



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

“Análisis Exploratorio Multivariado de la Fecundidad en Ecuador”

PROYECTO INTEGRADOR

Previo la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

Wendy Alexandra Yagual Peñafiel

GUAYAQUIL - ECUADOR

Año: 2021

DEDICATORIA

A Dios, porque a pesar de las adversidades me dio la fuerza necesaria para llegar hasta el final. A mi familia, que me apoyó en todo momento sin perder la fe en mí.

AGRADECIMIENTOS

Quiero empezar agradeciendo a Dios, porque nada de esto hubiese sido posible sin la sabiduría y fortaleza que me da día a día.

A mi familia por brindarme su apoyo para continuar hasta el final, pero sobre todo a mi mami que dejó todo a un lado para estar conmigo en este proceso.

A mi hermano, que confió en mis habilidades y me dio ánimo hasta su último día de vida.

A la ESPOL y a sus docentes por formarme como profesional, especialmente a mi tutora Eva Mera por su guía y confianza durante este proyecto.

A mis amigos que me brindaron palabras de aliento y apoyaron durante esta jornada, gracias por hacer más llevadera la etapa universitaria.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Yo, Wendy Alexandra Yagual Peñafiel doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual."

Wendy Yagual P.

Wendy Alexandra Yagual Peñafiel

EVALUADORES

A handwritten signature in blue ink, appearing to read 'Eva Mera Intriago', is centered within a light beige rectangular box.

Ph.D. Sandra García Bustos

PROFESOR DE LA MATERIA

Mgtr. Eva Mera Intriago

PROFESOR TUTOR

RESUMEN

Conforme han transcurrido los años, la tasa de fecundidad en Ecuador ha ido decreciendo hasta aproximarse al promedio global de 2.4 nacimientos por mujer (The World Bank, 2019), lo cual nos lleva a querer conocer que factores son los que influyen en dicho comportamiento. Si bien es cierto ya existen varios estudios referentes a la fecundidad, pero para países en vías en desarrollo como Ecuador no es muy común encontrar una gran variedad o son muy antiguos. Por consecuencia, se decidió hacer un estudio del número de hijos que tienen las mujeres ecuatorianas de 15 a 49 años según su edad, etnia, estado civil, nivel de instrucción, si sabe leer y escribir, región de residencia y área de residencia. Donde se emplean tres tipos de modelos: Regresión Poisson Zero-Truncado, Regresión Logística Binaria y Regresión Logística Multinomial, empleando bases de datos de nacimientos fetales del INEC desde el año 2015 al 2019 y usando el lenguaje de programación R. Los modelos se ajustaron correctamente a los datos y concordaron en que los factores determinantes en el número de hijos que tiene una mujer son la edad, etnia, estado civil, si sabe leer y escribir, nivel de instrucción alcanzado, región de residencia y área de residencia. Además, considerando el valor promedio global de fecundidad, el presente estudio determinó que la propensión de que una mujer ecuatoriana tenga más de dos hijos es más alta para las que no saben leer ni escribir, habitan en el área rural y no tienen ningún nivel de instrucción o su nivel de instrucción más alto es el centro de alfabetización.

Palabras claves: Fecundidad, Factores influyentes, Variables sociodemográficas, Propensión.

ABSTRACT

As the years have passed, the fertility rate in Ecuador has been decreasing as it approaches the global average of 2.4 births per woman (The World Bank, 2019), which leads us to want to know what factors influence such behavior. Although it is true that there are already several studies regarding fertility, for developing countries like Ecuador it is not very common to find a great variety of studies or they are outdated. Consequently, I decided to carry out a study of the number of children that Ecuadorian women between the ages of 15 and 49 have based on their age, ethnicity, marital status, level of education, if they can read and write, region of residence and area of residence. There are three types of models used: Zero-Truncated Poisson Regression, Binary Logistic Regression and Multinomial Logistic Regression, using INEC fetal birth databases from 2015 to 2019 and using the R programming language. The models were adjusted accordingly to the data and it agreed that the determining factors in the number of children a woman has are age, ethnicity, marital status, if she can read and write, level of education attained, region of residence and area of residence. In addition, considering the global average fertility value, the present study determined that the propensity for an Ecuadorian woman to have more than two children is higher for those who cannot read or write, live in rural areas and do not have any level of instruction or your highest instructional level is the literacy center.

Key words: Fertility, Influencing factors, Sociodemographic variables, Propensity.

ÍNDICE GENERAL

RESUMEN.....	vi
ABSTRACT	vii
ÍNDICE GENERAL	viii
ÍNDICE DE GRÁFICOS.....	x
ÍNDICE DE TABLAS.....	xi
ÍNDICE DE CUADROS.....	xii
CAPÍTULO 1.....	1
1. Introducción.....	1
1.1 Descripción del problema.....	3
1.2 Justificación del problema	4
1.3 Objetivos	5
1.3.1 Objetivo General	5
1.3.2 Objetivos Específicos.....	5
1.4 Marco Teórico	5
1.4.1 Fecundidad	6
1.4.2 Natalidad.....	7
1.4.3 Modelos de Regresión Multivariados	7
1.4.4 Software.....	14
1.4.5 Revisión Bibliográfica	15
CAPÍTULO 2.....	17
2. METODOLOGÍA y datos	17
2.1 Datos y Variables	17
2.1.1 Variables Discretas	18

2.1.2	Variables Continuas	19
2.1.3	Variables Categóricas	19
2.2	Regresión Poisson	20
2.3	Regresión Logística.....	20
2.3.1	El modelo	21
2.3.2	Estimación de Máxima Verosimilitud.....	23
2.3.3	Bondad de Ajuste.....	24
CAPÍTULO 3.....		27
3.	RESULTADOS Y ANÁLISIS	27
3.1	Análisis Univariado	27
3.2	Análisis Descriptivo	31
3.3	Análisis Bivariado	33
3.4	Modelo de Regresión Poisson Zero-Truncado	40
3.5	Modelos de Regresión Logística	45
3.5.1	Modelo de Regresión Logística Binaria (Variables Originales).....	45
3.5.2	Modelo de Regresión Logística Binaria (Variables Recodificadas)	49
3.5.3	Modelo de Regresión Logística Multinomial	53
CAPÍTULO 4.....		57
4.	CONCLUSIONES Y RECOMENDACIONES.....	57
4.1	Conclusiones.....	57
4.2	Recomendaciones.....	59
BIBLIOGRAFÍA.....		61

ÍNDICE DE GRÁFICOS

Gráfico 3.1 Histograma de frecuencias del número de hijos por mujer.....	32
Gráfico 3.2 Histograma de frecuencias de la edad de la madre	33
Gráfico 3.3 Diagrama de cajas de la etnia de la madre vs el número de hijos.....	35
Gráfico 3.4 Diagrama de cajas del estado civil de la madre vs el número de hijos.....	36
Gráfico 3.5 Diagrama de cajas del nivel de instrucción de la madre vs el número de hijos	37
Gráfico 3.6 Diagrama de cajas de si la madre sabe leer y escribir vs el número de hijos	38
Gráfico 3.7 Diagrama de cajas de la región de residencia de la madre vs el número de hijos	39
Gráfico 3.8 Diagrama de cajas del área de residencia de la madre vs el número de hijos	40

ÍNDICE DE TABLAS

Tabla 3.1 Número de hijos según la etnia de la madre.....	27
Tabla 3.2 Número de hijos según el estado civil de la madre.....	28
Tabla 3.3 Número de hijos según el nivel de instrucción de la madre	29
Tabla 3.4 Número de hijos acorde si la madre sabe leer y escribir.....	29
Tabla 3.5 Número de hijos según la región de residencia de la madre.....	30
Tabla 3.6 Número de hijos según el área de residencia de la madre	30
Tabla 3.7 Número de hijos por intervalos	31
Tabla 3.8 Estadísticas descriptivas del número de hijos por mujer.....	32
Tabla 3.9 Estadísticas descriptivas de la edad de la mujer.....	33
Tabla 3.10 Número de hijos acorde a los grupos de edades de la mujer	34
Tabla 3.11 Estimaciones del modelo de regresión Poisson Zero-Truncado	41
Tabla 3.12 Estimaciones del modelo de regresión logística binaria con las variables originales	46
Tabla 3.13 Estimaciones del modelo de regresión logística con las variables recodificadas	50
Tabla 3.14 Coeficientes del modelo de regresión logística multinomial.....	53
Tabla 3.15 Valores p del modelo de regresión logística multinomial.....	54
Tabla 3.16 Odds ratio del modelo de regresión logística multinomial	54

ÍNDICE DE CUADROS

Cuadro 3.1 Análisis de devianza para el modelo con las variables originales	49
Cuadro 3.2 Análisis de devianza para el modelo con variables recodificadas	52

CAPÍTULO 1

1. INTRODUCCIÓN

“La fecundidad es una de las variables demográficas más importantes para evaluar la tendencia del crecimiento de la población”. (INEC, 2017) La fecundidad hace referencia a lo que es real, es decir al número de hijos que tiene una mujer durante su vida reproductiva. A diferencia de la fertilidad que se refiere a lo que es posible, o sea la capacidad de una mujer para poder tener hijos. Por lo tanto, se puede ser fértil y no tener hijos; o, haber fecundado y dejar de ser fértil con el tiempo. (Pérez, 2010) Si se desea calcular la tasa de fecundidad, se debe hacer un estudio en función de las mujeres que se encuentran en edad reproductiva, es decir, mujeres entre 15 a 49 años. (Knoema, 2020)

Desde la década de 1960, varios países en vías de desarrollo experimentaron un rápido decrecimiento en la fecundidad, de hecho, algunos de ellos lograron llegar a 2.1 nacimientos que es el número de nacimientos esperados a lo largo de la vida fértil de una mujer. No obstante, para algunos países esta disminución ha sido más lenta o incluso se ha estancado, como en el caso de Ecuador. Entre 1955 y 1969, Ecuador se mantuvo en 7.0 nacimientos promedio por mujer, luego disminuyó a 3.8 hasta el año 1989, y culminó en aproximadamente 3.3 nacimientos para el año 2004. (Ishida, Stupp, & Ordoñez, 2011)

La fecundidad es un indicador importante de la transición demográfica de un país. En consecuencia, los estudios de fecundidad se consideran de suma importancia para llegar a una estimación de la población futura, analizando por qué variables se ve modelada. La razón de su relevancia en el campo demográfico es porque se

considera que los niños que nacen están supuestos a ser el sustento para el futuro del país. (Lanchimba & Medina, 2007).

En un estudio realizado a las mujeres ecuatorianas en el año 2001, asociaron el comportamiento reproductivo de la mujer con los factores socioeconómicos, desarrollando indicadores relacionados a la edad, etnia, zona de residencia, nivel de instrucción, características económicas y estado civil. Estas variables mostraron altos valores en el número de hijos para las mujeres que habitan en el área rural, auto declaradas indígenas, con baja instrucción y ubicadas en los niveles económicos más bajos hasta fines de los años 70'. Sin embargo, entre 1982 y 2001 se observó una disminución en los grupos de mujeres por edad, exceptuando al grupo de 15-19 años, indicando la presencia de fecundidad temprana. Al comparar la encuesta ENDEMAIN del 2004 con el censo del 2001, se apreció un descenso más intenso para el área rural que casi le permitió estabilizarse. Lo contrario al área urbana donde incrementó la fecundidad en la mayoría de las edades. (INEC, 2001)

(Buizza & Villafuerte, 2020) estudiaron el descenso de la fecundidad, empleando datos derivados de la ENSANUT 2018 para Ecuador. Obteniendo como resultado que los grupos sociales con más recursos son capaces de controlar eficazmente el nivel de fecundidad al tener dos hijos o menos, lo contrario de los grupos como: minorías étnicas (especialmente indígenas), áreas rurales (especialmente de la Amazonía), hogares pobres y mujeres que abandonan el sistema escolar o se casan en edad temprana. Además, a partir del año 2016, observan una reducción de fecundidad más significativa en las adolescentes. Ellos concluyen que la causa principal de la disminución de la fecundidad entre los años 2004 y 2018, es el aumento en el uso de métodos anticonceptivos modernos, sobre todo para las mujeres más jóvenes entre 15 y 24 años.

