H_trabajadas","ingtot","fa_jas16","nivel_alcanzad
o",
"log_ingtot","logfinal")
attach(dataempleo)

dataempleo$Genero=factor(dataempleo$Sexo)
is.factor(dataempleo$Genero)

dataempleo$niv.educativo=factor(dataempleo$nivel_alcanzado)
is.factor(dataempleo$niv.educativo)

attach(dataempleo)
detach(dataempleo)

write.dta(dataempleo,file="datanueva.dta",version=11L)

##### Regresión Cuantílica#####

quantreg2575 <- rq(log_ingtot ~ Genero+Edad+T_H_trabajadas+niv.educativo+Genero,
data = dataempleo, tau=c(0.25, 0.5,0.75))
summary(quantreg2575)

quantreg.all <- rq(log_ingtot ~ Genero+Edad+T_H_trabajadas+niv.educativo+Genero,
tau = seq(0.05, 0.95, by = 0.05), data=dataempleo)
quantreg.plot <- summary(quantreg.all)
plot(quantreg.plot)

```

Programación en STATA

```
ssc install usespss
usespss using "IngMenExcluyprac2500noremun5variables.sav"

qreg log_ingt p107 p108 p209t nivel_al, quantile(25,50,75)

gen primaria=0
replace primaria=1 if nivel_al==1

gen secundaria=0
replace secundaria=1 if nivel_al==2

gen sup_no_univ=0

gen sup_univ=0
replace sup_univ=1 if nivel_al==4

qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(5)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(10)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(15)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(20)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(25)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(30)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(35)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(40)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(45)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(50)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(55)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(60)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(65)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(70)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(75)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(80)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(85)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(90)
qreg log_ingt p107 p108 p209t primaria secundaria sup_no_univ sup_univ, quantile(95)
```