

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

TRABAJO FINAL DE LA MATERIA INTEGRADORA

**“Análisis de la estructura en la distribución
de los ingresos de los ecuatorianos usando
modelo de regresión cuantílica robusta”**

INFORME DE PROYECTO INTEGRADOR

Previa a la obtención del Título de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

Presentado por:

WILSON ABDÓN ORDÓÑEZ RAMOS

GUAYAQUIL – ECUADOR

AÑO: 2017

AGRADECIMIENTOS

“Agradezco en primer lugar a mis padres Wilson Ordóñez Vicuña y Marilyn Ramos Grande por su apoyo de toda las etapas de mi vida, y también agradezco a mis compañeros y profesores de la Universidad por su apoyo por mi pasó en la ESPOL”.

DECLARACIÓN EXPRESA

"La responsabilidad del contenido desarrollado en la presente propuesta de la materia integradora corresponde exclusivamente a Wilson Abdón Ordóñez Ramos y el patrimonio intelectual del mismo a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL"

.....
Wilson Abdón Ordóñez Ramos

RESUMEN

En este trabajo se describe e infiere la estructura de la distribución de los ingresos que obtiene los ecuatorianos por su trabajo ajustados a la horas de trabajo y a la inflación para representar mejor su poder adquisitivo que recibe debido a las horas dedicado a su actividad laboral, en este estudio se destaca que ha habido un aumentos significativo de los ingresos desde el 2007 hasta el 2014, en la desigualdad medidos en ratios cuantílicos ha disminuido significativamente desde 2007 hasta el 2014 y aumentado desde el 2015 al 2016. En el lado de la distribución por sexo, se observa que existe una brechas entre los ingresos a favor del hombre y esta brecha es mayor entre las personas que recibe los menores ingresos por sus horas de trabajos en comparación aquellos que recibe un mayor ingresos, además se observa que en la zona geográfica Costa la desigualdad es mayor en las áreas urbanas que rurales, hecho que no se observa en otras zonas geográficas, aunque esta diferencia ha ido disminuyendo con los años y por último tenemos que aquellos con educación superior recibe mayor ingresos por su trabajo en comparación a aquellos que no tiene este nivel de instrucción.

Palabras clave: ingresos, ratios de desigualdad, regresión cuantilica robusta, distribución, trabajo, INEC, ENEMDU. Ecuador.

ABSTRACT

This paper describes and infers the structure of the income distribution that Ecuadorians obtain by their work adjusted to the hours of work and inflation to get better representation of their purchasing power that receive due to the hours dedicated to their work activity, in This study highlights that there has been a significant increase in income from 2007 to 2014, the inequality measured in quantile ratio has declined significantly from 2007 to 2014 and increased from 2015 to 2016. On the distribution side by Sex, it is observed that there is a gap between income in favor of men and this gap is higher among people who receives the lowest income for their work hours compared to those who receive a higher income, in addition it is observed that in the geographical area Coastal, the inequality is greater in urban than rural areas, this phenomenon is not observed in other geographical areas, although this difference has been declining over the years and lastly it is observed that those with higher education receive higher income for their work compared to those who do not have this level of education.

Key words: income, ratios of inequality, robust quantile regression, distribution, labor, INEC, ENEMDU. Ecuador.

ÍNDICE GENERAL

AGRADECIMIENTOS	ii
DECLARACIÓN EXPRESA	iii
RESUMEN	iv
ABSTRACT	iv
ÍNDICE GENERAL.....	vi
ABREVIATURAS	viii
ÍNDICE DE TABLAS Y FIGURAS	ix
CAPÍTULO 1	1
1. INTRODUCCIÓN.....	1
1.1 Descripción del problema.	1
1.2 Objetivos del Proyecto.....	1
Objetivos Generales	1
Objetivos específicos	1
1.3 Justificación	2
CAPÍTULO 2	3
2. METODOLOGÍA.....	3
2.1 Origen y Estructuración de los datos	3
2.2 Ajustes a los valores de los Ingresos.....	4
2.2.1 Ajuste a la Inflación.....	4
2.2.2 Ajuste a la Estacionalidad.....	4
2.3 Regresión cuantílica	5
2.4 Modelo Robusto para la regresión cuantílica usando clase generalizada de distribuciones asimétricas.....	5
2.5 Familias de Distribuciones Asimétricas	6
2.5.1 Definición de la Normal Asimétrica	6
2.5.2 Definición de la Familia de Distribuciones Asimétrica.....	7
2.6 Estimación de los parámetros del modelo de regresión cuantílica robusta usando distribuciones asimétricas.	7
2.7 Estimación de la variancia de los parámetros de una regresión usando errores con distribuciones asimétricas.....	9
2.8 Estimación de los parámetros con los datos ponderados por observaciones.....	10
2.9 Análisis de los ingresos y desigualdad con regresión cuantílica...	11

CAPÍTULO 3	14
3. ANÁLISIS DESCRIPTIVO	14
3.1 Distribución e Histograma de la variable $lh=\log(\text{INGH})$	14
3.2 Descripción General de los cuantiles por año.....	15
3.3 Descripción General de los cuantiles por año y Sexo.....	15
3.4 Descripción General de los cuantiles por Año y Áreas	16
3.5 Descripción General de los cuantiles por Año y Zonas	17
3.6 Descripción General de los cuantiles por Año y Educación.....	18
3.7 Descripción General de los cuantiles por Año y Años de Experiencia	19
3.8 Cuantiles por Áreas, Zona, Sexo y Años de la Encuesta	20
3.9 Cúantiles por Nivel de Instrucción, Área, Sexo y Años de la Encuesta	23
3.10 Cúantiles por Nivel de Instrucción, Años de ocupación, Área, Sexo y Años de la Encuesta	23
CAPÍTULO 4	25
4. INFERENCIA CON MODELOS ESTADÍSTICO	25
4.1 Modelo 1: Ingresos con respecto al Tiempo (INGH~ANIO)	25
4.2 Modelo 2: Ingresos con respecto al tiempo, Zona geográfica y Área rural o urbana (INGH~ANIO*ZONA*AREA)	28
4.3 Modelo 3: Ingresos con respecto al Tiempo, Zona, Área y Sexo (INGH~ANIO*ZONA*AREA*SEXO).....	31
CONCLUSIONES	33
5. CONCLUSIONES	33
RECOMENDACIONES	33
6. RECOMENDACIONES.....	33
BIBLIOGRAFÍA	34
APÉNDICES	35
APÉNDICE A	36
Archivos en el CD	36

ABREVIATURAS

ENEMDU	Encuesta Nacional de Empleo, Desempleo y Subempleo
INEC	Instituto Nacional de Estadística y Censos
IPC	Índice del Precio al Consumidor
v.a.	Variable Aleatoria

ÍNDICE DE TABLAS Y FIGURAS

Tabla 2.1: Diccionario de Variables.....	3
Tabla 2.2: κ , $hu v$ y función de densidad de algunos de los miembros de la familia asimétrica.....	7
Tabla 2.3: Distribución Condicional de U dada Y por cada distribución específica de SKD.....	9
Tabla 3.1 Valores de los parámetros	15
Tabla 4.1 Valores de los parámetros	25
Tabla 4.2 Estimación de los ingresos por Cuantiles y años	25
Tabla 4.3 Valores de los parámetros	28
Tabla 4.4 Valores de los parámetros	29
Tabla 4.5 Valores de los parámetros	31
Tabla 4.6 Valores de los parámetros	31
Figura 3.1 Histograma y Ajuste de Distribución de la variable Ingresos en escala logarítmica.....	14
Figura 3.2 Ingresos por Cuantiles y Años	15
Figura 3.3 Ratios de Desigualdad por Años.....	15
Figura 3.4 Ingresos por Cuantiles, Años y Sexo	16
Figura 3.5 Ingresos por Cuantiles, Años y Sexo	16
Figura 3.6 Ingresos por Cuantiles, Años y Áreas	17
Figura 3.7 Ratios de Desigualdad por Años y Áreas.....	17
Figura 3.8 Ingresos por Cuantiles, Años y Zonas	18
Figura 3.9 Ratios de Desigualdad por Años y Zonas	18
Figura 3.10 Ingresos por Cuantiles, Años y Educación.....	19
Figura 3.11 Ratios de Desigualdad por Años y Educación	19
Figura 3.12 Ingresos por Cuantiles, Años y Años de experiencias	20
Figura 3.13 Ratios de Desigualdad por Años y Años de experiencias.....	20
Figura 3.14 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Región	21
Figura 3.15 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Nivel de Instrucción.....	23
Figura 3.16 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Nivel de Instrucción.....	24
Figura 4.1 Porcentaje de cambio en los ingresos por años consecutivos	27
Figura 4.2 Porcentaje de Cambio de ratios entre años consecutivos.	27
Figura 4.3 Porcentaje cambio en los ingresos en los cuantiles debido al área urbana y rural	30
Figura 4.4 Porcentaje de cambio entre los ratios de desigualdad de las Áreas Urbana con respecto al Rural.....	30
Figura 4.5 Proporción de los Ingresos entre Hombre/Mujer.....	32

CAPÍTULO 1

1. INTRODUCCIÓN

1.1 Descripción del problema.

Con el propósito de conocer las causas de las desigualdades en los ingresos de los ecuatorianos y describir la manera que se distribuye los ingresos, aquello para fundamentar los planes de desarrollo del país, es importante conocer y medir los ingresos y la desigualdad entre las regiones geográficas sea de la zona rural o urbana, y los distintos grupos humanos objeto de este estudio, a fin de verificar si existen diferencias significativas entre los ingresos y la desigualdad de estos grupos, y a la vez identificar qué variables podrían afectar al aumento o disminución de los ingresos y la desigualdad.

El proyecto se propone seleccionar y aplicar modelos de regresión cuantílica robusta que permitirán describir la desigualdad y los ingresos de los grupos humanos por diferentes regiones del país sea de la zona rural o urbana. Además se verificara si existen diferencias significativas entre los grupos de diversa edad, educación y sexo.

1.2 Objetivos del Proyecto.

Objetivos Generales

Determinar las variables significativas que afecta la desigualdad de ingresos y los ingresos en sí de los ecuatorianos en la población objetivo, contribuyendo al mejoramiento del plan de desarrollo del país.

Objetivos específicos

- Determinar los modelos de regresión cuantílica seleccionando el más adecuado a la muestra recogida.
- Analizar los ratios de desigualdad en base de los modelos de regresión cuantílica estableciendo las variables significativas.
- Establecer las relaciones más significativas entre las variables describiendo el nivel de desigualdad entre ingresos.

1.3 Justificación

En este proyecto propone usar y seleccionar modelos de **regresión cuantílica robusta** que permitirá describir la desigualdad en ingresos de los diferentes grupos humanos por diferentes ciudades del país y además verificar si existe diferencia significativa entre los grupos de diferentes edad, educación y Género.

CAPÍTULO 2

2. METODOLOGÍA

2.1 Origen y Estructuración de los datos

Los datos fueron obtenidos de las bases de datos generadas por la “Encuesta Nacional de Empleo, Desempleo y Subempleo” (INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016) (ENEMDU) que lo realizan desde el año 2000, para este trabajo se usó los datos del ENEMDU realizados entre el 2007 y el 2016, debido a que las bases de datos de estos años contienen todas las variables que es de nuestro interés estudiar, estas variables se especifica en la Taba #1. Entre 2007 y el 2016 existen 34 archivos “.sav” donde se guarda las bases de datos de los ENEMDU realizados, todos estos archivos fueron depurado y estructurado en uno solo, al final se analizó 478 000 observaciones que corresponde a la población objetico que son las personas que percibe algún tipo de ingreso mensual.