En el presente estudio se desea identificar las variables sociodemográficas

determinantes en el número de hijos que tienen las mujeres ecuatorianas entre los años 2015 y 2019. Empleando variables como la edad, estado civil, nivel de instrucción, etnia, área de residencia, entre otros. Y a su vez, analizar la relación entre la edad de la mujer ecuatoriana con el número total de hijos que ha tenido durante su vida fértil.

En este capítulo se describe y justifica cual es el problema, se mencionan los objetivos y se expone el marco teórico de lo que es relevante para el estudio. En el capítulo 2 se encuentra la metodología de los datos y modelos estadísticos a emplear. En el capítulo 3 se describen los resultados de los análisis estadísticos: univariados, bivariados y multivariados. Y, finalmente se presenta el capítulo 4 con las conclusiones y recomendaciones obtenidas del estudio de fecundidad en el Ecuador.

1.1 Descripción del problema

El modelamiento de la fecundidad ayuda a identificar las tendencias pasadas, presentes o futuras del crecimiento poblacional. Para los países en vías de desarrollo su estudio es esencial para poder planear, monitorear y evaluar los programas de salud y bienestar familiar no solo para la población actual sino también la futura. (Pandey & Kaur, Modelling fertility: an application of count regression models, 2015) Por consiguiente, hay dos posibles problemas que pueden presentarse: sobrepoblación o infrapoblación, lo cual puede ser medido a través de la Tasa Global de Fecundidad (TGF) o el número de nacimientos por mujer.

(The World Bank, 2021) indica que la tasa global de fecundidad en el año 1960 mundialmente era aproximadamente de 5.0 nacimientos por mujer, valor que descendió considerablemente hasta llegar a 2.4 en el año 2019. Esto demuestra que históricamente se han presentado diferentes fenómenos sociales y

económicos que intervinieron en el comportamiento de dicha variable, ya sea la introducción de la mujer al campo educativo y laboral, o el progreso de los sistemas de salud, urbanización y modernización. (INEC, 2001) Ecuador no ha sido la excepción, el país ha demostrado el mismo comportamiento. Su tasa de fecundidad en el año 1960 era de 6.7, valor mucho mayor a la tasa promedio de ese año. Sin embargo, dicho valor disminuyó hasta aproximarse a 2.4 nacimientos por mujer para el año 2019, que era el promedio o mínimo necesario para que una población se mantenga estable. Dicho esto, el presente estudio desea medir el comportamiento del número de hijos que tiene una mujer explicado a través de variables sociodemográficas como edad, etnia, estado civil, nivel de instrucción, si sabe leer y escribir, región y área de residencia.

1.2 Justificación del problema

En un mundo donde la fecundidad juega un papel importante en el campo demográfico, es sustancial estudiar el comportamiento del número de hijos que tiene una mujer. Si se modela dicha variable de respuesta a través de otras variables características de la mujer en edad fértil, se puede tener una estimación de la población futura, considerando que las nuevas generaciones son las que se encargan de relevar a las antiguas. Esta información ayuda a un país en la toma de acciones sociales según las características de las mujeres jóvenes, de mediana edad o mayores porque les permite identificar cuales de ellas tienen el mayor número de hijos, procrean más hijos de los que pueden mantener, se encuentran solas, requieren de educación, etc. Conociendo estas características se pueden tomar ciertas medidas como programas del uso de métodos anticonceptivos, ayuda social, económica y educativa.

Por lo tanto, se puede interpretar el perfil de las mujeres que intervienen en el estudio de la fecundidad, aportando información para la implementación de las medidas necesarias para cuidar la salud e integridad de las madres y los niños.

Una mujer debería ser capaz de tener el número de hijos que puede mantener y educar.

1.3 Objetivos

1.3.1 Objetivo General

Evaluar las variables determinantes en el comportamiento reproductivo de las mujeres ecuatorianas de 15 a 49 años durante el período del 2015 al 2019 a través de modelos estadísticos multivariados.

1.3.2 Objetivos Específicos

- Analizar el comportamiento de los factores sociodemográficos de las mujeres ecuatorianas entre 15 y 49 años.
- Determinar las variables representativas por las que se ve afectada la fecundidad de las mujeres que habitan en el Ecuador.
- Construir modelos estadísticos que expliquen el comportamiento del número de hijos de las mujeres ecuatorianas durante el período del 2015 al 2019 a través de un conjunto de variables sociodemográficas de la mujer.

1.4 Marco Teórico

En esta sección se describen algunas de las metodologías clásicas que se usan para medir la fecundidad y entrar en contexto del estudio de fecundidad demográfico clásico, además de otras metodologías multivariadas con las que se desea aportar al estudio. Es importante conocer los conceptos básicos de la fecundidad como la tasa de fecundidad general y natalidad, sin embargo, estas no serán empleadas. Además, se habla de la regresión Poisson pero como los datos corresponden a las mujeres que han tenido hijos, la variable de respuesta se encuentra restringida y por ello se emplean otras metodologías claves como la regresión Poisson Zero-Truncado. También se mencionan los modelos de

regresión logística existentes, aunque para el estudio solo se emplean dos de ellos, el de regresión logística binomial y regresión logística multinomial. Donde el binomial tiene dos categorías “Uno o dos hijos” y “Más de dos hijos”; y, el multinomial tiene cinco categorías “Un hijo”, “Dos hijos”, “Tres hijos”, “Cuatro hijos” y “Más de cuatro hijos” para aportar información más detallada al estudio.

En el primer y segundo apartado se habla de las tasas de fecundidad y natalidad, en el tercero de los modelos de regresión multivariados, seguido del software utilizado y por último la revisión bibliográfica.

1.4.1 Fecundidad

La fecundidad hace referencia al número de hijos que tiene una mujer durante su periodo fértil. Se mide a través de la tasa de fecundidad o la tasa de fecundidad general.

1.4.1.1 Tasa de Fecundidad General (TFG)

La tasa de fecundidad general se mide a través de la relación de dos variables, el número de nacidos vivos por año y el número de mujeres en edad fértil.

Su fórmula es la siguiente:

$$TFG = \frac{\text{Número de nacidos vivos por año}}{\text{Número de mujeres entre 15 a 49 años}} * 1000$$

Este indicador aporta información de la transición demográfica de un país e incluso permite estimar su población futura, dado que los niños por nacer son considerados el sustento del futuro. (Ortega, 2018) Es más preciso que la tasa de natalidad porque se mide acorde a un grupo específico de edad, el de mujeres en edad fértil, que por ende tienen mayor probabilidad de dar a luz.

1.4.2 Natalidad

Cuando se habla de natalidad, se habla del número de nacimientos ocurridos en una población durante un periodo determinado. Se mide a través de la tasa de natalidad o tasa bruta de natalidad, la cual generalmente es muy alta para los países subdesarrollados o de tercer mundo.

1.4.2.1 Tasa Bruta de Natalidad (TN)

La tasa bruta de natalidad se mide a través de la relación de dos variables, el número total de nacimientos por año y la población total.

Su fórmula es la siguiente:

$$\text{Tasa bruta de natalidad} = \frac{\text{Número total de nacimientos por año}}{\text{Población total}} * 1000$$

Este indicador se encarga de determinar la tasa de decrecimiento de una población. Es fácil de medir e interpretar, pero puede traer consigo cierto error o imprecisión al no tener restricciones de edad o sexo. (Ortega, 2018)

1.4.3 Modelos de Regresión Multivariados

En este apartado se describe matemáticamente los modelos de regresión Poisson y Logística respectivamente.

1.4.3.1 Regresión Poisson

La regresión Poisson modela la relación entre una variable de conteo, que sigue una distribución Poisson, y un conjunto de variables predictoras de cualquier tipo. Los posibles valores que puede tomar la variable de

respuesta son enteros no negativos. Como es el caso del número de hijos de una mujer entre 15 y 49 años que a lo largo de su vida pudo tener 0 hijos, 1 hijo, etc.

La distribución Poisson modela la probabilidad de que y eventos ocurran en un tiempo determinado. Su distribución de probabilidad es (The Pennsylvania State University, 2018):

$$P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}; y = 0,1, \dots$$

Donde λ es la cantidad promedio de ocurrencias para el evento que se está midiendo.

Se determina λ a través de k conjunto de predictores $\mathbf{X} = (X_1, X_2, \dots, X_k)$, lo cual se escribe como:

$$\lambda = e^{(\mathbf{X}\beta)}$$

Entonces, el modelo de regresión Poisson para una observación i dada es:

$$P(Y_i = y_i|\mathbf{X}_i\beta) = \frac{e^{-e^{(\mathbf{X}_i\beta)}} e^{(\mathbf{X}_i\beta)^{y_i}}}{y_i!}$$

Para encontrar los coeficientes del modelo de regresión Poisson, se usa el método de máxima verosimilitud. Donde para una muestra de tamaño n se encuentra dado por:

$$L(\beta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \frac{e^{-e^{(\mathbf{X}_i\beta)}} e^{(\mathbf{X}_i\beta)^{y_i}}}{y_i!}$$

Aplicando logaritmo:

$$P(\beta) = \sum_{i=1}^n y_i \mathbf{X}_i \beta - \sum_{i=1}^n e^{(\mathbf{X}_i \beta)} - \sum_{i=1}^n \log(y_i!)$$

1.4.3.1.1 El modelo

Para la distribución Poisson, se define como y a la probabilidad de observar un conteo específico. Siendo su ecuación la siguiente (Pandey & Kaur, 2015):

$$f\left(\frac{y_i}{x_i}\right) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}; y_i = 0, 1, \dots$$

Este modelo parametriza la relación entre el parámetro λ y los predictores x , que para este estudio resultan ser siete. Además, incorpora la heterogeneidad observada a través de:

$$E(y_i/x_i) = \lambda_i = e^{x_i\beta}; i = 1, 2, \dots, n$$

Y como se conoce que la media y la varianza para una distribución Poisson es la misma, entonces:

$$Var(y_i/x_i) = \lambda_i$$

Si se presenta equidispersión en el modelo (media igual a la varianza), habría restricciones en su uso. Por ello, se podrían aplicar modelos con especificaciones más generales como el de clasificador de Bayes. Cuando se usan otros modelos alternativos, se lo hace con el fin de considerar esa variación extra existente y la heterogeneidad que no se logra observar.

Una función de enlace se encarga de relacionar la media de la función con la variable predictora lineal. Esta función se usa para transformar la relación de no lineal a una lineal, por ello el modelo también puede ser llamado modelo log-lineal. Se expresa de la siguiente manera:

$$\log(\mu_i) = \log(e^{x_i\beta}) = x_i\beta$$

Que para este estudio sería:

$$\log(\mu_i) = x_i\beta = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7$$

Donde:

- X_1, \dots, X_7 son las variables predictoras.

- β_0 es el intercepto.
- β_1, \dots, β_7 son los coeficientes desconocidos del modelo que acompañan a las variables predictoras.

Para estimar los parámetros se puede aplicar la estimación de máxima verosimilitud; o, cuasi verosimilitud máxima.

1.4.3.2 Regresión Poisson Zero-Truncado

El modelo Zero-Truncado es un caso especial de la regresión Poisson donde los conteos observados son estrictamente positivos, o sea no pueden tomar el valor de cero. Como el modelo de regresión Poisson permite que la variable de respuesta tome valores desde cero a infinito, quitar el cero como posible valor genera una infradispersión en los datos. (UCLA, 2016). Los datos que disponemos son solo de nacimientos, es por esto que el modelo que se utilizará para modelar los datos del estudio es el número de nacimientos de las mujeres que han tenido al menos un hijo.

La distribución Zero-Truncada se define por una función de probabilidad condicional donde $y > 0$ (Van der Heijden, Cruyff, & Van Houwelingen, 2003):

$$P(y_i | Y_i > 0, \lambda) = \frac{P(y_i | \lambda)}{P(Y_i > 0 | \lambda)} = \frac{e^{-\lambda} \lambda^{y_i}}{y_i! (1 - e^{-\lambda})}; \quad y_i = 1, 2, \dots$$

Asumiendo que hay un estimador $\hat{\lambda}$ para λ que puede usarse para determinar la probabilidad de un individuo no observado $p_0 = e^{-\lambda}$. El número de individuos no observados se denotaría por:

$$f_0 = \frac{\hat{p}_0}{1 - \hat{p}_0} N_{obs}$$

Siendo N_{obs} el número de individuos observados en la muestra. Por lo que la estimación de la población \hat{N} vendría dada por:

$$\hat{N} = \hat{f}_0 + N_{obs}$$

Siendo Y_1, \dots, Y_N una muestra aleatoria de una distribución Poisson Zero-Truncada con el parámetro $\lambda_i, i = 1, \dots, N$; y, y_1, \dots, y_N las realizaciones de la muestra aleatoria, se denota el logaritmo como:

$$\log(\lambda_i) = \beta^T x_i$$

Donde $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ y $x_i = (1, x_{i1}, \dots, x_{ip})^T$

1.4.3.3 Regresión Logística

La regresión logística es una técnica multivariante que modela la relación entre una variable categórica y un conjunto de variables predictoras cualitativas o cuantitativas. Se encarga de estimar la probabilidad o propensión de caer dentro de uno de los niveles de la variable dependiente, según los predictores dados. (The Pennsylvania State University, 2018) A continuación se describen algunos tipos de regresión logística como: Regresión Logística Binaria, Multinomial, Nominal y Ordinal.

1.4.3.4 Regresión Logística Binaria

La regresión logística binaria o binomial se utiliza cuando la variable de respuesta es una variable cualitativa dicotómica, es decir, una variable a la que se le asigna los valores de 0 o 1. Es importante mencionar que para este tipo de modelo, las categorías o niveles de la variable dicotómica deben ser mutuamente excluyentes. (Huamaní, 2014)

Su modelo es el siguiente (The Pennsylvania State University, 2018):

$$\pi(\mathbf{X}) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

Donde:

- π es la probabilidad de suceso, es decir, la probabilidad de que una observación se encuentre dentro de una categoría de la variable dicotómica Y .
- e representa a la función exponencial.
- $\beta_0, \beta_1, \dots, \beta_k$ son los parámetros desconocidos del modelo.
- X_0, X_1, \dots, X_k son las variables predictoras del modelo.

Los parámetros se estiman usando el método de máxima verosimilitud, que para una muestra de tamaño n se encuentra dado por:

$$L(\beta; y, \mathbf{X}) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

$$= \prod_{i=1}^n \left(\frac{e^{(X_i \beta)}}{1 + e^{(X_i \beta)}} \right)^{y_i} \left(\frac{1}{1 + e^{(X_i \beta)}} \right)^{1-y_i}$$

Aplicando logaritmo:

$$P(\beta) = \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)]$$

$$= \sum_{i=1}^n [y_i \mathbf{X}_i \beta - \log(1 + e^{(X_i \beta)})]$$

Además, para determinar la relación que existe entre los predictores y la

variable de respuesta, se emplean los odds ratio. Que para dos variables predictoras X_1 y X_2 se expresa como:

$$\theta = \frac{(\pi/(1/\pi))_{X=X_1}}{(\pi/(1/\pi))_{X=X_2}}$$

Entonces para la regresión logística binaria, se denota la razón de probabilidades de suceso como:

$$\frac{\pi}{1 - \pi} = e^{X\beta}$$

Si se aplica logaritmo a la expresión, resulta más sencillo la interpretación:

$$\log\left(\frac{\pi}{1 - \pi}\right) = \log(e^{X\beta})$$

$$\log\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right) = b_0 + \sum_{i=1}^n b_i x_i$$

Donde b_0, b_1, \dots, b_n son los parámetros estimados del modelo.

Es importante mencionar que los odds ratio pueden tomar cualquier valor que no sea negativo. Si estos toman el valor de 1, se puede concluir que no hay asociación entre la variable predictora y la variable de respuesta. Si son mayor que 1, se concluye que los odds de suceso son mayores para los niveles más altos de la variable predictora continua; o, para un nivel específico del factor. Lo contrario a si son menor que 1, donde los odds de suceso serán menores.

1.4.3.5 Regresión Logística Multinomial

La regresión logística multinomial es una extensión de la regresión logística binaria que se emplea cuando la variable de respuesta es

politómica, es decir, cuando tiene más de dos categorías. (Dueñas, 2010)

Su modelo es el siguiente (Dueñas, 2010):

$$P[Y = 1|X] = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_s)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_s)}}$$

Donde $P[Y = 1|X]$ hace referencia a la probabilidad de que Y sea 1.

Resulta más sencillo interpretar si se transforma el modelo logarítmicamente obteniendo una función lineal:

$$\log \left[\frac{P(Y = 1|X)}{1 - P(Y = 1|X)} \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_s$$

1.4.3.6 Regresión Logística Nominal

La regresión logística nominal se usa cuando hay tres categorías o más, pero los niveles no siguen un orden.

1.4.3.7 Regresión Logística Ordinal

Esta regresión también se usa cuando hay tres categorías o más; sin embargo, la diferencia recae en que ésta si considera el orden en los niveles.

1.4.4 Software

El software utilizado en este estudio es R, el cual se encarga de facilitar la manipulación de datos, cálculos, gráficos y demás empleando un lenguaje de programación. (Venables & Smith, 2021)

1.4.5 Revisión Bibliográfica

Hay varios estudios sobre la fecundidad y sus variables determinantes, por ello a continuación se resumen algunos de ellos.

(Pandey & Kaur, 2015) estudiaron el comportamiento de la fecundidad en India a través de un modelo de Regresión Poisson, donde la variable número de hijos es la variable de respuesta; y, educación, ocupación, índice de riqueza, religión, lugar de residencia y casta son las variables predictoras. El estudio reveló que la educación de las mujeres, los esposos con mayor nivel educacional, la participación de mujeres en actividades comerciales y el estatus económico son factores influyentes en la reducción de nacimientos. A diferencia de las normas sociales y culturales de la sociedad india que contribuyen al aumento de nacimientos.

(Chowdury & Moloy, 2014) realizaron un estudio donde relacionaban el número de hijos con las variables demográficas, socioeconómicas y culturales de las mujeres de Bangladesh. Para ello hicieron un análisis bivariado y luego emplearon un modelo de Regresión Logística Binaria donde la variable dependiente era el número de hijos; y, las variables independientes eran la edad, edad al primer matrimonio, edad al primer nacimiento, número ideal de hijos, hijos previamente muertos, uso de anticonceptivos, estatus laboral actual, educación, lugar de residencia, religión, división y exposición a los medios de comunicación. La variable de respuesta fue transformada a una variable dicotómica: 2 hijos o menos y 3 hijos o más. Del análisis bivariado se obtuvo que la variable uso de anticonceptivos no es relevante por lo que se la removió. Sin embargo, el modelo de regresión logística indicó que el resto de las variables analizadas sí influyen en el número de hijos. Por ende, se llegó a la conclusión de que las mujeres más educadas o con un mejor estatus socioeconómico tienen menos hijos, a diferencia de aquellas cuyos esposos

no terminan la educación primaria o se casan a una edad temprana que terminan teniendo 3 o más.

(Canto, Ortiz, & Ortiz, 2018) llevaron a cabo un estudio para determinar el efecto de las preferencias de fecundidad en el número de hijos para las mujeres trabajadoras en Perú. Para ello emplearon un modelo de regresión logística ordinal, donde la variable dependiente era el número total de hijos nacidos vivos (variable con tres categorías o más donde importa el orden de los niveles) y las principales variables independientes eran el número de hijos deseados, edad al primer matrimonio, edad al tener el primer hijo, índice de riqueza, método anticonceptivo utilizado y el tipo de trabajadora (dependiente o independiente). Como resultado concluyeron que las variables analizadas influyen en gran medida en el número de hijos nacidos vivos, haciendo especial énfasis en la introducción de la mujer en el ámbito laboral. Pues las mujeres que se encuentran trabajando ajustan su comportamiento reproductivo según su vida laboral, dándole a esta una mayor importancia por sobre la realización del ideal reproductivo.

(Ortíz, 2006) analizó los factores sociodemográficos y socioeconómicos de la fecundidad adolescente en México. Donde mediante un ajuste de regresión logística, con una variable dependiente dicotómica que representa si una adolescente es madre o no, y diferentes variables explicativas demográficas y socioeconómicas, se obtuvo como resultado final que factores como la edad, el estado civil, el nivel de escolaridad y la etnia resultan ser determinantes para que una adolescente sea madre. Factores como el nivel de escolaridad resulta ser un determinante más importante que otras variables como la zona de residencia y nivel de pobreza, puesto que las mujeres con mayores pretensiones académicas y posteriormente laborales retrasan acciones como el matrimonio y su primer hijo.

CAPÍTULO 2

2. METODOLOGÍA Y DATOS

En este capítulo se presentan los métodos empleados en la construcción del modelo estadístico que explica el comportamiento del número de hijos que tiene una mujer ecuatoriana en edad fértil, utilizando las bases de datos de nacimientos fetales del INEC desde el año 2015 al 2019.

Para modelar el total de número de hijos de una mujer ecuatoriana en edad fértil se emplearon tres tipos de modelos. Uno de ellos es el modelo de regresión Poisson Zero-Truncado que explica el comportamiento de una variable cuantitativa de conteo que no puede tomar el valor de 0 en base a un conjunto de variables predictoras cuantitativas o cualitativas. También un modelo de regresión logística binaria, el cual se encarga de modelar una variable de respuesta dicotómica en base a una combinación lineal de variables predictoras. Y por último, el modelo de regresión logística multinomial, el cual se encarga de modelar una variable de respuesta politómica en base a una combinación lineal de variables predictoras.

2.1 Datos y Variables

Para la realización de los cálculos pertinentes, se validaron los datos provenientes de cinco bases de datos de los nacimientos fetales del INEC: ENV_2015, ENV_2016, ENV_2017, ENV_2018 y ENV_2019. Las cuales contienen 47 variables que se dividen en tres secciones: inscripción del niño, características del niño y características de la madre. La sección más relevante para lo que se quiere medir en el estudio es la de los datos de la madre, por lo que se seleccionó un subconjunto de 10 variables del conjunto total, las cuales se detallan a

continuación. Cabe mencionar que la población fue de 962,331 mujeres con edades entre 15 y 49 años.

2.1.1 Variables Discretas

Las variables discretas son variables cuantitativas que pueden asumir un número finito o contable de valores. Se utilizaron tres de ellas en el estudio:

- ¿Cuántos hijos vivos tiene actualmente?
- ¿Cuántos hijos que nacieron vivos han muerto?
- ¿Cuántos hijos nacieron muertos?

Estas tres variables fueron sumadas con el fin de obtener la variable de respuesta, o también llamada dependiente, que es la siguiente:

- Número de hijos.

Esta variable dependiente o de respuesta se utilizará en las tres modelizaciones a realizar, donde sus agrupaciones se hicieron dependiendo del modelo. Esto se describe a continuación:

Para el modelo de regresión Poisson Zero-Truncado, la variable se estudió como una variable discreta de tipo numérica que no puede tomar el valor de 0. Sin embargo, para los modelos de regresión logística se realizaron transformaciones. Para el modelo de regresión logística binaria se la transformó a dicotómica de tipo factor, generando dos grupos: mujeres con más de dos hijos y mujeres con uno o dos hijos. Pero para el modelo de regresión logística multinomial se la transformó a politómica de tipo factor, generando cinco grupos: mujeres con un hijo, dos hijos, tres hijos, cuatro hijos y más de cuatro hijos.

2.1.2 Variables Continuas

Las variables continuas son variables cuantitativas que pueden tomar un número infinito o incontable de valores entre dos cualesquiera. Para el estudio se empleó una de este tipo:

- Edad de la madre.

Esta variable originalmente permitía un rango de edades desde 8 a 56 años. Sin embargo, en este estudio se seleccionaron solo las mujeres de entre 15 a 49 años.