Tabla 2.1: Diccionario de Variables

Variables Usada	Código	Nombres Originales/Formula
Sexo de la Persona	SEXO	En las bases de datos está con el código: SEXO o P02
Edad de la Persona	EDAD	EDAD o P03
Región geográfica donde reside	ZONA	RN
Área Rural o Urbana	AREA	AREA
Hora Total que trabaja a la semana	HORT	=P51A+P51B+P51C Suma de las Horas trabajadas en todas las actividades
Ingresos por cada 160 horas de trabajos	INGH	=INGRL/(HORT)*160

VARIABLES USADA	CÓDIGO	NOMBRES ORIGINALES/FORMULA
		Este resultado es ajustado por medio del IPC al su equivalente para enero del 2016 y ajustado en base a la mediana de cada mes
Nivel máximo de instrucción	EDUC	P03, NIVINST,P10A
Años de experiencia ejerciendo su ocupación principal	ANT1	P45
Factor de Expansión	FACT	FEXP
Año de la Encuesta ENEMDU	ANIO	(En la nombre de las bases de datos Indica)
Mes de la Encuesta	MES	(En el nombre de las bases de datos indica)

Fuente: Elaboración Propia

2.2 Ajustes a los valores de los Ingresos.

2.2.1 Ajuste a la Inflación

El poder adquisitivo que te otorga cierta cantidad de dinero aumenta o disminuye de manera dinámica, debido a que el nivel general de los precios de los bienes, medido con el Índice del Precio al Consumidor (IPC) (INEC, Índice de Precios al Consumidor, 2014), no se mantiene constante y fluctúa a través del tiempo (inflación o deflación). Con el propósito de reducir este efecto en el análisis, se ajustó los valores de los ingresos proporcionalmente al IPC tal como se muestra en la expresión 2.1.

$$\text{Ingresos ajustado a la Inflación} \quad (2.1)$$

$$\text{Ingresos} \frac{IPC_{mes-1,año}}{IPC_{01,2016}}$$

2.2.2 Ajuste a la Estacionalidad

Los meses en que se realizó las encuestas pueden presentar características estacionarias, con el objetivo de reducir este efecto se propone ajustar los Ingresos al mes de diciembre usando proporciones de medianas, tal como se indica en la expresión 2.2

$$\begin{aligned} & \text{Ingresos ajustado a la estacionaridad} \\ & = \text{Ingresos} \frac{\text{Mediana}(\text{Ingresos}_{\text{diciembre}})}{\text{Mediana}(\text{Ingresos}_{\text{mes}})} \end{aligned} \quad (2.2)$$

2.3 Regresión cuantílica

La regresión cuantílica es un tipo de regresión que busca predecir los cuantiles condicionado de una variable aleatoria Y , dado las variables dependientes X (x_1, x_2, x_3, \dots) a través de una relación lineal, tal como se muestra en la expresión 2.3. (Koenker, 2005)

$$Q_p(Y|x_1, x_2, \dots, x_n) = \beta_{0p} + \beta_{1p}x_1 + \beta_{2p}x_2 + \dots + \beta_{np}x_n \quad (2.3)$$

Donde $Q_p(Y|x_1, x_2, \dots, x_n)$ es el cuantil p de la v.a. Y , dado las variables dependientes x_i 's, y $\beta_{0p}, \beta_{1p}, \dots, \beta_{np}$ son los coeficientes de la regresión, las cuales varían según el cuantil p que se desea predecir.

Se puede usar varios modelos estocásticos para estimar los coeficientes lineales de los cuantiles condicionales, en este trabajo se optó por un modelo de Robusto en base de distribuciones asimétrica.

2.4 Modelo Robusto para la regresión cuantílica usando clase generalizada de distribuciones asimétricas.

Este modelo es propuesto en base al trabajo de Wichitaksorn (Wichitaksorn, 2014), donde se utilizan una re-parametrización de una familia de distribuciones asimétricas, así definiendo el parámetro de asimétrica con p (un número entre 0 y 1), esta re-parametrización se realiza de manera que al calcular el cuantil coincidente con el parámetro p de la variable aleatoria es igual a la moda ($Q_p(e) = \text{Moda}(e)$), dicha parametrización facilita aplicar estas distribuciones a una regresión cuantílica.

$$y_i = \beta_0 + x_i^t \beta_p + \varepsilon_{ip} \quad (2.4)$$

ε_{pi} es el error del modelo mostrado de regresión en la expresión 2.4 que son v.a independientes e igualmente distribuida con distribución

asimétrica, parámetro de asimetría p , moda 0, y otros parámetros θ propias de cada distribución asimétrica, en particular ($\varepsilon_{pi} \sim SKN(0, p, \theta)$), con esto tenemos que el cuantil p de ε_{pi} es 0, y el cuantil p la variable y_i dado x_i es igual a la combinación lineal de las variables dependientes x_i , tal como se muestra en la expresión 2.5.

$$Q_p(y_i|x_i) = \beta_0 + x_i^t \beta_p + Q_p(\varepsilon_{ip}) = \beta_0 + x_i^t \beta_p + 0 \quad (2.5)$$

En base de la representación estocástica mostrada en la expresión 2.4 obtenemos una relación lineal del cuantil p de la v.a y_i condicionado por x_i .

2.5 Familias de Distribuciones Asimétricas

En esta subsección se describirá a las distribuciones pertenecientes a la familia asimétrica que se usará en nuestro análisis.

2.5.1 Definición de la Normal Asimétrica

Decimo que X tiene una distribución normal-asimétrica con locación μ , escala $\sigma > 0$ y parámetro de asimetría $p \in (0,1)$, si esta sigue la siguiente función de distribución

$$f(x|\mu, \sigma, p) = \frac{4p(1-p)}{\sqrt{2\pi\sigma^2}} e^{-2\rho_p^2 \left(\frac{x-\mu}{\sigma}\right)} \quad (2.6)$$

Sea \mathbb{I} la función indicador, entonces $\rho_p(u) = u p - \mathbb{I}(u < 0)$.

Una propiedad de la distribución normal-asimétrica es que es cerrada bajo la transformación de escala y locación, y además se puede expresar esta distribución en relación a la distribución normal estandar con la expresión 2.7 (Christian E. Galarza, 2016):

$$X = \mu + \sigma I | Z | \quad (2.7)$$

Donde $Z \sim N(0,1)$ y la v.a. I es discreta y tiene la siguiente función de probabilidad:

$$P\left(I = -\frac{1}{2(1-p)}\right) = p \quad \& \quad P\left(I = \frac{1}{2p}\right) = 1 - p \quad (2.8)$$

2.5.2 Definición de la Familia de Distribuciones Asimétrica

Sea $Z \sim N(0,1)$ y U una variable aleatoria positiva, con U independiente de Z , y I una variable aleatoria discreta tal como se especifica en la expresión 2.8. Entonces la familia de distribuciones asimétrica $Y \sim SKD(\mu, \sigma, p, \nu)$ se define:

$$Y = \mu + [\sigma\kappa(U)^{1/2}]I|Z| \quad (2.9)$$

La función de distribución de la v.a. U con parámetro ν se representa con $h(u|\nu)$, y κ es una función de v.a..

Las funciones $h(\cdot|\nu)$ y κ se definen según la familia de distribución asimétrica que se quiere formar, esto se muestra en la tabla 2. (Christian E. Galarza, 2016)

Tabla 2.2: κ , $h(u|\nu)$ y función de densidad de algunos de los miembros de la familia asimétrica.

Distribución	$\kappa(u)$	$h(u \nu)$	$f(y \mu, \sigma, \nu)$
t de student asimétrica	u^{-1}	$G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$	$\frac{4p(1-p)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}\sigma} \left(\frac{4}{\nu}\rho_p^2\left(\frac{y-\mu}{\sigma}\right) + 1\right)^{-\frac{\nu+1}{2}}$
Laplace Asimétrica	u	$Exp(2)$	$\frac{2p(1-p)}{\sigma} e^{-2\rho_p^2\left(\frac{y-\mu}{\sigma}\right)}$
Slash Asimétrica	u^{-1}	$Beta(\nu, 1)$	$\nu \int_0^1 u^{\nu-1} \phi_{skd}\left(y \mu, u^{-\frac{1}{2}}\sigma, p\right) du$
Normal Contaminada Asimétrica	u^{-1}	$\nu\mathbb{I}(u = \gamma) + (1-\nu)\mathbb{I}(u = 1)$ $0 \leq \nu, \gamma \leq 1$	$\nu\phi_{skd}\left(y \mu, \gamma^{-\frac{1}{2}}\sigma, p\right) + (1-\nu)\phi_{skd}(y \mu, \sigma, p)$

Fuente: Christian E. Galarza, 2016

$G(\alpha, \beta)$ denota la distribución Gamma con parámetro de forma $\alpha > 0$ y parámetro de tasa $\beta > 0$, $Exp(\beta)$ Denota la distribución exponencial con media β , $Beta(\alpha, \beta)$ denota la distribución beta y ϕ_{skd} denota la función de probabilidad de la distribución asimétrica normal definida en la expresión 2.6.

2.6 Estimación de los parámetros del modelo de regresión cuantílica robusta usando distribuciones asimétricas.

Sea el modelo de regresión:

$$Y_i \sim SKD(x_i^t \beta_p, \sigma, p, \nu) \quad (2.10)$$

Donde Y_1, Y_2, Y_3, \dots son independiente e idénticamente distribuida, este modelo puede ser representando en relación a la distribución normal asimétrica condicionándolo a una variable aleatoria \mathbf{U} en base lo mostrado en la expresión 2.9, así deducimos la expresión 2.11.

$$(Y_i | U_i = u_i) \sim SKN \left(x_i^t \beta_p, \sqrt{\kappa(u_i) \sigma}, p \right) \quad (2.11)$$

$$U_i \sim h(u_i | \nu) \quad (2.12)$$

A partir de ahí se puede estimar los parámetros usando algoritmo EM, donde tenemos datos observados obtenidos de \mathbf{Y} y datos no observados de la v.a. \mathbf{U} . Se inicia con una estimación de los parámetros $\theta^{(0)}$ cualquiera.

- Paso E: Se define la función Q .

$$Q(\theta | \theta^{(k)}) \propto -n \log \sigma - 2 \sum_{i=1}^n \kappa^{-1}(\widehat{u}_i) \xi_i^2 z_i^2 + \sum_{i=1}^n E[\log h(u_i | \nu) | y_i, \theta^{(k)}] \quad (2.13)$$

- Paso M: Encontrar $\theta = (\beta_p, \sigma^2, \nu)$ que maximiza la función Q , que resulta ser:

$$\begin{aligned} \widehat{\beta}_p^{(k+1)} &= (\mathbf{X}^T \boldsymbol{\Omega}^{(k)} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{(k)} \mathbf{y} \\ \widehat{\sigma}^2^{(k+1)} &= \frac{4}{n} (\mathbf{y} - \mathbf{X} \widehat{\beta}_p^{(k+1)})^T \boldsymbol{\Omega}^{(k)} (\mathbf{y} - \mathbf{X} \widehat{\beta}_p^{(k+1)}) \end{aligned} \quad (2.14)$$

Donde $\boldsymbol{\Omega}$ es una matriz diagonal, cuyos elementos son

$\xi_i^2 \kappa^{-1}(\widehat{u}_i)$ con $\xi_i = (1 - p) \mathbb{I}\{y_i \leq x_i^t \beta_p\} + (p) \mathbb{I}\{y_i > x_i^t \beta_p\}$, para $i = 1, \dots, n$. \mathbf{X} la matriz de diseño y \mathbf{y} las observaciones.