2.1.3 Variables Categóricas

Las variables categóricas son variables cualitativas que poseen un número finito de categorías o propiedades, sin necesariamente llevar un orden. Las que se usaron son las siguientes:

- Autoidentificación étnica de la madre.
- Estado civil y/o conyugal de la madre.
- ¿Sabe leer y escribir?
- Nivel de instrucción alcanzado.
- Provincia de residencia habitual de la madre.
- Área de residencia de la madre.

Donde región de residencia fue formada en base a la variable provincia de residencia habitual de la madre. Como se conoce que Ecuador tiene 24 provincias, el factor provincia de residencia contenía 24 niveles. Es por ello, que se agruparon los datos según la región con la finalidad de tener resultados por región.

Además, para la variable etnia se unificaron los niveles negra, afrodescendiente y mulata en un solo nivel denominado “Afrodescendiente”.

En el estado civil se unificaron los niveles casada, unida y unión de hecho generando uno llamado “Casada/Unida”.

Para el nivel de instrucción se unificaron los niveles educación básica y primaria para generar uno llamado “Educación básica”; y, ciclo post bachillerato se unió con superior no universitario en el nivel denominado “Superior no universitario”.

2.2 Regresión Poisson

Es una técnica estadística multivariante que modela una variable dependiente, cuyos datos de conteo ocurren en un tiempo o espacio determinado, en base a un conjunto de variables independientes de cualquier tipo. Además, se conoce que su variable de respuesta sigue una distribución Poisson y que sus valores esperados pueden ser modelados linealmente a través de parámetros desconocidos. (NCSS)

Esta técnica estadística modela la influencia de diferentes variables predictoras sobre el número de hijos que tiene una mujer ecuatoriana en su edad fértil.

2.3 Regresión Logística

Es una técnica estadística multivariante que busca explicar la relación entre una variable dependiente de tipo dicotómica o politómica, con un conjunto de variables independientes de tipo cualitativo o cuantitativo. Si las variables predictoras son

de tipo categórico que contienen c niveles, se deberán generar c-1 variables dummy, o también llamadas variables ficticias, para que queden bien representadas las variables en el modelo. (Poma & Mercedes)

Aquí se describirá el modelo de regresión específico, luego se mostrará el método que se emplea para estimar los parámetros, para finalmente comprobar el ajuste del modelo a través de pruebas de bondad de ajuste.

2.3.1 El modelo

Para crear el modelo, primero se define cual es la variable dependiente, o de respuesta, denotada por Y . Además, se elige una colección de variables independientes o predictoras, denotadas por el vector $X = (X_1, X_2, \dots, X_p)$ donde para este estudio $p = 1, \dots, 7$. Luego se denota la probabilidad condicional de que la respuesta esté presente como: $P(Y = 1 / X) = \pi$. Finalmente, al aplicar logaritmo sobre la probabilidad de que el evento esté presente, o sea $Y = 1$, se obtiene (Mashfiqul Huq & Deluar, 2014):

$$\text{logit}(\pi) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p ; 0 \leq \pi \leq 1$$

Donde la función queda expresada como:

$$\pi_i = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$$

Que para este estudio en específico es:

$$\pi_i = \frac{e^{(\beta_0 + \beta_1 \text{edad_mad} + \beta_2 \text{etnia} + \beta_3 \text{est_civil} + \beta_4 \text{niv_inst} + \beta_5 \text{sabe_leer} + \beta_6 \text{reg_res} + \beta_7 \text{area_res})}}{1 + e^{(\beta_0 + \beta_1 \text{edad_mad} + \beta_2 \text{etnia} + \beta_3 \text{est_civil} + \beta_4 \text{niv_inst} + \beta_5 \text{sabe_leer} + \beta_6 \text{reg_res} + \beta_7 \text{area_res})}}$$

Donde:

- $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ son los parámetros desconocidos del modelo que se desean estimar.

- *total_hij* es una variable tipo factor que indica cuantos hijos tiene una mujer.

La cual tiene dos niveles: $\begin{cases} 1; \text{ Dos hijos o menos} \\ 0; \text{ Más de dos hijos} \end{cases}$

- *edad_mad* es una variable cuantitativa continua que indica la edad de la madre.
- *etnia* es una variable tipo factor que indica la etnia de la madre.

La cual tiene 6 niveles:

1. Afrodescendiente
2. Blanca
3. Indígena
4. Mestiza
5. Montubia
6. Otra

- *est_civil* es una variable tipo factor que indica el estado civil de la madre.

La cual tiene 5 niveles:

1. Soltera
2. Casada/Unida
3. Divorciada
4. Separada
5. Viuda

- *niv_inst* es una variable tipo factor que indica el nivel de instrucción de la madre.

La cual tiene 8 niveles:

1. Centro de alfabetización
2. Secundaria
3. Educación básica
4. Bachillerato
5. Superior no universitario
6. Superior universitario
7. Postgrado

8. Ninguno

- *sabe_leer* es una variable tipo factor que responde a la pregunta ¿Sabe leer y escribir?

La cual tiene dos niveles: $\begin{cases} 1; \text{ Sí} \\ 0; \text{ No} \end{cases}$

- *reg_res* es una variable tipo factor que indica la región de residencia de la madre.

La cual tiene 4 niveles:

1. Costa
2. Sierra
3. Amazónica
4. Otra

- *area_res* es una variable tipo factor que indica el área de residencia de la madre.

La cual tiene dos niveles: $\begin{cases} 1; \text{ Urbana} \\ 0; \text{ Rural} \end{cases}$

2.3.2 Estimación de Máxima Verosimilitud

Para estimar los parámetros del modelo se aplica una técnica estadística llamada máxima verosimilitud, donde la probabilidad de la variable binaria Y se expresa como (Mashfiqul Huq & Deluar, 2014) :

$$P(Y = 1) = p = \frac{1}{1 + e^{-X\beta}}$$
$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = X\beta$$

Donde:

- β es el vector de coeficientes o parámetros desconocidos.
- X es la matriz de variables independientes, donde la primera columna es una columna de unos.

$$\prod_{i=1}^n P(Y = y_i) = \prod_{i=1}^n p_i(1 - p_i)^{1-y_i}$$

Al carecer del valor de la variable de respuesta, no se pueden estimar los parámetros del modelo directamente. Por ello, el método de máxima verosimilitud se encarga de proveer una solución.

Las observaciones de la variable de respuesta se consideran como un ensayo de Bernoulli, donde para el las i^{th} observaciones $P(Y = y_i) = p_i(1 - p_i)^{1-y_i}$

La función de máxima verosimilitud asumiendo que las n observaciones son independientes, sería:

$$L = \prod_{i=1}^n P(Y = y_i) = \prod_{i=1}^n p_i(1 - p_i)^{1-y_i}$$

Reemplazando y aplicando logaritmo a la función de verosimilitud:

$$\text{Log}L = \sum_{i=1}^n y_i \log \left(\frac{1}{1 + e^{-X'\beta}} \right) + \sum_{i=1}^n (1 - y_i) \log \left(\frac{e^{-X'\beta}}{1 + e^{-X'\beta}} \right) \dots$$

Si se maximiza esta expresión, se pueden estimar los parámetros $\beta_0, \beta_1, \dots, \beta_7$ para el estudio.

2.3.3 Bondad de Ajuste

Para comprobar que tan bien se ajustan los valores predichos a los observados se aplican diferentes pruebas de bondad de ajuste.

Donde la hipótesis nula es (Iglesias, 2012):

$$H_0: p_j = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}}; \quad \forall j = 1, \dots, J$$

Y, la hipótesis alternativa es (Iglesias, 2012):

$$H_1: p_j \neq \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}}; \text{ para algún } j$$

2.3.3.1 Estadístico de la Devianza

Se utiliza el estadístico de Wilks o Devianza para medir la bondad de ajuste del modelo a los datos observados, definido como (Iglesias, 2012):

$$D = -2 \log \left(\frac{\hat{L}_c}{\hat{L}_s} \right) = -2(\log \hat{L}_c - \log \hat{L}_s)$$

Donde $\hat{L}_c = L(\hat{\beta}_0, \hat{\beta}_1)$ para el modelo ajustado; y, $\hat{L}_s = L(\hat{\beta}_0, \hat{\beta}_1)$ para el modelo saturado.

2.3.3.2 Estadístico Chi Cuadrado de Pearson X^2

Este estadístico se encarga de comparar las frecuencias observadas con las esperadas, definido como (Iglesias, 2012):

$$X^2 = \sum_{j=1}^J \frac{(y_j - n_j \hat{p}_j)^2}{n_j \hat{p}_j (1 - \hat{p}_j)} = \sum_{j=1}^J \frac{n_j (y_j - \hat{y}_j)^2}{\hat{y}_j (n_j - \hat{y}_j)}$$

Donde se rechaza la hipótesis nula a un nivel de significancia α si $X^2 = X^2_{j-(R+1); \alpha}$ para el modelo ajustado; y, $\hat{L}_s = L(\hat{\beta}_0, \hat{\beta}_1)$ para el modelo saturado.

2.3.3.3 Medida R^2

Hay otras medidas que se usan para medir la bondad de ajuste, las cuales son similares al R^2 que se usa en los modelos de regresión lineal. Uno de ellos sigue la siguiente ecuación (Iglesias, 2012):

$$R_g^2 = 1 - \left(\frac{\hat{L}_g}{L_0} \right)^{\frac{2}{n}}$$

Donde \hat{L}_g es el logaritmo de la verosimilitud cuando el modelo se evalúa en $(\hat{\beta}, \hat{\beta}_1)$; y, \hat{L}_0 es el logaritmo de la verosimilitud cuando el modelo solo tiene a la constante.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

En esta sección se realiza un análisis univariado exploratorio de las variables que se consideran en el estudio, para luego hacer uno bivariado y por último se presentan los resultados obtenidos de los modelos estadísticos empleados en el estudio de la fecundidad de las mujeres ecuatorianas de entre 15 a 49 años.

3.1 Análisis Univariado

En la tabla 3.1 se observa que el 88.60% de las mujeres ecuatorianas se autodefinen como mestizas, siendo éste el porcentaje más alto para la población. Seguido de la raza indígena representada por el 6.79%. Además, las mujeres consideradas de raza negra y afroecuatoriana componen un porcentaje muy similar, 1.27% y 1.14% respectivamente. Dejando a las etnias restantes con porcentajes muy bajos, por debajo del 1%.

Tabla 3.1 Número de hijos según la etnia de la madre.

Etnia	n	%
Afroecuatoriana/Afrodescendiente	10936	1.14%
Blanca	6487	0.67%
Indígena	65384	6.79%
Mestiza	852579	88.60%
Montubia	5487	0.57%
Mulata	4055	0.42%
Negra	12211	1.27%
Otra	3458	0.36%
Sin información	1734	0.18%

Total	962331	100%
--------------	---------------	-------------

Fuente: INEC

Elaboración propia

En la tabla 3.2 se observa que el 43.90% de la población son madres solteras. Seguido de porcentajes muy similares para los grupos de mujeres casadas y unidas con el 27.68% y 24.27% respectivamente. Además, el 2.30% de las madres se encuentran en una unión de hecho y el 1.59% se han divorciado. Dejando a las viudas y separadas con valores muy pequeños que no llegan ni al 1%.

Tabla 3.2 Número de hijos según el estado civil de la madre.

Estado Civil	n	%
Casada	266375	27.68%
Divorciada	15338	1.59%
Separada	646	0.07%
Soltera	422440	43.90%
Unida	233512	24.27%
Unión de hecho	22132	2.30%
Viuda	1752	0.18%
Sin información	136	0.01%
Total	962331	100%

Fuente: INEC

Elaboración propia

En la tabla 3.3 se observa que el 29.24% de la población son mujeres con educación media o de bachillerato. Valor cercano al de las mujeres que estudiaron hasta la secundaria con un 24.00%. Luego hay tres grupos con porcentajes similares, los cuales son el de las mujeres con educación superior universitario, educación básica y primaria. Cuyos valores son 15.87%, 14.70% y 11.29% respectivamente. El 3.32% corresponde a las mujeres con estudios superiores no universitarios. Además, las mujeres que culminaron sus estudios hasta el centro

de alfabetización, ciclo post bachillerato, postgrado o simplemente carecieron de un nivel de estudio, representan a una parte muy pequeña de la población.

Tabla 3.3 Número de hijos según el nivel de instrucción de la madre.

Nivel de Instrucción	n	%
Centro de alfabetización	1901	0.20%
Ciclo Post - Bachillerato	1216	0.13%
Educación básica	141473	14.70%
Educación media/ Bachillerato	281392	29.24%
Ninguno	7652	0.80%
Postgrado	2480	0.26%
Primaria	108655	11.29%
Secundaria	230939	24.00%
Superior no universitario	31920	3.32%
Superior universitario	152716	15.87%
Sin información	1987	0.21%
Total	962331	100%

Fuente: INEC

Elaboración propia

De la tabla 3.4 se concluye que el 99.09% de las mujeres que tienen hijos si saben leer y escribir, a diferencia del grupo cuya respuesta fue no que representa al 0.64% de la población.

Tabla 3.4 Número de hijos acorde si la madre sabe leer y escribir.

¿Sabe leer y escribir?	n	%
No	6127	0.64%
Sí	953583	99.09%
Sin información	2621	0.27%
Total	962331	100%

Fuente: INEC

Elaboración propia

De la tabla 3.5 se concluye que el 55.34% de las mujeres habitan en la región costa, lo cual es más de la mitad de la población. Luego le sigue la sierra con el 34.45%, seguido de la región amazónica con un 7.10%. Y finalmente el grupo de la región insular se ve representado con el 0.14%.

Tabla 3.5 Número de hijos según la región de residencia de la madre.

Región de Residencia	n	%
Costa	532594	55.34%
Sierra	331516	34.45%
Amazónica	68281	7.10%
Insular	1325	0.14%
Zonas no delimitadas	1	0.00%
Exterior	445	0.05%
Sin información	28169	2.93%
Total	962331	100%

Fuente: INEC

Elaboración propia

De la tabla 3.6 se concluye que hay más mujeres que residen en el área urbana, correspondiendo al 77.25% de la población. A diferencia de las mujeres del área rural que constituyen el 22.75% restante.

Tabla 3.6 Número de hijos según el área de residencia de la madre.

Área de Residencia	n	%
Urbana	743398	77.25%
Rural	218933	22.75%
Sin información	0	0.00%
Total	962331	100%

Fuente: INEC

Elaboración propia

De la tabla 3.7 se concluye que hay el 86.43% de las mujeres ecuatorianas tienen entre uno a cuatro hijos; y, menos del 1% tienen entre ocho a 19 hijos.

Tabla 3.7 Número de hijos por intervalos.

Número de hijos	n	%
[1-4)	831790	86.43%
[4-8)	122821	12.76%
[8-12)	7171	0.75%
[12-16)	526	0.05%
[16-19]	23	0.00%
Total	962331	100%

Fuente: INEC

Elaboración propia

3.2 Análisis Descriptivo

Se calcularon las estadísticas descriptivas como media, varianza, mediana, mínimo, máximo y cuartiles para la variable de respuesta número de hijos (discreta). También se graficaron los histogramas de frecuencias no solo para la variable de respuesta sino también para la variable predictora edad de la madre (continua) con el fin de analizar el comportamiento de sus datos.

La tabla 3.8 muestra que para que una mujer sea parte del estudio de fecundidad, debe tener mínimo un hijo. La mediana o segundo cuartil indica que el 50% de los datos se encuentran sobre y por debajo de 2 hijos por mujer, y el 75% por debajo de 3, lo cual se aprecia en el gráfico 3.1. A su vez, se observa que la media o número promedio de hijos que tiene una mujer durante su vida fértil es de 2.

Tabla 3.8 Estadísticas descriptivas del número de hijos por mujer.

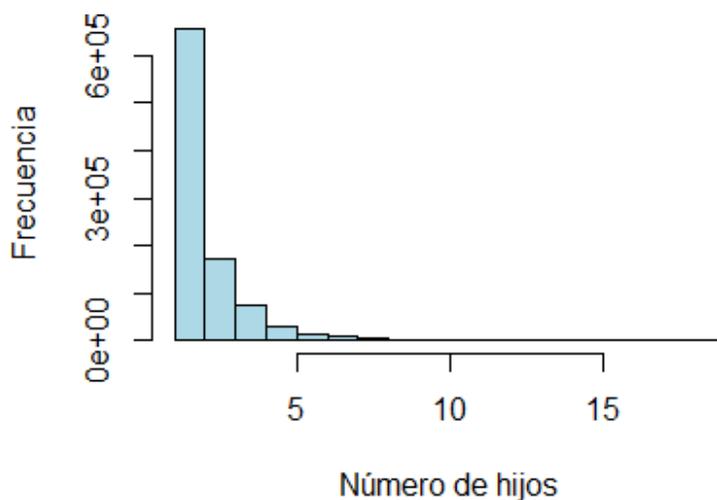
Mínimo	Q1	Mediana	Q3	Máximo	Media	Varianza	Desviación
1.00	1.00	2.00	3.00	19.00	2.19	1.93	1.39

Fuente: INEC

Elaboración propia

El gráfico 3.1 muestra un histograma de frecuencias asimétrico, ya que la mayoría de los datos se encuentran hacia la izquierda, o sea tienen un sesgo positivo.

Gráfico 3.1 Histograma de frecuencias del número de hijos por mujer.



Fuente: INEC

Elaboración propia

La tabla 3.9 muestra que el rango de edad definido para este estudio es mínimo 15 años y máximo 45 años. La mediana o segundo cuartil indica que el 50% de los datos se encuentran sobre y por debajo de 29 años, y el 75% por debajo de 34 años, lo cual se aprecia en el gráfico 3.2. A su vez, se observa que la media o número promedio de edad de una mujer durante su vida fértil es de 29 años.

Tabla 3.9 Estadísticas descriptivas de la edad de la mujer.

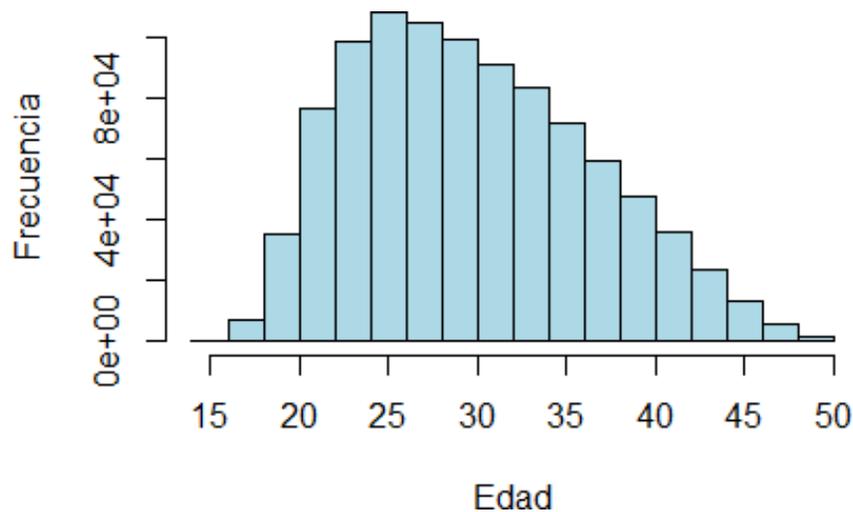
Mínimo	Q1	Mediana	Q3	Máximo	Media	Varianza	Desviación
15.01	24.42	29.02	34.48	49.00	29.75	44.51	6.67

Fuente: INEC

Elaboración propia

El gráfico 3.2 muestra un histograma de frecuencias de la edad de la mujer con una media de aproximadamente 30 años y un ligero sesgo positivo.

Gráfico 3.2 Histograma de frecuencias de la edad de la madre.



Fuente: INEC

Elaboración propia

3.3 Análisis Bivariado

Para esta sección se elaboró una tabla bivariada de frecuencias; y, diagramas de caja donde se observa la distribución de las variables de estudio con sus valores aberrantes.

La tabla 3.10 muestra que las mujeres de entre 33 a 39 años tienen más de cuatro hijos, mientras que, las de 21 a 27 años tienden a tener un solo hijo. Además, no es muy común que las que tienen entre 15 y 21 años tengan cuatro hijos o más.

Tabla 3.10 Número de hijos acorde a los grupos de edades de la mujer.

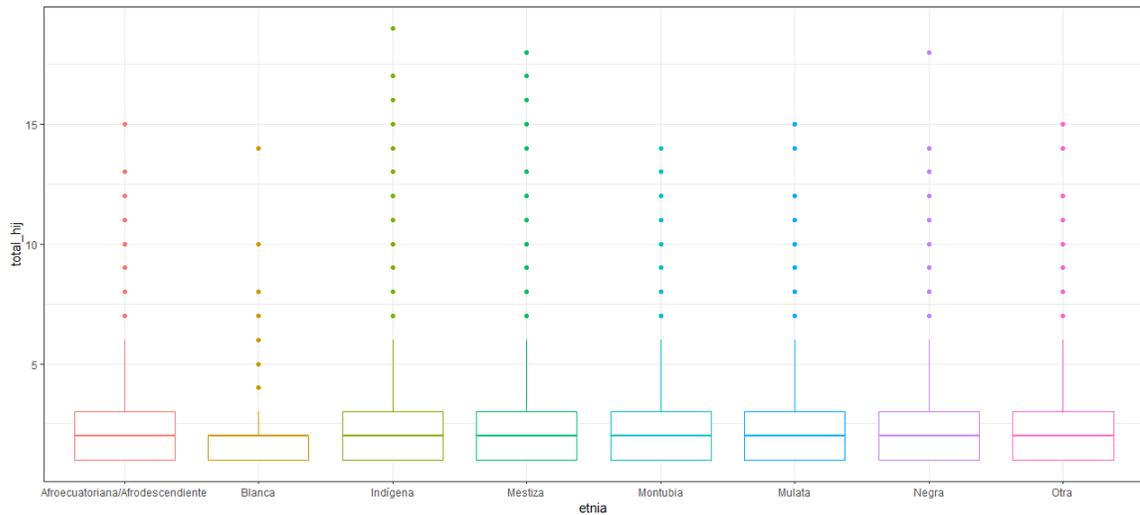
Edad	Un hijo	Dos hijos	Tres hijos	Cuatro hijos	Más de cuatro hijos	Total
De 15 a 21 años	67670	6812	442	40	3	74967
De 21 a 27 años	181720	91605	23958	4448	869	302600
De 27 a 33 años	77403	109838	64296	23087	9768	284392
De 33 a 39 años	29915	61410	56502	27588	22233	197648
De 39 a 45 años	8962	20033	25496	15339	20072	89902
Sin información	1035	1872	2821	2131	4963	12822
Total	366705	291570	173515	72633	57908	962331

Fuente: INEC

Elaboración propia

El gráfico 3.3 muestra los diagramas de caja para cada nivel del factor etnia acorde al número de hijos, donde no se toman en consideración los datos faltantes o sin información. Se observa que el diagrama de cajas del grupo de mujeres que se autodenominan blancas toma como valor máximo 3 hijos e indica que el 75% de sus datos se encuentra aproximadamente por debajo de 2 hijos por mujer. Las demás etnias tienen un mismo comportamiento, su valor máximo es 6 y el tercer cuartil es de 3 hijos por mujer. El valor atípico más alto es de 19 hijos y corresponde al grupo de mujeres que se autodenominan indígenas.

Gráfico 3.3 Diagrama de cajas de la etnia de la madre vs el número de hijos.