El parámetro ν es calculado de manera numérica maximizando la función de log-verosimilitud en base de la distribución del error de la regresión.

$$\hat{\nu}^{(k+1)} = \arg \max_{\nu} \sum_{i=1}^n \log f \left(y_i | \hat{\beta}_p^{(k+1)}, \hat{\sigma}^2^{(k+1)}, \nu \right) \quad (2.15)$$

- Iterar el paso M hasta que la función de log-verosimilitud sea el máximo posible.

Las funciones κ^{-1} depende la distribución asimétrica que suponemos para el error.

Tabla 2.3: Distribución Condicional de U dada Y por cada distribución específica de SKD.

Distribución	Distribución de U	Distribución Condicional de $U Y$	$\kappa^{-1}(\hat{u}_i)$
t de student asimétrica	$G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$	$G\left(\frac{\nu+1}{2}, \frac{\nu+4\xi_i^2 z_i^2}{2}\right)$	$\frac{\nu+1}{\nu+4\xi_i^2 z_i^2}$
Laplace Asimétrica	$Exp(2)$	$GIG\left(\frac{1}{2}, 2\xi_i^2 z_i^2, \frac{1}{2}\right)$	$\frac{1}{2\xi_i z_i }$
Slash Asimétrica	$Beta(\nu, 1)$	$TG\left(\nu + \frac{1}{2}, 2\xi_i^2 z_i^2, 1\right)$	$\left[\frac{\nu + \frac{1}{2}}{2\xi_i^2 z_i^2}\right] \frac{\mathcal{F}\left(1 \nu + \frac{3}{2}, 2\xi_i^2 z_i^2\right)}{\mathcal{F}\left(1 \nu + \frac{1}{2}, 2\xi_i^2 z_i^2\right)}$
Normal Contaminada Asimétrica	$\nu \mathbb{I}(u = \gamma) + (1 - \nu) \mathbb{I}(u = 1)$ $0 \leq \nu, \gamma \leq 1$	$\frac{a \mathbb{I}(u = \gamma) + b \mathbb{I}(u = 1)}{a + b}$	$\frac{a\gamma + b}{a + b}$

Fuente: Christian E. Galarza, 2016

En la tabla 3, $\mathcal{F}(x|\alpha, 1/\beta)$ es la función de acumulación de la distribución $Gamma(\alpha, 1/\beta)$. Además las expresiones a y b están dada por $a = \nu \phi\left(y_i | x_i^t \beta_p, \frac{\gamma^{-1} \sigma^2}{4\xi_i^2}\right)$ y $b = (1 - \nu) \phi\left(y_i | x_i^t \beta_p, \frac{\sigma^2}{4\xi_i^2}\right)$. La noatación $TG(a, b, t)$ representa la variable aleatoria con distribución $Gamma(a, 1/b)$ truncado hacia la derecha hasta el valor t . $GIG(\nu, a, b)$ denotan la distribución Inversa Gaussiana Generalizada [Barndorff-Nielsen & Shephard (2001)], cuya función de densidad es dada por:

$$f(x|\nu, a, b) = \frac{(b/a)^\nu}{2K_\nu(ab)} x^{\nu-1} e^{-\frac{1}{2}\left(\frac{a^2+b^2}{x} + x\right)}, x > 0, \nu \in \mathbb{R}, a, b > 0 \quad (2.16)$$

Con $K_\nu(\cdot)$ Siendo la función modificada de Bessel de tercera clase.

2.7 Estimación de la variancia de los parámetros de una regresión usando errores con distribuciones asimétricas.

La estimación de la matriz varianza se obtiene con la inversa de la matriz de información del modelo, a continuación se muestra la fórmula para la matriz de varianza de los estimadores:

$$S = \begin{bmatrix} \frac{2}{\sigma} WX & \frac{2}{\sigma} W(z) - 1 \end{bmatrix} \quad (2.17)$$

$$I = S^T S$$

$$Var(\theta|y) = (I)^{-1}$$

Donde z es el error de la regresión, I la matriz de información, W es una matriz diagonal con elementos $z_i \xi_i^2 \kappa^{-1}(\widehat{u}_i)$, el factor $\xi_i^2 \kappa^{-1}(\widehat{u}_i)$ ha sido calculado para la matriz Ω de la última iteración del paso M del algoritmo EM. (Christian E. Galarza, 2016)

2.8 Estimación de los parámetros con los datos ponderados por observaciones.

Existe ocasiones que los datos se pondera por observaciones, por esta razón se usará la aproximación de Kish para calcular el tamaño de la muestra efectiva y los grados de libertad para la varianza (Koenker, 2005).

$$n = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2} \quad (2.18)$$

Donde w_i es la ponderación asignada por observación, n el tamaño de muestra efectiva.

En base al tamaño de muestra efectiva, se ajusta las ponderaciones de manera que su suma de igual a dicho tamaño.

$$w'_i = n \frac{w_i}{\sum_{i=1}^N w_i} \quad (2.19)$$

$$\sum_{i=1}^N w'_i = n$$

Donde w'_i son las ponderaciones ajustadas al tamaño de muestra efectiva n .

Con las nuevas ponderaciones w'_i se procede a estimar los parámetros de la regresión de manera parecida a la mostrada en la sección 2.7, lo

que cambia es la matriz diagonal Ω en la expresión 2.14, la matriz diagonal W en la expresión 2.17 y la función de verosimilitud mostrada en el 2.15 de la manera que se muestra en las expresiones 2.20, 2.21, 2.22 respectivamente. A parte que el tamaño de la muestra n se reemplaza con el tamaño de la muestra efectiva calculado en la expresión 2.18.

$$\Omega = \text{diag}[w_i' \xi_i^2 \kappa^{-1}(\widehat{u}_i)] \quad (2.20)$$

$$W = \text{diag}[w_i' z_i \xi_i^2 \kappa^{-1}(\widehat{u}_i)] \quad (2.21)$$

$$\widehat{\mathbf{v}}^{(k+1)} = \arg \max_{\mathbf{v}} \sum_{i=1}^n w_i' \log f(y_i | \widehat{\beta}_p^{(k+1)}, \widehat{\sigma}^2^{(k+1)}, \mathbf{v}) \quad (2.22)$$

2.9 Análisis de los ingresos y desigualdad con regresión cuantílica.

En el análisis de los ingresos mensuales es importante el comportamiento de los diferentes cuantiles de ingresos, pues suele representar diferentes clases sociales y por lo tanto podría tener comportamientos diferenciados, por esta razón es idóneo analizar los ingresos con modelo de regresión cuantílica pues permite una interpretación intuitiva. Para el modelo cuantílico se supondrá el logaritmo de base 10 de los ingresos como variables de respuesta, $ly = \log_{10} y$ (Donde y es la variable “ingresos”)

$$Q_p[ly] = X\beta + \varepsilon_{ip} \quad (2.23)$$

Con la estimación de los coeficientes del modelo cuantílico, junto a sus varianzas, se podría realizar varias inferencias sobre la importancia de una variable o combinación lineal de variables para los ingresos en diferentes cuantiles, esto es posible suponiendo los coeficientes de la regresión sigue una distribución normal multivariada.

$$\beta_p \sim N(\hat{\beta}_p, \Sigma_{\hat{\beta}}^p) \quad (2.24)$$

$$K = C\beta_p \sim N(C\hat{\beta}_p, S = C\Sigma_{\hat{\beta}}^p C')$$

Donde C es una matriz que representa combinación lineal de los coeficientes.

También supondremos que las estimaciones de coeficientes para distintos cuantiles son independientes entre sí.

$$Var(\beta_p, \beta_q) = 0 \quad (2.25)$$

Con esta suposición se podrá realizar inferencias sobre las diferencias de los ingresos entre los cuantiles.

$$Q_p[ly] - Q_q[ly] = \log_{10} Q_p[ly] - \log_{10} Q_q[ly] = \log_{10} \frac{Q_p[ly]}{Q_q[ly]} = \log_{10} r_{p:q} \quad (2.26)$$

$$\log_{10} r_{p:q} = X(\beta_p - \beta_q) + (\varepsilon_{ip} - \varepsilon_{iq})$$

$$(\beta_p - \beta_q) \sim N(\hat{\beta}_p - \hat{\beta}_q, \Sigma_{\hat{\beta}}^p + \Sigma_{\hat{\beta}}^q)$$

Como se muestra en la expresión 2.26, a ser ly logaritmo de ingresos y el logaritmo una función creciente, tenemos que la diferencia entre los cuantiles de los logaritmos de ingresos es igual al logaritmo del ratio entre estos cuantiles, y este ratio $r_{p:q}$ mide la proporción de ingresos recibe los de los cuantiles superiores con respecto a los de inferiores, y se interpreta como una medida de desigualdad, en este trabajo se usará los ratios 90:10, 90:50, 50:10.

Para verificar si una variable o valores de coeficientes es influyente o sea distinto a 0 en el modelo, se usa las pruebas de hipótesis estadísticas, donde se establece como hipótesis nula que los valores estimados de los coeficientes sigue una distribución normal con media 0, esta hipótesis se rechaza cuando la probabilidad de obtener la estimación que se obtuvo suponiendo una normal con media 0 es menor a 0.05 (Nivel de Significancia).

$$P\left(z \geq |k_i| \mid \mu = 0, \sigma = \sqrt{S_k^{(i,i)}}\right) \leq 0.05 \quad (2.24)$$

En el caso multivariado, la hipótesis nula es que los coeficientes sigue una distribución normal multivariada con medias 0, el estadístico de prueba es la distancia de mahalanobis $d = \sqrt{\beta' \Sigma^{-1} \beta}$, si suponemos a β como una normal multivariada con media 0 y matriz de covarianzas Σ , entonces la distancia estimada d sigue una distribución Chi con grados de libertad igual al número de elementos en β , por lo tanto se calcula la probabilidad de obtener una distancia mayor a la que se estimó, si esta probabilidad es menor a 0.05, se rechaza la hipótesis nula de que los coeficientes sea igual 0.

$$P\left(D \geq \sqrt{\hat{\beta}' \Sigma^{-1} \hat{\beta}} \mid g.l = N. \text{Elementos en } \beta\right) \leq 0.05 \quad (2.24)$$

CAPÍTULO 3

3. ANÁLISIS DESCRIPTIVO

Con el objetivo de conocer cómo se comporta y evolucionan los ingresos en sus cuantiles a través de los años y los diferentes grupos, se calcula de manera empírica estos mismos para cada grupo y años.