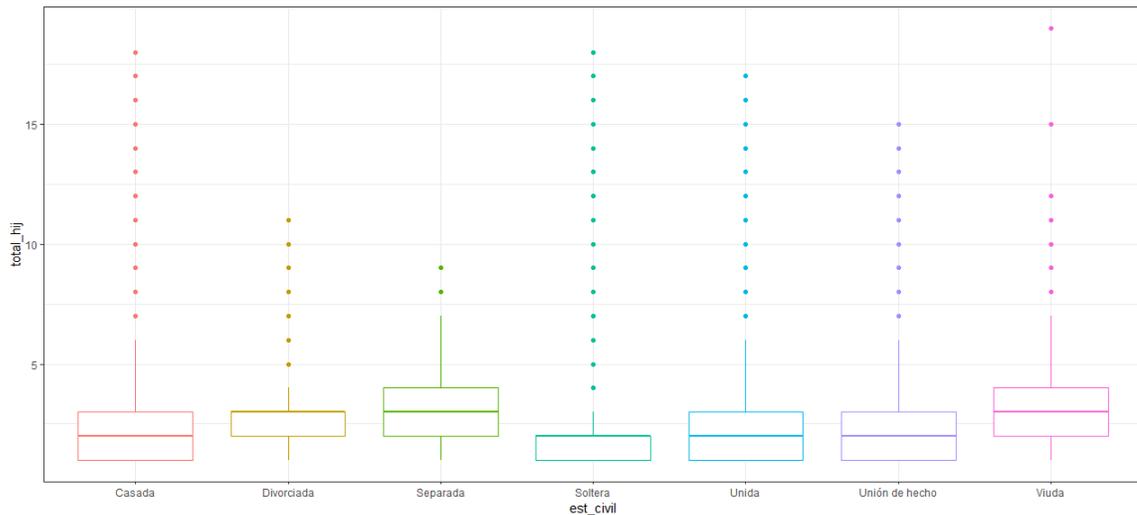


Fuente: INEC

Elaboración propia

En el gráfico 3.4 se observa que si hay diferencias entre las medianas de los grupos del estado civil de la madre. El grupo de mujeres solteras son las que tienen menos hijos porque el 50% y 75% de sus datos coinciden en que tienen 1 o 2 hijos. Además, el 75% de las mujeres separadas y viudas tienen 4 hijos o menos. Y, las casadas, unidas o con unión de hecho muestran un comportamiento similar.

Gráfico 3.4 Diagrama de cajas del estado civil de la madre vs el número de hijos.

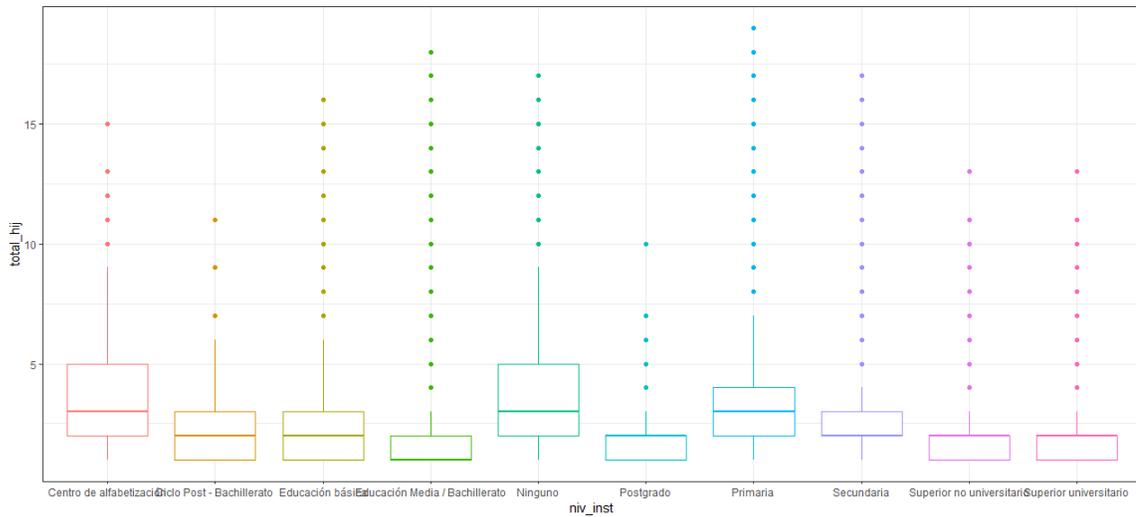


Fuente: INEC

Elaboración propia

En el gráfico 3.5 se observa que las medianas del nivel de instrucción de la madre varían por grupo. Las mujeres con ningún nivel de instrucción o que han estudiado en un centro de alfabetización tienen como máximo 9 hijos, aunque el 75% de ellas tiene 5 o menos. El valor atípico más alto es de 19 hijos y corresponde al grupo de mujeres que han terminado la primaria.

Gráfico 3.5 Diagrama de cajas del nivel de instrucción de la madre vs el número de hijos.

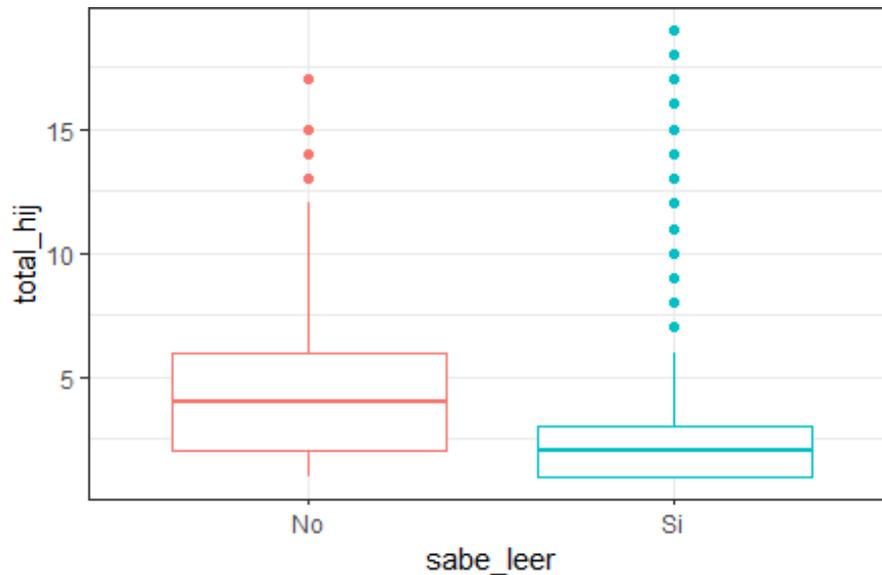


Fuente: INEC

Elaboración propia

El gráfico 3.6 muestra que las medianas entre grupos difieren entre sí. Además de que los bigotes para el grupo de mujeres que no saben leer y escribir son más largos, por lo que el valor máximo de hijos que tienen es 12. A diferencia del grupo de mujeres que si saben leer y escribir cuyo valor máximo es de 6. El valor atípico más alto es de 19 hijos y corresponde al grupo de mujeres que saben leer y escribir. El 50% de las mujeres que saben leer y escribir tienen entre 1 y 3 hijos, mientras que las que no tienen entre 2 y 6 hijos.

Gráfico 3.6 Diagrama de cajas de si la madre sabe leer y escribir vs el número de hijos.

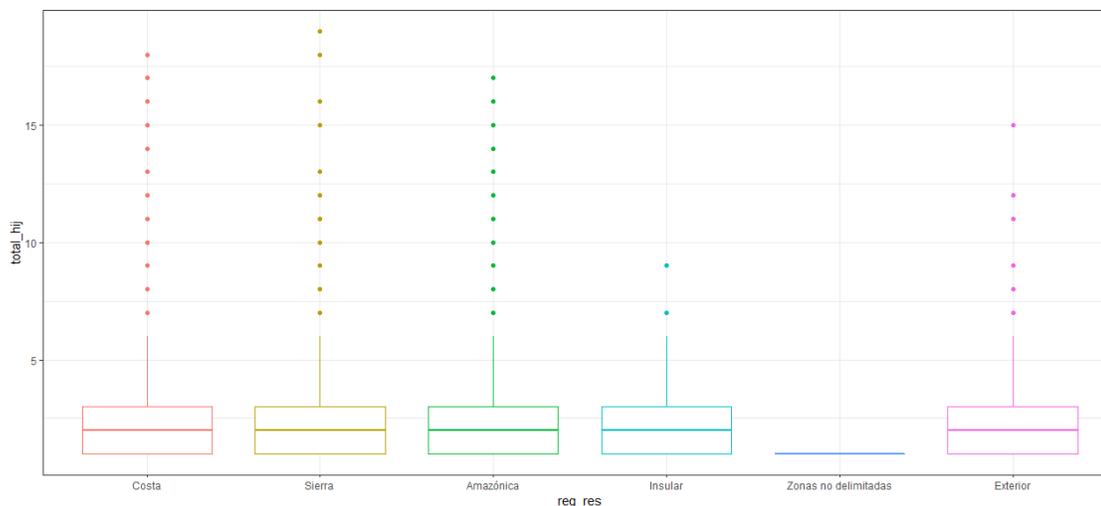


Fuente: INEC

Elaboración propia

El gráfico 3.7 muestra que las medianas entre grupos no difieren entre sí, a excepción del de zonas no delimitadas. De igual forma, los bigotes se extienden hasta el mismo valor, lo cual indica que las mujeres que habitan en la Costa, Sierra, Oriente, Amazonía, Insular o Exterior tienen como máximo 6 hijos. Además, el valor atípico más alto, que es de 19 hijos aproximadamente, se encuentra en el grupo de las mujeres que habitan en la sierra.

Gráfico 3.7 Diagrama de cajas de la región de residencia de la madre vs el número de hijos.

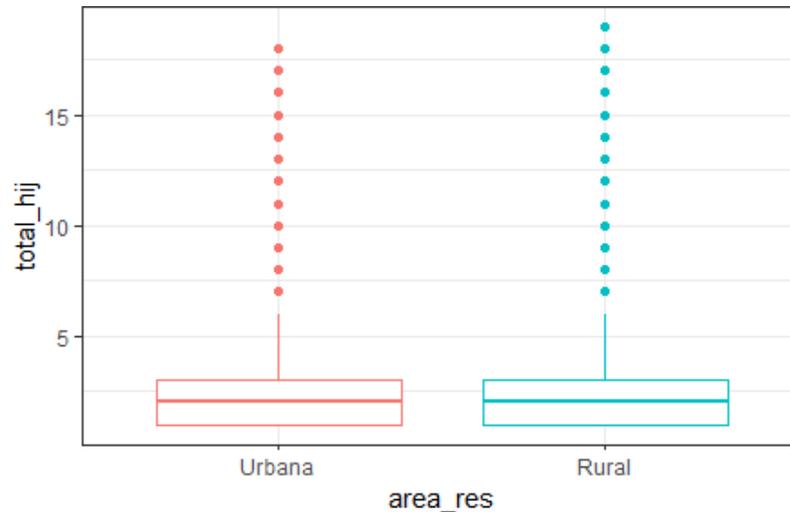


Fuente: INEC

Elaboración propia

El gráfico 3.8 muestra que no hay diferencias significativas entre los grupos de mujeres que habitan en el área urbana o rural. Ambos grupos muestran el mismo comportamiento, donde el 75% de sus datos se encuentran por debajo de 3 hijos por mujer aproximadamente, aunque como máximo tengan 6 hijos. El valor atípico más alto es de 19 hijos y corresponde al grupo de mujeres que habitan en el área rural.

Gráfico 3.8 Diagrama de cajas del área de residencia de la madre vs el número de hijos.



Fuente: INEC

Elaboración propia

3.4 Modelo de Regresión Poisson Zero-Truncado

Como las bases de datos del INEC corresponden a los nacimientos fetales, los datos corresponden a mujeres que han fecundado. Eso quiere decir que la variable de respuesta número de hijos no puede tomar el valor de cero, y por dicha razón se decidió construir un modelo de regresión Zero-Truncado. Para emplear este modelo se recodificaron las variables predictoras etnia, estado civil, nivel de instrucción y región de residencia. Cabe recalcar que la variable dependiente o de conteo es el número de hijos que tiene una mujer y se explica a través de las variables independientes edad, etnia, estado civil, nivel de instrucción, región de residencia y área de residencia. Además, para seleccionar un buen modelo se hizo una prueba Chi cuadrado donde se contrasta el modelo de regresión Poisson Zero-Truncado y un modelo de regresión Binomial Negativa.

En la tabla 3.11 se presentan los coeficientes del mejor modelo: Regresión Poisson Zero-Truncado. Donde se observa que el hecho de que el máximo nivel de instrucción de una mujer sea el centro de alfabetización, no son factores determinantes en el número de hijos que tiene, tomando como nivel de referencia a una mujer sin instrucción. A diferencia del resto de variables predictoras que sí son factores determinantes del modelo.

Tabla 3.11 Estimaciones del modelo de regresión Poisson Zero-Truncado.

Variable	Estimación \hat{Q}	Error estándar	Valor z	Valor p	$e^{\hat{Q}}$
Intercepto	-1.286	0.009	-148.962	< 2e-16	0.28
Edad	0.073	0.000	552.544	< 2e-16	1.08
Nivel de referencia para Etnia: Mestiza					
Etnia: Afrodescendiente	0.120	0.005	25.200	< 2e-16	1.13
Etnia: Blanca	-0.133	0.012	-11.138	< 2e-16	0.88
Etnia: Indígena	0.164	0.003	47.435	< 2e-16	1.18
Etnia: Montubia	0.055	0.010	5.217	1.82E-07	1.06
Etnia: Otra	0.073	0.014	5.275	1.33E-07	1.08
Nivel de referencia para Estado civil: Soltera					
Estado civil: Casada/Unida	0.073	0.002	39.083	< 2e-16	1.08
Estado civil: Divorciada	0.075	0.006	12.501	< 2e-16	1.08
Estado civil: Separada	0.079	0.027	2.933	0.003352	1.08
Estado civil: Viuda	0.121	0.014	8.609	< 2e-16	1.13
Nivel de referencia para Nivel de instrucción: Ninguno					
Nivel de instrucción: Bachillerato	-0.467	0.012	-38.317	< 2e-16	0.63
Nivel de instrucción: Centro de alfabetización	-0.022	0.017	-1.298	0.194165	0.98
Nivel de instrucción: Educación básica	-0.123	0.012	-10.215	< 2e-16	0.88
Nivel de instrucción: Postgrado	-1.107	0.025	-43.752	< 2e-16	0.33
Nivel de instrucción: Secundaria	-0.267	0.012	-22.069	< 2e-16	0.77
Nivel de instrucción: Superior no universitario	-0.494	0.013	-37.688	< 2e-16	0.61
Nivel de instrucción:	-0.828	0.012	-67.324	< 2e-16	0.44

Superior universitario					
Nivel de referencia para Sabe leer: No					
Sabe leer: Si	-0.043	0.013	-3.301	9.63E-04	0.96
Nivel de referencia para Región de residencia: Costa					
Región de residencia: Amazónica	0.077	0.003	22.300	< 2e-16	1.08
Región de residencia: Otra	-0.153	0.021	-7.446	9.61E-14	0.86
Región de residencia: Sierra	-0.203	0.002	-100.484	< 2e-16	0.82
Nivel de referencia para Área de residencia: Rural					
Área de residencia: Urbana	-0.047	0.002	-22.129	< 2e-16	0.95

Fuente: Elaboración propia

Además, también se obtuvo los siguientes datos del modelo:

- Log - likelihood: -1173,281
- Grados de libertad: 929,399

En la tabla 3.11 se observa que el promedio del número de hijos de una mujer autodenominada mestiza, cuyo estado civil es soltera, carece de educación, que no sabe leer ni escribir, de la región Costa y que habita en el área rural es de 0.28.

Si una mujer se autodenomina afrodescendiente (grupo que incluye a los mulatos y negros), su media del número de hijos es 1.13 veces la media de una mestiza, manteniendo el resto de las variables constantes.

Si una mujer se autodenomina blanca, su media del número de hijos es 0.88 veces la media de una mestiza, manteniendo el resto de las variables constantes.

Si una mujer se autodenomina indígena, su media del número de hijos es 1.18 veces la media de una mestiza, manteniendo el resto de las variables constantes.

Si una mujer se autodenomina montubia, su media del número de hijos es 1.06 veces la media de una mestiza, manteniendo el resto de las variables constantes.

Si una mujer se autodenomina de otra raza, su media del número de hijos es 1.08 veces la media de una mestiza, manteniendo el resto de las variables constantes.

Si una mujer es casada o unida, su media del número de hijos es 1.08 veces la media de una mujer soltera, manteniendo el resto de las variables constantes.

Si una mujer es divorciada, su media del número de hijos es 1.08 veces la media de una mujer soltera, manteniendo el resto de las variables constantes. El mismo comportamiento presentan las mujeres que se han separado.

Si una mujer se ha quedado viuda, su media del número de hijos es 1.13 veces la media de una mujer soltera, manteniendo el resto de las variables constantes.

Para una mujer cuyo máximo nivel de instrucción es el centro de alfabetización, su media del número de hijos será similar a la de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer que tiene un nivel de instrucción básico, su media del número de hijos es 0.88 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer que tiene un nivel de instrucción medio o de bachillerato, su media del número de hijos es 0.63 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer que tiene un postgrado, su media del número de hijos es 0.33 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer cuyo nivel de instrucción más alto es la secundaria, su media del número de hijos es 0.77 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer cuyo nivel de instrucción más alto es superior no universitario, su media del número de hijos es 0.61 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Para una mujer cuyo nivel de instrucción más alto es superior universitario, su media del número de hijos es 0.44 veces la media de una mujer que carece de educación, manteniendo el resto de las variables constantes.

Si una mujer sabe leer y escribir, su media del número de hijos es 0.96 veces la media de que no posee dicha capacidad, manteniendo el resto de las variables constantes.

Para una mujer cuya región de residencia es la Amazonía, su media del número de hijos es 1.08 veces la media de una que reside en la Costa, manteniendo el resto de las variables constantes.

Para una mujer que reside en otra región, su media del número de hijos es 0.86

veces la media de una que reside en la Costa, manteniendo el resto de las variables constantes.

Para una mujer que reside en la región Sierra, su media del número de hijos es 0.82 veces la media de una que reside en la Costa, manteniendo el resto de las variables constantes.

Si una mujer reside en el área urbana, su media del número de hijos es 0.95 veces la media de una que habita en el área rural, manteniendo el resto de las variables constantes.

3.5 Modelos de Regresión Logística

Se construyeron dos modelos de regresión logística donde la variable dependiente número de hijos (dicotómica) se explica a través de las variables independientes edad, etnia, estado civil, nivel de instrucción, región de residencia y área de residencia. La diferencia entre ambos modelos es que en el primero se emplearon las variables con los niveles provenientes del INEC, pero en el segundo se recodificaron agrupando a las variables cuya respuesta salió similar. Al implementar los modelos se estableció como éxito de ocurrencia que una mujer tenga dos hijos o menos. Además, se construyó un tercer modelo donde se aplica regresión logística multinomial donde los grupos son “Un hijo”, “Dos hijos”, “Tres hijos”, “Cuatro hijos” y “Más de cuatro hijos”.

3.5.1 Modelo de Regresión Logística Binaria (Variables Originales)

En la tabla 3.12 se presentan los coeficientes del modelo empleado y los valores de los odds ratio. Si los odds ratio son mayor que uno, la propensión de ocurrencia del evento es mayor en relación con la categoría de referencia. Lo contrario a si son menor que uno, donde la propensión de ocurrencia disminuye. Donde la probabilidad de fallo se define como $P(Y = 0) =$

"Más de dos hijos"; y, la probabilidad de éxito se define como $P(Y = 1) =$
 "Dos hijos o menos".

Tabla 3.12 Estimaciones del modelo de regresión logística binaria con las variables originales.

Variable	Estimación \hat{Q}	Error estándar	Valor z	Valor p	$e^{\hat{Q}}$
Intercepto	5.909	0.039	150.923	< 2e-16	368.241
Edad	-0.208	0.001	-377.698	< 2e-16	0.812
Nivel de referencia para Etnia: Mestiza					
Etnia: Afroecuatoriana/ Afrodescendiente	-0.267	0.024	-10.949	< 2e-16	0.766
Etnia: Blanca	0.376	0.035	10.785	< 2e-16	1.457
Etnia: Indígena	-0.294	0.012	-23.892	< 2e-16	0.745
Etnia: Montubia	-0.156	0.035	-4.476	7.61E-06	0.855
Etnia: Mulata	-0.221	0.040	-5.492	3.98E-08	0.802
Etnia: Negra	-0.263	0.023	-11.371	< 2e-16	0.769
Etnia: Otra	-0.059	0.045	-1.308	1.91E-01	0.942
Nivel de referencia para Estado civil: Soltera					
Estado civil: Casada	-0.143	0.007	-20.393	< 2e-16	0.867
Estado civil: Divorciada	-0.377	0.021	-18.359	< 2e-16	0.686
Estado civil: Separada	-0.284	0.097	-2.917	0.00354	0.753
Estado civil: Unión de hecho	-0.155	0.018	-8.613	< 2e-16	0.857
Estado civil: Unida	-0.153	0.007	-21.878	< 2e-16	0.858
Estado civil: Viuda	-0.836	0.066	-12.704	< 2e-16	0.433
Nivel de referencia para Nivel de instrucción: Ninguno					
Nivel de instrucción: Centro de alfabetización	0.087	0.075	1.151	0.24964	1.091
Nivel de instrucción: Ciclo Post - Bachillerato	1.733	0.092	18.763	< 2e-16	5.659
Nivel de instrucción: Educación básica	0.452	0.052	8.712	< 2e-16	1.571
Nivel de instrucción: Educación Media / Bachillerato	1.082	0.052	20.912	< 2e-16	2.949
Nivel de instrucción: Postgrado	3.206	0.079	40.432	< 2e-16	24.676
Nivel de instrucción:	-0.016	0.052	-0.309	7.57E-01	0.984

Primaria					
Nivel de instrucción: Secundaria	0.732	0.052	14.188	< 2e-16	2.080
Nivel de instrucción: Superior no universitario	1.280	0.054	23.899	< 2e-16	3.598
Nivel de instrucción: Superior universitario	2.366	0.052	45.454	< 2e-16	10.659
Nivel de referencia para Sabe leer: No					
Sabe leer: Si	0.173	0.058	2.969	2.99E-03	1.188
Nivel de referencia para Región de residencia: Costa					
Región de residencia: Amazónica	-0.082	0.012	-7.041	1.90E-12	0.921
Región de residencia: Exterior	0.257	0.124	2.064	0.03904	1.293
Región de residencia: Insular	0.532	0.072	7.377	1.62E-13	1.702
Región de residencia: Sierra	0.592	0.006	92.996	< 2e-16	1.808
Región de residencia: Zonas no delimitadas	5.220	16.181	0.323	0.74701	184.884
Nivel de referencia para Área de residencia: Rural					
Área de residencia: Urbana	0.071	0.007	10.389	< 2e-16	1.074

Fuente: Elaboración propia

Además, también se obtuvo los siguientes datos del modelo:

- Null deviance: 1161,468
- Residual deviance: 839,084
- Pseudo R^2 : 0.2776
- AIC : 839,146

Se observa en la tabla 3.12 que el hecho de que una mujer se autodefiniera con otra etnia, su máximo nivel de instrucción sea el centro de alfabetización o la primaria, o habite en zonas no delimitadas, no influye significativamente en tener dos hijos o menos. A diferencia del resto de variables predictoras que sí influyen significativamente en la variable de respuesta.

Las mujeres que se autodenominan afrodescendientes, indígenas, montubias, mulatas, negras o de otra raza tienen menor propensión de tener dos hijos o menos a comparación de las que se autodenominan mestizas. Mientras que la propensión aumenta para las mujeres que se autodenominan blancas.

La propensión de que una mujer casada, divorciada, separada, unida, viuda o en unión de hecho tenga dos hijos o menos es más baja que la de una mujer soltera.

Teniendo como referencia el nivel de instrucción ninguno, la variable que reduce la propensión de tener dos hijos o menos es la primaria. Al contrario de las mujeres que tienen cualquier otro tipo de educación cuya propensión aumenta.

Las mujeres que saben leer y escribir tienen mayor propensión de tener dos hijos o menos que las que no.

La propensión de tener dos hijos o menos aumenta para las mujeres que habitan en la región Sierra, Insular, exterior y zonas no delimitadas, con respecto a la región Costa. Mientras que la propensión disminuye para las mujeres que habitan en la Amazonía.