3.1 Distribución e Histograma de la variable $l_h = \log(\text{INGH})$

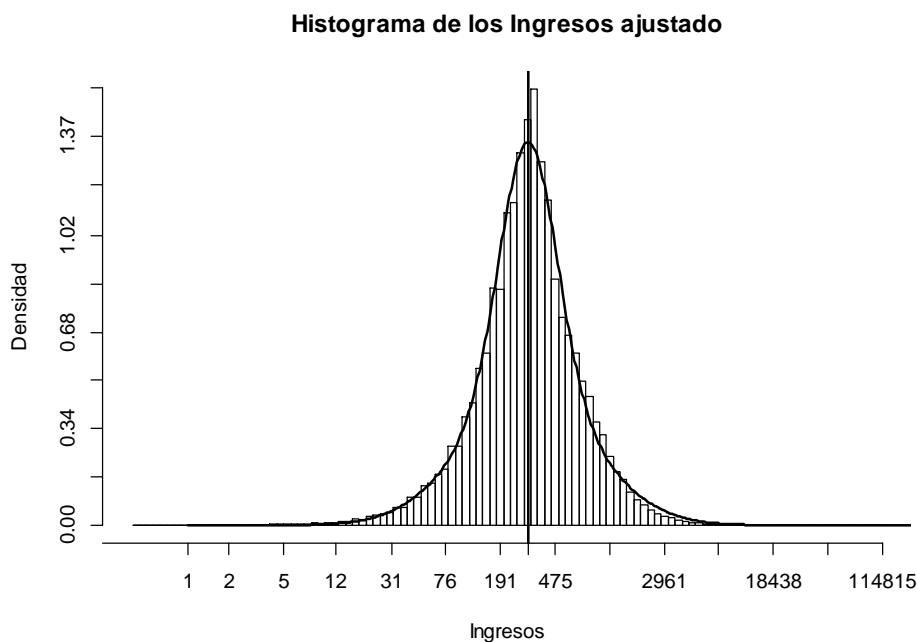


Figura 3.1 Histograma y Ajuste de Distribución de la variable Ingresos en escala logarítmica

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

Vemos en la figura 3.1 que la distribución de los logaritmos de los ingresos sigue una distribución normal contaminada con los siguientes parámetros.

$$ly = \log_{10} y \sim NC(\mu, \sigma, \gamma, \mu) \quad (3.1)$$

Y función de densidad:

$$f(y) = v\phi\left(y|\mu, \gamma^{-\frac{1}{2}}\sigma\right) + (1-v)\phi(y|\mu, \sigma) \quad (3.2)$$

En la tabla 3.1 se muestra los valores de los parámetros.

PARÁMETROS	VALORES	10^VALORES
------------	---------	------------

ν	0,5357	
γ	0,1886	$1/\gamma = 5.30$
σ	0,2060	
μ	2,4797	$10^\mu = 301,82$

Tabla 3.1 Valores de los parámetros

3.2 Descripción General de los cuantiles por año

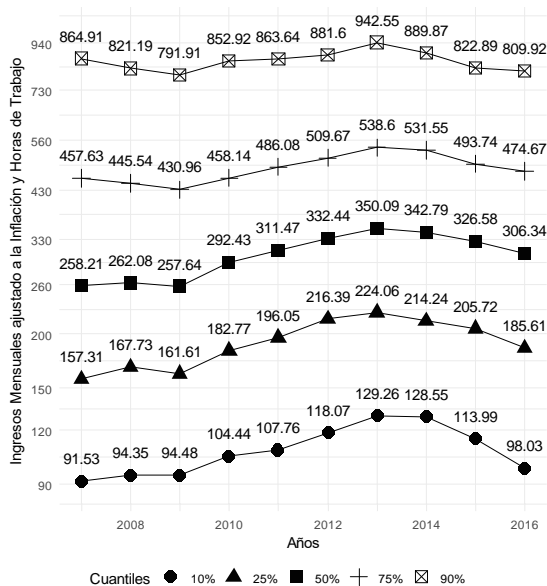


Figura 3.2 Ingresos por Cuantiles y Años

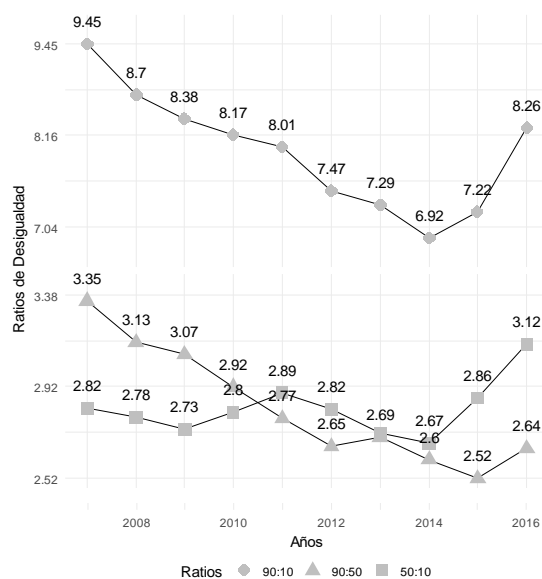


Figura 3.3 Ratios de Desigualdad por Años

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

Interpretaciones

- En la figura 3.2 se observa un aumento general de los Ingresos de los Ecuatorianos, en todos los cuantiles hasta el 2013 y desde el 2014 ha disminuido los ingresos.
- En la figura 3.3 se observa que el ratio de desigualdad entre los 10% más rico y los 10% más pobres (ratio 90:10), hasta el 2014, desde ahí ha aumentado este ratio, y lo mismo sucede si comparamos el ratio de desigualdad entre quienes tienen ingreso mediano con respecto a los 10% más pobres (ratio 50:10).

3.3 Descripción General de los cuantiles por año y Sexo

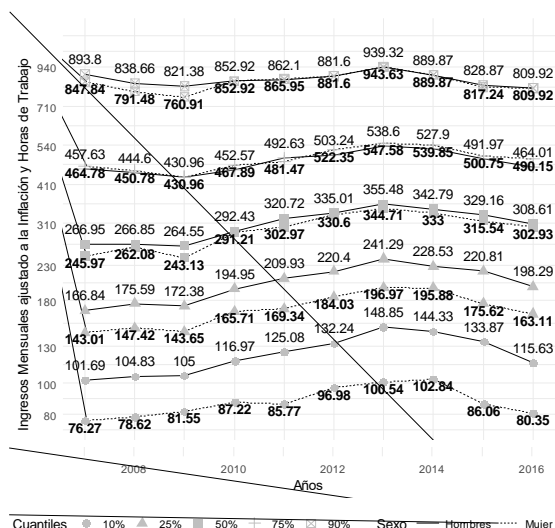


Figura 3.4 Ingresos por Cuantiles, Años y Sexo

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

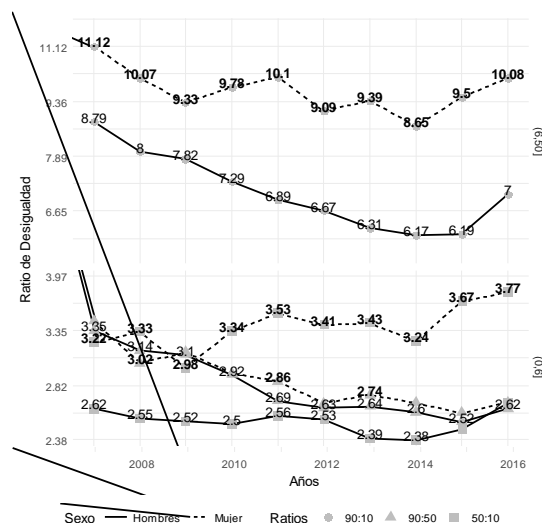


Figura 3.5 Ingresos por Cuantiles, Años y Sexo

Interpretaciones

- En la figura 3.4 se observa que los Hombres gana más que las mujeres en los cuantiles del 10%, pero esta brecha disminuye proporcionalmente para los cuantiles altos.
- En la figura 3.5 se observa que el ratio de desigualdad 90:10 entre mujeres es mucho mayor que entre hombres, el ratio con respecto a la mediana 90:50 se comporta igual entre hombres y mujeres, pero existe una gran diferencia en el ratio 50:10, donde se observa que la mujer con ingreso mediana gana aproximadamente 1.2 veces más que el 10% de las mujeres con menos ingresos.

3.4 Descripción General de los cuantiles por Año y Áreas

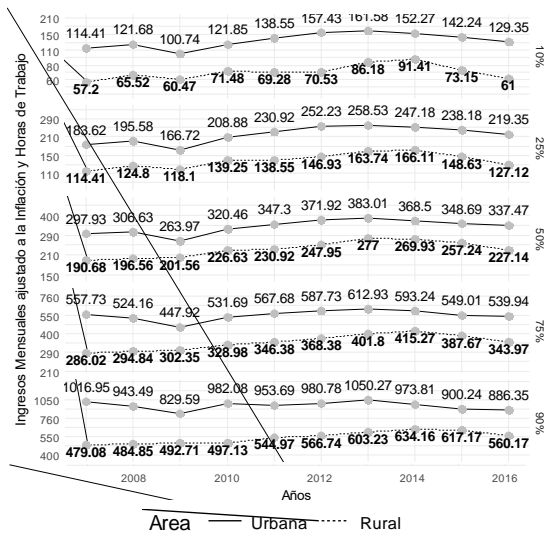


Figura 3.6 Ingresos por Cuantiles, Años y Áreas

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

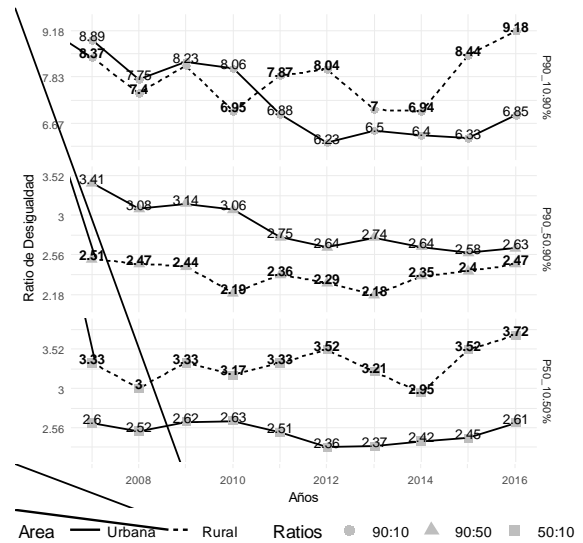


Figura 3.7 Ratios de Desigualdad por Años y Áreas

Interpretaciones

- En la figura 3.6 se observamos que hay diferencias en los ingresos entre quienes viven en áreas urbanas y rurales en todos los cuantiles, los que viven en áreas rurales tiene menor ingresos con respecto quienes viven en áreas urbanas.
- En la figura 3.7 se observamos que la desigualdad en áreas rurales es mayor que en áreas urbanas, así se observa en los ratios 90:10, 90:50 y 50:10

3.5 Descripción General de los cuantiles por Año y Zonas

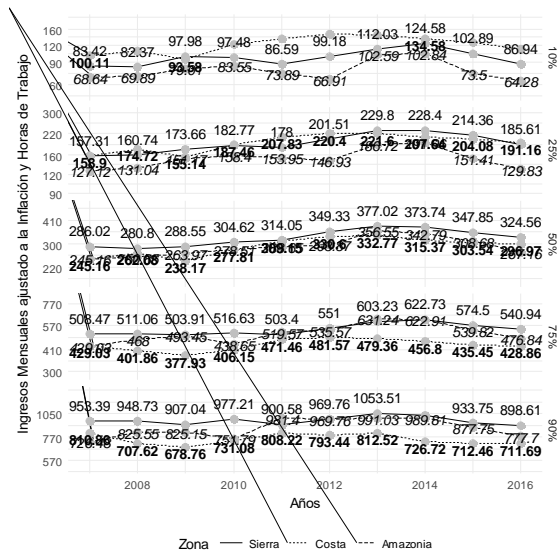


Figura 3.8 Ingresos por Cuantiles, Años y Zonas

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

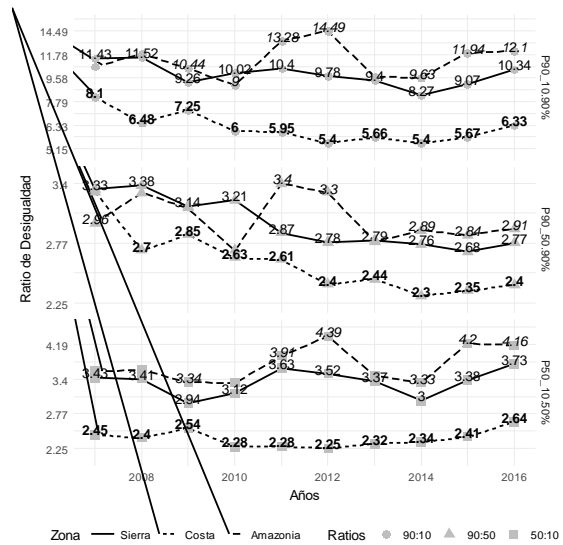


Figura 3.9 Ratios de Desigualdad por Años y Zonas

Interpretaciones

- En la figura 3.8 se observa que los ingresos en las zona Sierra se tiene mayor ingresos que la zona Costa y Amazonía en los cuantiles altos, en el cuantiles mediano se observamos que no existe mucha diferencia y en los cuantiles más bajos vemos que en la Costa se tiene mayor ingresos
- En la figura 3.9 se observa que en la Amazonía existe menor desigualdad en ingresos en comparación de otras zonas según todos los ratios

3.6 Descripción General de los cuantiles por Año y Educación

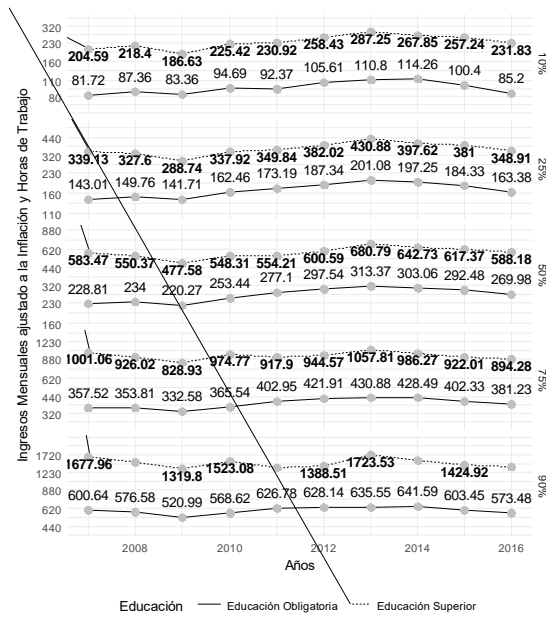


Figura 3.10 Ingresos por Cuantiles, Año y Educación

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

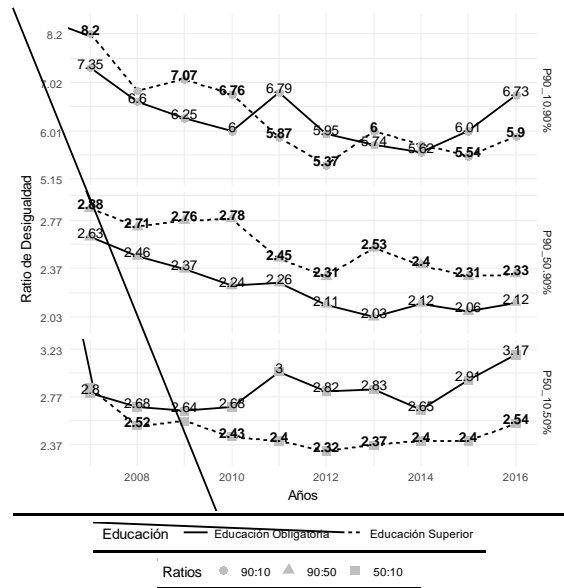


Figura 3.11 Ratios de Desigualdad por Año y Educación

Interpretaciones

- En la Figura 3.10 observamos que los ingresos de las personas con educación superior es mayor que las personas sin educación superior en todos los cuantiles.
- En la Figura 3.11 se observa que la desigualdad en ingresos entre lo que tiene educación superior e inferior de aquellos que no en todos los ratios.

3.7 Descripción General de los cuantiles por Año y Años de Experiencia

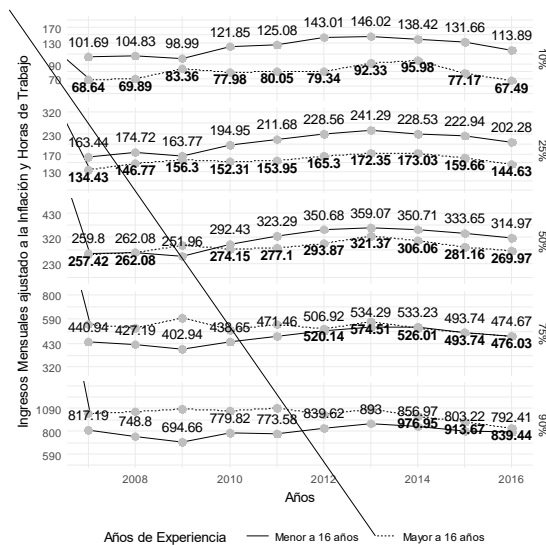


Figura 3.12 Ingresos por Cuantiles, Años y Años de experiencias

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

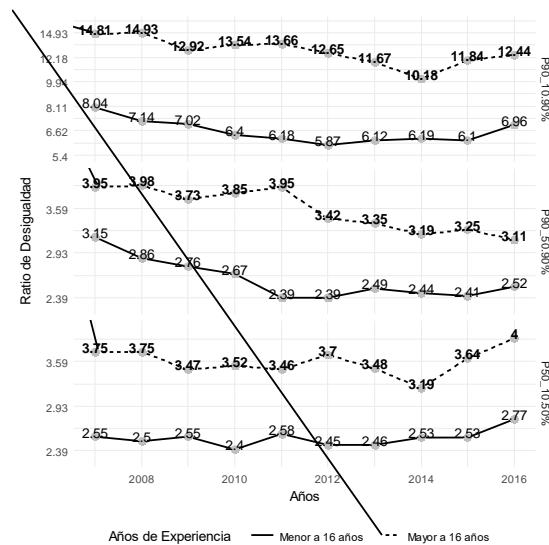


Figura 3.13 Ratios de Desigualdad por Años y Años de experiencias

Interpretaciones

- En la figura 3.12 se observa que las personas con menos años de experiencia tiende a obtener mayores ingresos en los cuantiles más bajos y mediano, mientras que en el cuantil 90% se observa que las personas con mayor años experiencia recibe mayores ingresos por su trabajo.
- En la figura 3.12 se observa una menor desigualdad en los ingresos por aquellos grupos que tienen menos años de experiencia en la ocupación en que trabaja.

3.8 Cuantiles por Áreas, Zona, Sexo y Años de la Encuesta

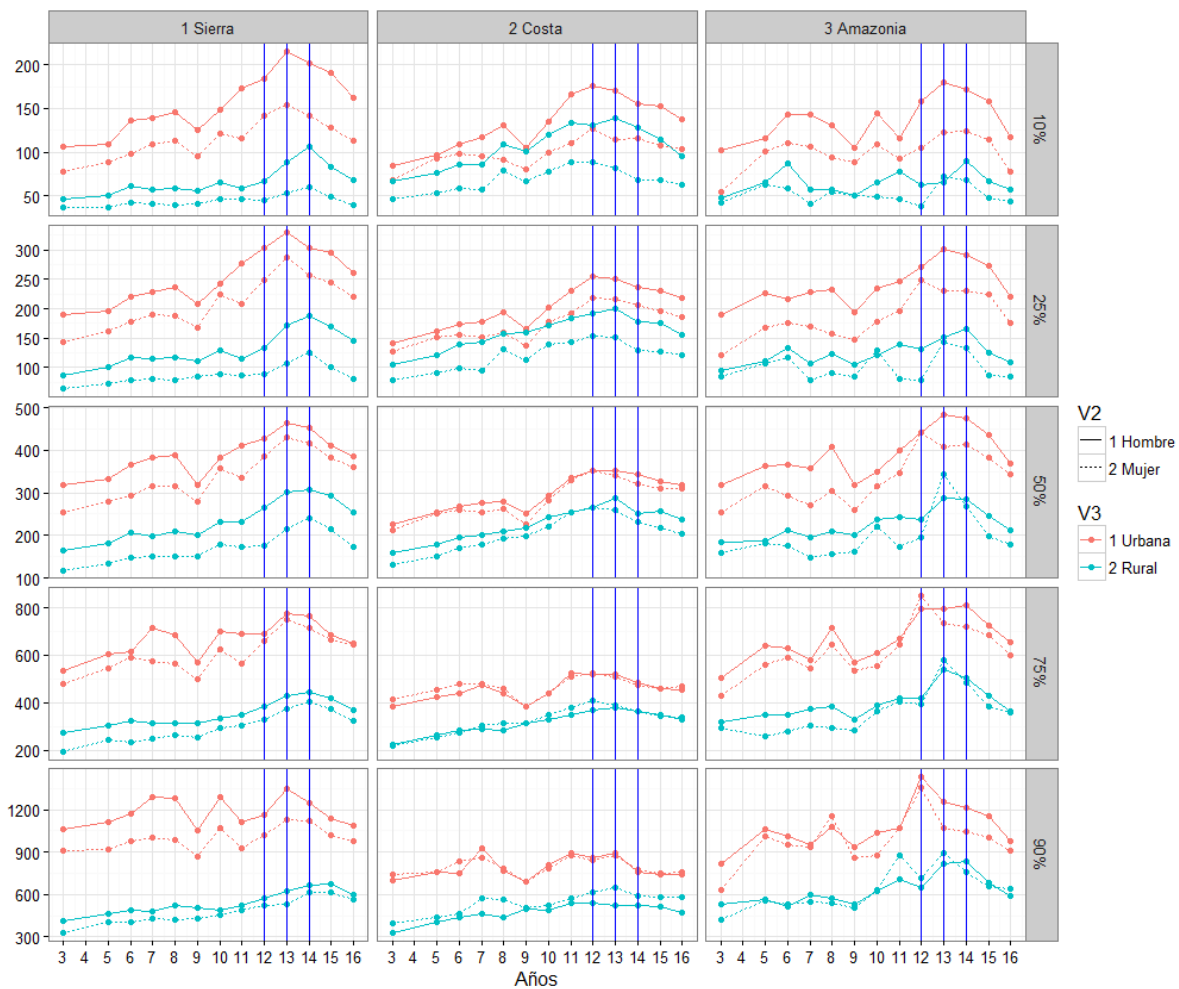


Figura 3.14 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Región

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

Interpretaciones

- Se Observa que los ingresos en la Costa es menor en comparación de la demás regiones en especial en cuantiles mayores
- La brecha entre los ingresos de Hombres y Mujeres disminuye en cuantiles mayores excepto en la Área Urbana de la Sierra donde la brecha entre Hombres y Mujeres no parece disminuir con los cuantiles.
- La brecha entre los ingresos de los que viven en Área Urbana y Rural se mantiene entre los cuantiles, es interesante notar que está brecha es menor en la zona Costa

- Se observa una bajada en los ingresos en general desde el 2012 o 2013.
- Se observa que la mediana de ingresos está entre 200 y 400 dólares en las zonas urbanas y rural entre 100 y 200 dólares (2016).

3.9 Cúantiles por Nivel de Instrucción, Área, Sexo y Años de la Encuesta

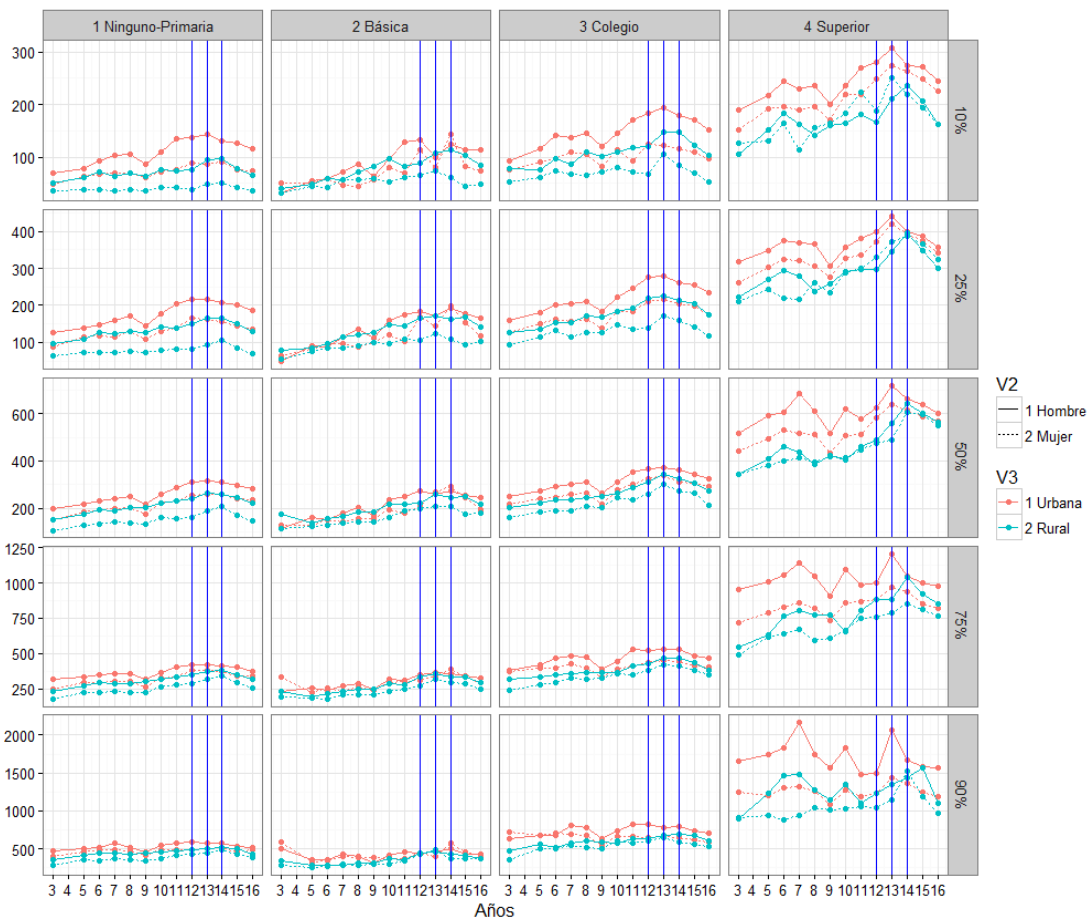


Figura 3.15 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Nivel de Instrucción

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

Interpretaciones

- Se observa que existe un aumento notable en los ingresos aquellas personas con educación Superior con respecto aquellos que no tienen este nivel de instrucción
- La brecha entre género es más grande en el grupo de personas con instrucción superior, aunque en la mayoría de los cuantiles se muestra que esta brecha está disminuyendo.

3.10 Cúantiles por Nivel de Instrucción, Años de ocupación, Área, Sexo y Años de la Encuesta

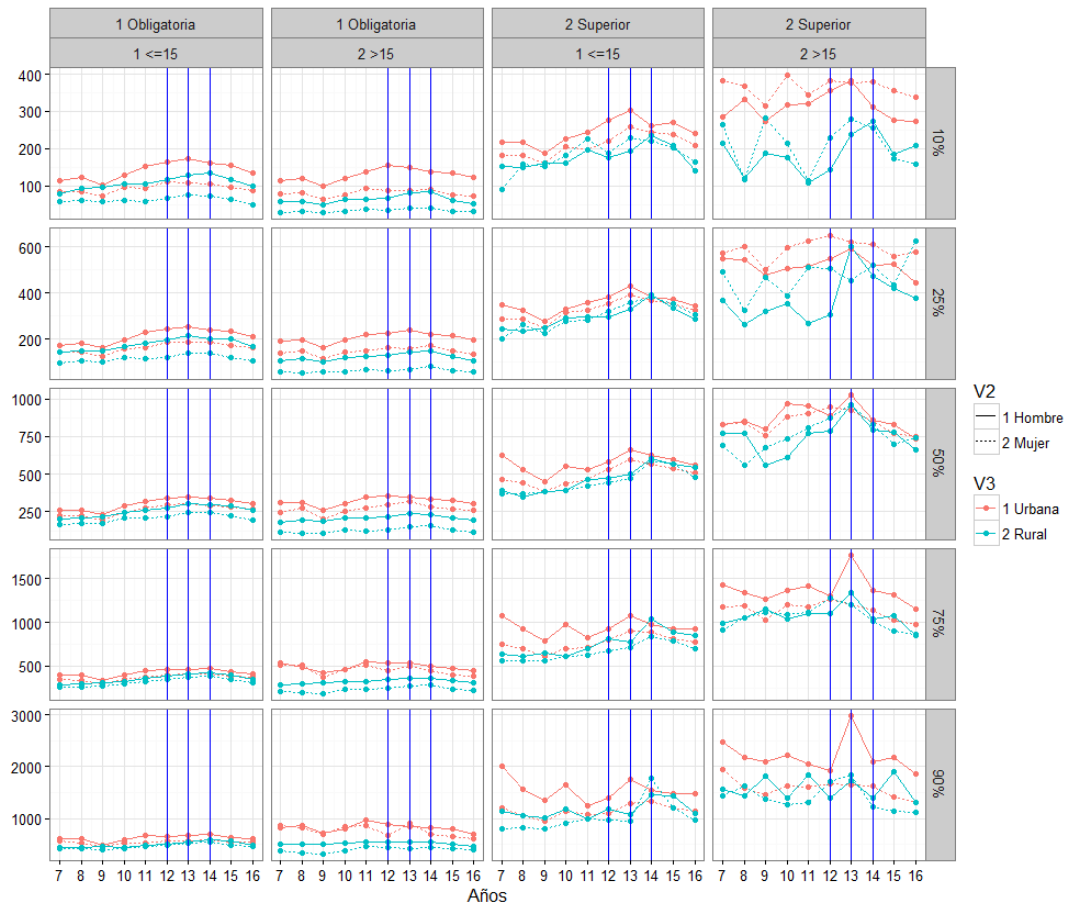


Figura 3.16 Gráficas de la evolución de los Ingresos desde el año 2003 separado por Género, Área, Cuantiles y Nivel de Instrucción

Fuente: INEC, La Encuesta de Empleo, Desempleo y Subempleo, 2007-2016

Interpretaciones

- Se observa que existe cambio apreciable de la distribución de los ingresos con la experiencia cuando se tiene instrucción de educación superior, esto se puede deberse que la mayoría de lo que tiene educación superior tiene ocupaciones más especializa

CAPÍTULO 4

4. INFERENCIA CON MODELOS ESTADÍSTICO

En este capítulo se analizará algunas hipótesis sobre la estructura de distribución de los Ingresos utilizando modelo cuantilico robusto cuyo error siga una distribución “Normal Contaminada Asimétrica” con parámetros ν , γ y σ . Para conocer cómo se comporta y evolucionan los ingresos en sus cuantiles a través de los años y de los diferente grupos.

4.1 Modelo 1: Ingresos con respecto al Tiempo (INGH~ANIO)

En este modelo se relaciona los ingresos monetario por años en diferentes cuantiles, contiene 10 variables independiente que se ha estimado. En la tabla 4.1 se muestra los parámetros de la distribución del error estimado por cada cuantil.

Tabla 4.1 Valores de los parámetros

	10%	25%	50%	75%	90%
ν	0,0729	0,1219	0,5764	0,1381	0,0516
γ	0,0237	0,0973	0,1592	0,1698	0,0420
σ	0,1000	0,2500	0,5000	0,7500	0,9000
lks	-140150,65	-110034,49	-89523,42	-104623,99	-129659,83

Fuente: Análisis de Datos del ENEMDU

Además con los coeficientes de este modelo podemos estimar con intervalo de confianza los ingresos promedio en cada cuantil por diferentes años.

Tabla 4.2 Estimación de los ingresos por Cuantiles y años

	10%	25%	50%	75%	90%
2007	\$ 97.31 (±2.06%)	\$152.56 (±2.87%)	\$260.64 (±2.93%)	\$508.01 (±2.50%)	\$870.35 (±1.68%)

	10%	25%	50%	75%	90%
2008	\$ 99.82 (±2.09%)	\$159.21 (±2.98%)	\$266.60 (±3.07%)	\$500.74 (±2.64%)	\$851.48 (±1.75%)
2009	\$ 97.47 (±0.81%)	\$153.26 (±1.18%)	\$258.52 (±1.29%)	\$484.10 (±1.14%)	\$819.54 (±0.78%)
2010	\$108.76 (±2.13%)	\$171.57 (±3.05%)	\$286.95 (±3.19%)	\$524.10 (±2.68%)	\$905.19 (±1.76%)
2011	\$113.33 (±1.74%)	\$182.30 (±2.43%)	\$308.99 (±2.65%)	\$552.03 (±2.28%)	\$913.24 (±1.56%)
2012	\$122.93 (±1.91%)	\$196.71 (±2.84%)	\$330.57 (±2.89%)	\$580.24 (±2.54%)	\$948.04 (±1.68%)
2013	\$130.32 (±2.00%)	\$208.20 (±2.94%)	\$348.84 (±3.12%)	\$610.49 (±2.72%)	\$992.94 (±1.81%)
2014	\$128.89 (±1.33%)	\$202.30 (±1.91%)	\$338.37 (±2.04%)	\$592.21 (±1.95%)	\$954.84 (±1.31%)
2015	\$118.37 (±0.87%)	\$189.77 (±1.29%)	\$320.23 (±1.42%)	\$556.88 (±1.35%)	\$899.47 (±0.91%)
2016	\$106.12 (±0.89%)	\$172.94 (±1.29%)	\$298.77 (±1.41%)	\$530.92 (±1.30%)	\$853.49 (±0.89%)

Fuente: Análisis de Datos del ENEMDU

Con este modelo podemos revisar los siguientes hipótesis:

H1: Los ingresos del año n con respecto al año n-1 no ha cambiado significativamente

En la Figura 4.1 observamos que:

- Hubo un aumento significativo en los ingresos de aprox. un 4% en los cuantiles 20% y 25% en el año 2008 con respecto al año 2007.
- Hubo una disminución significativa entre del 10% al 5% en todos los cuantiles de los ingresos en el año 2016 con respecto al 2015 y en el año 2015 con respecto al 2014.
- Hubo un aumento significativo entre 4% al 12% de los ingresos en todos los cuantiles desde 2010 hasta el 2013.

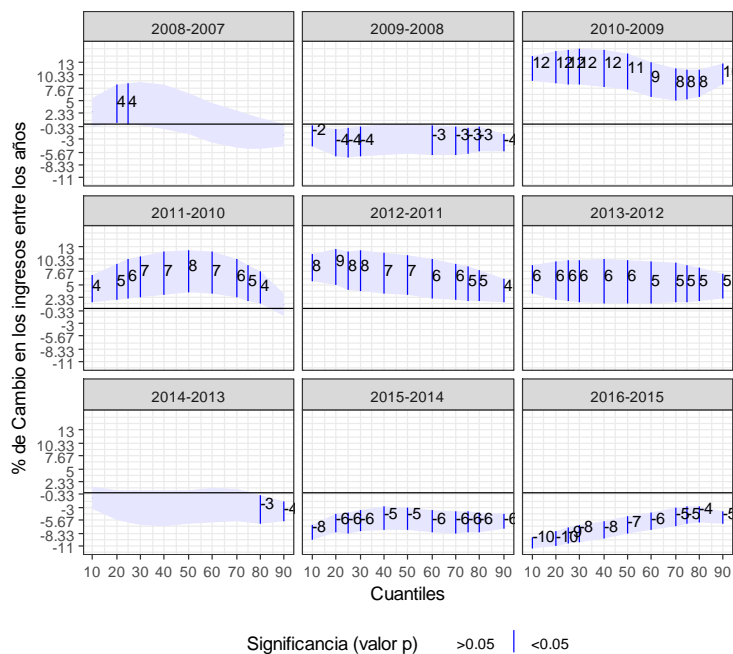


Figura 4.1 Porcentaje de cambio en los ingresos por años consecutivos
Fuente: Análisis de Datos del ENEMDU

H2: La proporciones para medir desigualdad de 90:10, 50:10, 90:50 del año n con respecto al año n-1 no ha cambiado significativamente (se verificará por cada años)

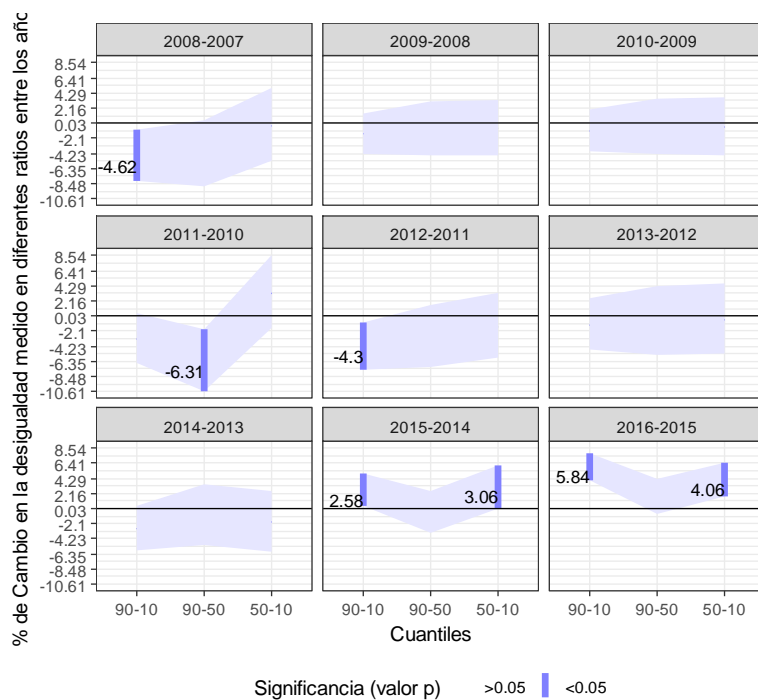


Figura 4.2 Porcentaje de Cambio de ratios entre años consecutivos.
Fuente: Análisis de Datos del ENEMDU

En la figura 4.2 vemos que si ha cambiado la desigualdad de manera significativa:

- El ratio 90:10 ha disminuido significativamente en el 2008 con respecto al 2007, en el 2012 con respecto al 2011, y aumentado en el 2015 con respecto al 2014 y en el 2016 con respecto al 2015.
- El ratio 90:50 ha disminuido significativamente en el 2011 con respecto al 2010
- El ratio 50:10 ha aumentado significativamente en el 2016 y 2015 respectivamente

4.2 Modelo 2: Ingresos con respecto al tiempo, Zona geográfica y Área rural o urbana (INGH~ANIO*ZONA*AREA)

En este modelo se relaciona los ingresos monetario por años en diferentes cuantiles. A continuación se muestra los parámetros de la distribución del error del modelo estimado por los principales cuantiles.

Tabla 4.3 Valores de los parámetros

	10%	25%	50%	75%	90%
ν	0,0692	0,1157	0,4687	0,1432	0,0566
γ	0,0242	0,0985	0,1946	0,1568	0,0380
σ	0,1177	0,2212	0,2093	0,2202	0,1211
lks	-117625	-91172	-73319	-85671	-108600

Fuente: Análisis de Datos del ENEMDU

La fórmula del modelo completo es:

$$\sim(\text{Interception})+ \text{ANIO}+ \text{ZONA}+ \text{AREA}+ \text{ANIO:ZONA}+ \text{ANIO:AREA}+ \text{REGI:AREA}+ \text{ANIO:ZONA:AREA}$$

A continuación revisaremos las hipótesis a considerar:

H1: Años, Zona y Área son independientes/dependientes entre sí por cada cuantil.

Para comprobar esta hipótesis, se va a testear si el conjunto de coeficientes asociados a la triple interacción ANIO:ZONA:AREA son

todos igual a 0 o al por lo contrario existe algún coeficiente distinto a 0, esto se hará por cada cuantíl.

Tabla 4.4 Valores de los parámetros

Interacción ANIO:REGI:AREA		
Cuantiles	Estadístico de prueba	Rechazo si es mayor:
10%	53.5186	5.373 (Se rechaza la Ho)
25%	41.4132	5.373 (Se rechaza la Ho)
50%	34.7499	5.373 (Se rechaza la Ho)
75%	28.1029	5.373 (Se rechaza la Ho)
90%	33.5479	5.373 (Se rechaza la Ho)

Fuente: Análisis de Datos del ENEMDU

Observamos en la tabla 4.4 que todos los cuantiles se rechaza la hipótesis de que los coeficientes asociados a la interacción ANIO:REGI:AREA son distintos a 0, demostrando así que las tres variables son dependientes entre sí para explicar los ingresos.

H2: No hay diferencias significativas entre los Ingresos Áreas Rural y Urbana. (Estimar en qué proporción son diferentes por Año y Cuantil)

En la Figura 4.3 se observa una disminución de las proporciones en el año 2009 en la zona de Costa y Sierra, mientras que en la Sierra existió una disminución de esta proporción en el 2014 y que la hipótesis de que no existe diferencia significativa en los ingresos de quienes viven en áreas rurales, con respecto a quienes viven en áreas urbanas se rechaza en todos los años, cuantiles y zonas geográficas, pues las estimaciones de las proporciones entre los ingresos de quienes viven en áreas urbanas con respecto a quienes viven en áreas rurales son todos significativos, la varianza de estas estimaciones es mayor en la Amazonía.

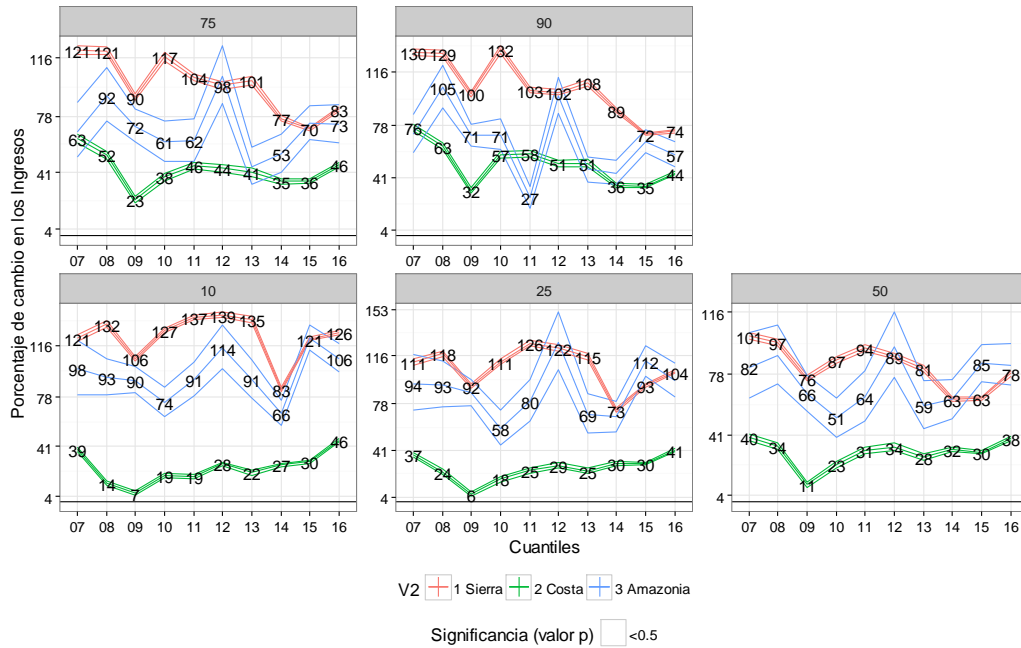


Figura 4.3 Porcentaje cambio en los ingresos en los cuantiles debido al área urbana y rural

Fuente: Análisis de Datos del ENEMDU

H3: No hay diferencias significativas entre las proporciones de desigualdad entre Áreas Rural y Urbana.

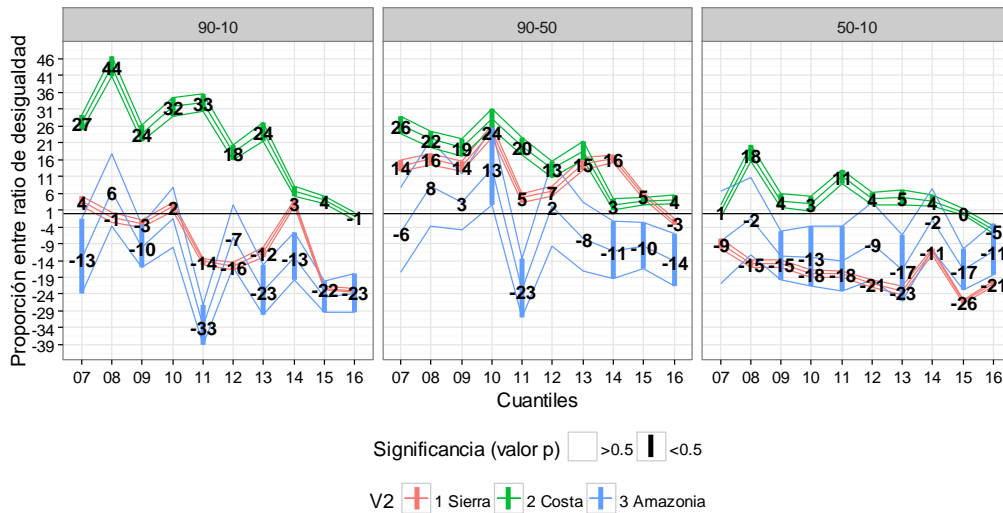


Figura 4.4 Porcentaje de cambio entre los ratios de desigualdad de las Áreas Urbana con respecto al Rural.

Fuente: Análisis de Datos del ENEMDU

Observamos que las proporciones de los ratios de desigualdad 90:10, 90:50 y 50:10 entre Áreas Urbana y Rurales disminuye en la Costa, en la Sierra también disminuye esto ratios, mientras en la Amazonía no se observa una tendencia clara.

4.3 Modelo 3: Ingresos con respecto al Tiempo, Zona, Área y Sexo (INGH~ANIO*ZONA*AREA*SEXO)

En este modelo se relaciona los ingresos monetario por años en diferentes cuantiles. A continuación se muestra los parámetros de la distribución del error del modelo estimado por los principales cuantiles.

Tabla 4.5 Valores de los parámetros

	10%	25%	50%	75%	90%
ν	0,0662	0,1100	0,4742	0,1399	0,0555
γ	0,0242	0,0993	0,1966	0,1571	0,0382
σ	0,1174	0,2217	0,2083	0,2207	0,1213
lks	-115033,3640	-89329,8482	-72464,3431	-85249,6537	-108157,3741

Fuente: Análisis de Datos del ENEMDU

La fórmula del modelo completo es:

~(Interception)+	ANIO:SEXO+
ANIO+	ZONA:SEXO+
ZONA+	AREA:SEXO+
AREA+	ANIO:ZONA:AREA+
SEXO+	ANIO:ZONA:SEXO+
ANIO:ZONA+	ANIO:AREA:SEXO+
ANIO:AREA+	ZONA:AREA:SEXO+
ZONA:AREA+	ANIO:ZONA:AREA:SEXO

A continuación revisaremos las hipótesis a considerar:

H1: Años, Zona, Área y Sexo son independientes/dependientes entre sí por cada cuantil.

Para comprobar esta hipótesis, se va a testear si el conjunto de coeficientes asociados a la triple interacción ANIO:ZONA:AREA:SEXO son todos igual a 0 o al por lo contrario existe algún coeficiente distinto a 0, esto se hará por cada cuantil.

Tabla 4.6 Valores de los parámetros

Interacción ANIO:REGI:AREA		
Cuantiles o a:	Estadístico de prueba	Rechazo si es may
10%	19.1450	5.373 (Se rechaza la Ho)
25%	13.0738	5.373 (Se rechaza la Ho)
50%	8.3049	5.373 (Se rechaza la Ho)

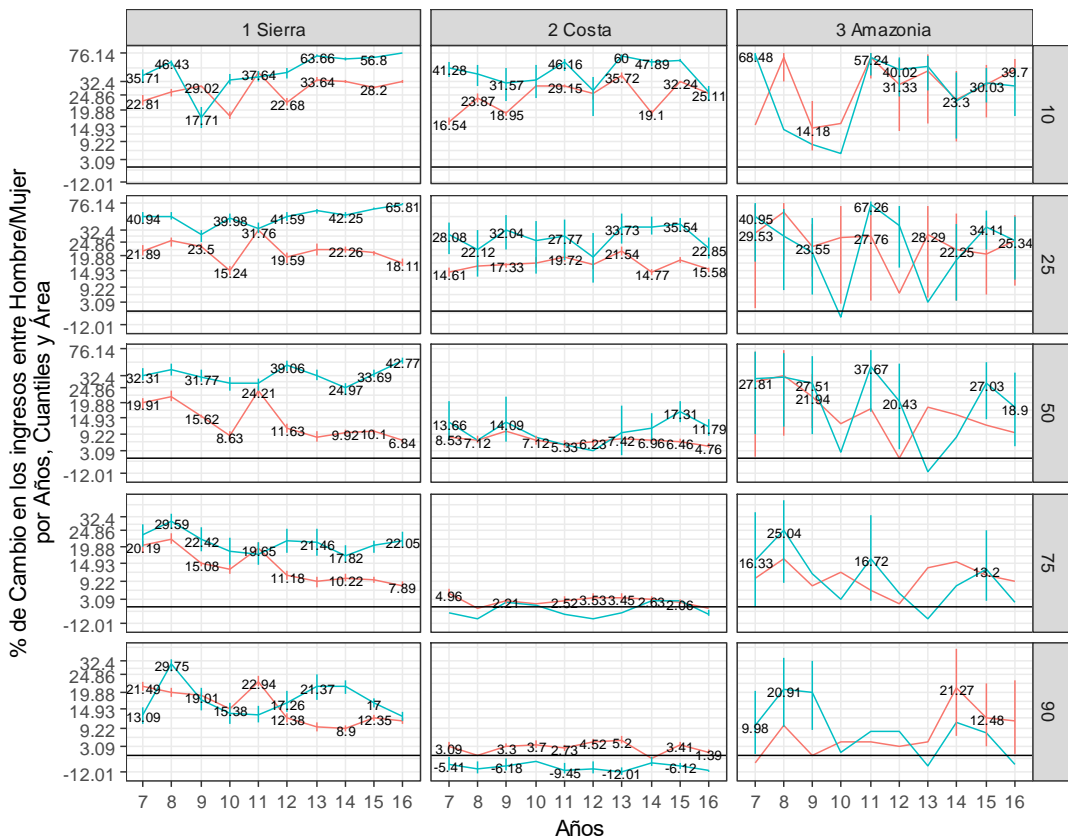
75%	5.4305	5.373 (Se rechaza la Ho)
90%	9.2244	5.373 (Se rechaza la Ho)

Fuente: Análisis de Datos del ENEMDU

Observamos en la tabla 4.4 que todos los cuantiles se rechaza la hipótesis de que los coeficientes asociados a la interacción ANIO:REGI:AREA son distintos a 0, demostrando así que las tres variables son dependiente entre sí para explicar los ingresos.

H2: No hay diferencias significativas entre los Ingresos de Sexo Masculino y Femenino. (Estimar en qué proporción son diferentes por Año, Área y Cuantil)

En la Figura 4.5 se observa que la hipótesis de que no existe diferencia significativa en los ingresos de Hombres y mujeres se rechaza en en la mayoría de los casos, excepto para el cuantil 10% en la zona Costa Rural, y el cúantil 75% y 90% en la zona de Amazonía durante ciertos años.



Significancia (valor p) >0.05 | <0.05 Área + 1 Urbana + 2 Rural
Figura 4.5 Proporción de los Ingresos entre Hombre/Mujer
Fuente: Análisis de Datos del ENEMDU

CONCLUSIONES

5. CONCLUSIONES

1. Los Ingresos recibido mensualmente por los ecuatorianos ha aumentado de manera generalizada en todos los cuantiles hasta el 2014, después de lo cual comenzó a disminuir
2. Se ha observa una disminución de la desigualdad desde el 2007, pero en el 2014 se ha visto un aumento en la desigualdad de los Ingresos.
3. El Sexo si es influyente en el ingreso que se recibe, en especial en los cuantiles más bajo, en los cuantiles alto la diferencia no es muy pronunciada
4. El Nivel de Instrucción es influye en los Ingresos, en especial se observa un salto significativo entre aquellos que tiene educación superior y aquellos que no.
5. Se observa que la desigualdad en área urbana en la costa es significativamente más alta que la desigualdad en la área rural, efecto que no observa igual en la Sierra y en el Oriente

RECOMENDACIONES

6. RECOMENDACIONES

1. Las Bases de datos ENEMDU de diferentes trimestre están muy dispersos, sería bueno ordenar mejor dichas bases de datos
2. En este trabajo se puede expandir para incluir otras variables como profesión o carga en el hogar.+
3. Se puede consolidar las bases de datos de años anteriores al 2007, aunque dichas bases de datos no contenía todos los variables usadas.

BIBLIOGRAFÍA

- Buchinsky, M. (1994, 03). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica*, 62.
- Christian E. Galarza, V. H. (2016, 06). Robust Quantile Regression using a Generalized Class of Skewed Distributions. *Instituto de Matemática, Estatística e Computação Científica*, 07.
- INEC. (2007-2016). La Encuesta de Empleo, Desempleo y Subempleo. *La Encuesta de Empleo, Desempleo y Subempleo*.
- INEC. (2014). Índice de Precios al Consumidor. *Índice de Precios al Consumidor*.
- Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, Inc.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
- Pereira, P. S. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics*, 11.
- T.J. Hastie, R. T. (1990). *Generalized Additive Models* (1 ed.). Chapman and Hall/CRC.
- Wichitaksorn, N. C. (2014, 12). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics*, 42.

APÉNDICES

APÉNDICE A
ARCHIVOS EN EL CD

Nombre del Archivo	Dirección	Descripción
200312 mercado_laboral_educacion_dic2003. sav	\bases de datos\sav	Base de datos del ENEMDU
200512 mercado_laboral_educacion_dic2005. sav		Base de datos del ENEMDU
200612 mercado_laboral_educacion_dic2006. sav		Base de datos del ENEMDU
200712 mercado_laboral_educacion_dic2007. sav		Base de datos del ENEMDU
200809 Septiembre_per2008.sav		Base de datos del ENEMDU
200810 Octubre_per2008.sav		Base de datos del ENEMDU
200812 mercado_laboral_educacion_dic2008. sav		Base de datos del ENEMDU
200903 empleo_subempleo_2009_mar_spss.s av		Base de datos del ENEMDU
200906 Base Datos Junio 2009.sav		Base de datos del ENEMDU
200909 septiembre2009_per.sav		Base de datos del ENEMDU
200912 mercado_laboral_educacion_dic2009. sav		Base de datos del ENEMDU
201012 mercado_laboral_educacion_dic2010. sav		Base de datos del ENEMDU
201112 mercado_laboral_educacion_dic2011. sav		Base de datos del ENEMDU
201212 mercado_laboral_educacion_dic2012. sav		Base de datos del ENEMDU
201312 mercado_laboral_educacion_dic2013. sav		Base de datos del ENEMDU
201409_EnemduBDD_15anios (2).sav		Base de datos del ENEMDU
201412 mercado_laboral_educacion_dic2014. sav		Base de datos del ENEMDU
201503 mercado_laboral_educacion_mar2015 .sav		Base de datos del ENEMDU
201506 mercado_laboral_educacion_jun2015. sav		Base de datos del ENEMDU
201509_EnemduBDD_15anios.sav		Base de datos del ENEMDU
201512_EnemduBDD_15anios.sav		Base de datos del ENEMDU
201603_EnemduBDD.sav		Base de datos del ENEMDU
201606_EnemduBDD.sav		Base de datos del ENEMDU
201609_EnemduBDD_per.sav		Base de datos del ENEMDU
dat.R		Código en R para estructurar las bases de datos
EM2.R		Código del algoritmo EM para la regresión cuantilica robusta

Nombre del Archivo	Dirección	Descripción
lqr_axuliares.R		Funciones auxiliares que se usó para la programación de la regresión cuantílica
prog.c.cpp		Código en C++ usado en lqr_axuliares
datos.RData		Datos estructurados del ENEMDU en .Rdata
DATOS.sav		Datos estructurados del ENEMDU en .sav
convsh.R	\bases de datos\	Códigos auxiliares para las pruebas de Hipótesis
descriptivo_area_general.R		Códigos para la estadística descriptivas
descriptivo_educación_general.R		Códigos para la estadística descriptivas
descriptivo_experiencia_general.R		Códigos para la estadística descriptivas
descriptivo_ingresos_general.R		Códigos para la estadística descriptivas
descriptivo_sexo_general.R		Códigos para la estadística descriptivas
descriptivo_zona_general.R		Códigos para la estadística descriptivas
gráficas.R		Códigos para la gráficas de los modelos
graficas_descriptivas.R		Códigos para la estadística descriptivas
hipotesis_modelos.R		Las Hipótesis de los modelos
hist_w.R		Código Histograma ponderado
leer_modelos.R		Leer los modelos
mod1.R		Código para ejecutar el modelo 1
mod2.R		Código para ejecutar el modelo 2
mod3.R		Código para ejecutar el modelo 3
mod1.RData		Datos del modelo 1
mod2.RData		Datos del modelo 2
mod3.RData		Datos del modelo 3
hipotesis.Rdata		Datos de las hipótesis en diferentes modelos
bibliografia.bib		\Escrituras\
bibliografia.xml	Bibliografía en formato .xml	
Poster.pdf	Poster en pdf	
Poster.pptx	Poster en pptx	
Proyecto Integrador.docx	Proyecto Integrador en .docx	
Proyecto Integrador.pdf	Proyecto Integrador en .pdf	