Las mujeres que habitan en el área urbana tienen mayor propensión de tener dos hijos o menos que las que habitan en el área rural.

Cuadro 3.1 Análisis de devianza para el modelo con las variables originales.

```
Analysis of Deviance Table
Model: binomial, link: logit
Response: total_hij
Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                929421    1161468
edad_mad  1    230176    929420     931292 < 2e-16 ***
etnia     7     6150    929413     925143 < 2e-16 ***
est_civil 6     5500    929407     919643 < 2e-16 ***
niv_inst  9    70501    929398     849141 < 2e-16 ***
sabe_leer 1         6    929397     849136 0.01705 *
reg_res   5     9944    929392     839191 < 2e-16 ***
area_res  1      108    929391     839084 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fuente: Elaboración propia

El cuadro 3.1 muestra que todas las variables del modelo son significativas, o sea que su inclusión en el modelo permite explicar el número de hijos que tiene una mujer. También se hizo una prueba del ratio de verosimilitud donde se obtuvo un valor $p = 0$, lo cual indica que el modelo si es significativo y por lo tanto se ajusta correctamente a los datos.

3.5.2 Modelo de Regresión Logística Binaria (Variables Recodificadas)

En la tabla 3.13 se presentan los coeficientes del modelo empleado y los valores de los odds ratio. Si los odds ratio son mayor que uno, la propensión de ocurrencia del evento es mayor en relación con la categoría de referencia. Pero si son menor que uno, la propensión de ocurrencia disminuye. Donde la probabilidad de fallo se define como $P(Y = 0) = \text{"Más de dos hijos"}$; y, la probabilidad de éxito se define como $P(Y = 1) = \text{"Dos hijos o menos"}$.

Tabla 3.13 Estimaciones del modelo de regresión logística con las variables recodificadas.

Variable	Estimación \hat{Q}	Error estándar	Valor z	Valor p	$e^{\hat{Q}}$
Intercepto	6.080	0.039	156.285	< 2e-16	436.98
Edad	-0.213	0.001	-403.565	< 2e-16	0.81
Nivel de referencia para Etnia: Mestiza					
Etnia: Afrodescendiente	-0.254	0.016	-16.238	< 2e-16	0.78
Etnia: Blanca	0.381	0.035	10.918	< 2e-16	1.46
Etnia: Indígena	-0.298	0.012	-24.428	< 2e-16	0.74
Etnia: Montubia	-0.155	0.035	-4.477	7.56E-06	0.86
Etnia: Otra	-0.062	0.045	-1.379	1.68E-01	0.94
Nivel de referencia para Estado civil: Soltera					
Estado civil: Casada/Unida	-0.156	0.006	-27.272	< 2e-16	0.86
Estado civil: Divorciada	-0.370	0.021	-17.999	< 2e-16	0.69
Estado civil: Separada	-0.294	0.097	-3.039	0.00237	0.75
Estado civil: Viuda	-0.838	0.066	-12.729	< 2e-16	0.43
Nivel de referencia para Nivel de instrucción: Ninguno					
Nivel de instrucción: Bachillerato	1.085	0.052	20.851	< 2e-16	2.96
Nivel de instrucción: Centro de alfabetización	0.099	0.076	1.305	0.19178	1.10
Nivel de instrucción: Educación básica	0.229	0.052	4.405	1.06E-05	1.26
Nivel de instrucción: Postgrado	3.270	0.080	41.098	< 2e-16	26.30
Nivel de instrucción: Secundaria	0.754	0.052	14.511	< 2e-16	2.12
Nivel de instrucción: Superior no universitario	1.315	0.054	24.440	< 2e-16	3.73
Nivel de instrucción: Superior universitario	2.412	0.052	46.075	< 2e-16	11.16
Nivel de referencia para Sabe leer: No					
Sabe leer: Si	0.152	0.058	2.609	9.08E-03	1.16
Nivel de referencia para Región de residencia: Costa					
Región de residencia: Amazónica	-0.074	0.012	-6.388	1.68E-10	0.93
Región de residencia: Otra	0.460	0.062	7.387	1.51E-13	1.58
Región de residencia: Sierra	0.588	0.006	94.051	< 2e-16	1.80
Nivel de referencia para Área de residencia: Rural					
Área de residencia: Urbana	0.077	0.007	11.200	< 2e-16	1.08

Fuente: Elaboración propia

Además, también se obtuvo los siguientes datos del modelo:

- Null deviance: 1161,468
- Residual deviance: 841,073
- Pseudo R^2 : 0.2759
- *AIC*: 841,119

En la tabla 3.13 se observa que una mujer que se autodefine con otra etnia, su máximo nivel de instrucción es el centro de alfabetización o la primaria, o habita en zonas no delimitadas, no son factores determinantes en el número de hijos que tiene.

Además, las mujeres que se autodenominan afrodescendientes, indígenas, montubias, o de otra raza tienen menor propensión de tener dos hijos o menos a comparación de las que se autodenominan mestizas. Mientras que la propensión aumenta para las mujeres que se autodenominan blancas.

La propensión de que una mujer casada o unida, divorciada, separada o viuda tenga dos hijos o menos es más baja que la de una mujer soltera.

Teniendo como referencia el nivel de instrucción ninguno, las mujeres que poseen cualquier tipo de educación tienen mayor propensión de tener dos hijos o menos.

Las mujeres que saben leer y escribir tienen mayor propensión de tener dos hijos o menos que las que no.

La propensión de tener dos hijos o menos aumenta para las mujeres que habitan en la Sierra u otra región, con respecto a la región Costa. Mientras que disminuye para las mujeres que habitan en la Amazonía.

Las mujeres que habitan en el área urbana tienen mayor propensión de tener dos hijos o menos que las que habitan en el área rural.

Cuadro 3.2 Análisis de devianza para el modelo con variables recodificadas.

```

Analysis of Deviance Table

Model: binomial, link: logit

Response: total_hij

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                929421    1161468
edad_mad  1   230176    929420    931292 < 2e-16 ***
etnia     5     6135    929415    925157 < 2e-16 ***
est_civil 4     282    929411    924875 < 2e-16 ***
niv_inst  7   73510    929404    851365 < 2e-16 ***
sabe_leer 1      5    929403    851360 0.02379 *
reg_res   3   10162    929400    841198 < 2e-16 ***
area_res  1    125    929399    841073 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Fuente: Elaboración propia

El cuadro 3.2 muestra que todas las variables del modelo son significativas, o sea que su inclusión en el modelo permite explicar el número de hijos que tiene una mujer. También se hizo una prueba del ratio de verosimilitud donde se obtuvo un valor $p = 0$, lo cual indica que el modelo si es significativo y por lo tanto se ajusta correctamente a los datos.

3.5.3 Modelo de Regresión Logística Multinomial

En la tabla 3.14 se presentan los coeficientes del modelo para cada variable según el grupo del número de hijos, donde el nivel de referencia es “Más de cuatro hijos”.

Tabla 3.14 Coeficientes del modelo de regresión logística multinomial.

	Un hijo	Dos hijos	Tres hijos	Cuatro hijos
Intercepto	12.95952	7.55631	4.23335	2.11684
Edad	-0.43886	-0.25498	-0.14562	-0.08273
Afrodescendiente	-0.72614	-0.74246	-0.64519	-0.44024
Blanca	0.59539	0.28964	0.12389	-0.21042
Indígena	-0.87474	-0.92066	-0.83952	-0.55408
Montubia	-0.18736	-0.14020	-0.11685	-0.25566
Otra etnia	-0.22173	-0.21740	-0.18905	-0.16029
Casada/Unida	-0.21202	0.06977	0.12121	0.06417
Divorciada	-0.45815	0.10586	0.39452	0.32872
Separada	1.34801	0.99542	1.05230	1.10763
Viuda	-1.59738	-0.94081	-0.43062	-0.18358
Bachillerato	2.50194	2.17400	1.82421	0.95655
Centro de alfabetización	0.23508	0.22217	0.36728	-0.27321
Educación básica	0.75033	0.78215	0.90308	0.43286
Postgrado	6.22327	4.24443	2.73661	0.80071
Secundaria	1.91994	1.83639	1.64629	0.83201
Superior no universitario	3.21159	2.78060	2.30978	1.17059
Superior universitario	5.46113	4.27390	3.17169	1.53280
Sabe leer	-0.05601	0.38041	0.15855	0.22829
Amazónica	-0.40080	-0.40489	-0.42302	-0.35856
Otra región	1.01886	0.91936	0.56640	0.21009
Sierra	1.20915	0.98572	0.63131	0.33505
Urbana	0.32349	0.30065	0.26546	0.19625

Fuente: Elaboración propia

Además, también se obtuvo los siguientes datos del modelo:

- Residual deviance: 2102,527

- AIC: 2102,711

Tabla 3.15 Valores p del modelo de regresión logística multinomial.

	Un hijo	Dos hijos	Tres hijos	Cuatro hijos
Intercepto	0.00	0.00	0.00	0.00
Edad	0.00	0.00	0.00	0.00
Afrodescendiente	0.00	0.00	0.00	0.00
Blanca	0.00	0.00	0.10	0.02
Indígena	0.00	0.00	0.00	0.00
Montubia	0.00	0.02	0.05	0.00
Otra etnia	0.01	0.01	0.02	0.08
Casada/Unida	0.00	0.00	0.00	0.00
Divorciada	0.00	0.00	0.00	0.00
Separada	0.00	0.00	0.00	0.00
Viuda	0.00	0.00	0.00	0.02
Bachillerato	0.00	0.00	0.00	0.00
Centro de alfabetización	0.04	0.03	0.00	0.01
Educación básica	0.00	0.00	0.00	0.00
Postgrado	0.00	0.00	0.00	0.00
Secundaria	0.00	0.00	0.00	0.00
Superior no universitario	0.00	0.00	0.00	0.00
Superior universitario	0.00	0.00	0.00	0.00
Sabe leer	0.53	0.00	0.04	0.00
Amazónica	0.00	0.00	0.00	0.00
Otra región	0.00	0.00	0.00	0.17
Sierra	0.00	0.00	0.00	0.00
Urbana	0.00	0.00	0.00	0.00

Fuente: Elaboración propia

Tabla 3.16 Odds ratio del modelo de regresión logística multinomial.

	Un hijo	Dos hijos	Tres hijos	Cuatro hijos
Intercepto	424860.800	1912.777	68.948	8.305
Edad	0.645	0.775	0.864	0.921
Nivel de referencia: Mestiza				
Afrodescendiente	0.484	0.476	0.525	0.644
Blanca	1.814	1.336	1.132	0.810
Indígena	0.417	0.398	0.432	0.575

Montubia	0.829	0.869	0.890	0.774
Otra etnia	0.801	0.805	0.828	0.852
Nivel de referencia: Soltera				
Casada/Unida	0.809	1.072	1.129	1.066
Divorciada	0.632	1.112	1.484	1.389
Separada	3.850	2.706	2.864	3.027
Viuda	0.202	0.390	0.650	0.832
Nivel de referencia: Ninguno				
Bachillerato	12.206	8.793	6.198	2.603
Centro de alfabetización	1.265	1.249	1.444	0.761
Educación básica	2.118	2.186	2.467	1.542
Postgrado	504.349	69.716	15.435	2.227
Secundaria	6.821	6.274	5.188	2.298
Superior no universitario	24.819	16.129	10.072	3.224
Superior universitario	235.363	71.801	23.848	4.631
Nivel de referencia: No				
Sabe leer	0.946	1.463	1.172	1.256
Nivel de referencia: Costa				
Amazónica	0.670	0.667	0.655	0.699
Otra región	2.770	2.508	1.762	1.234
Sierra	3.351	2.680	1.880	1.398
Nivel de referencia: Rural				
Urbana	1.382	1.351	1.304	1.217

Fuente: Elaboración propia

En la tabla 3.16 las variables que se encuentran en negrita, así como los valores resaltados, son las variables cuyos coeficientes resultaron negativos en la tabla 3.14 para el modelo planteado, así mismo se muestra la celda en gris cuando esa opción de respuesta no fue estadísticamente significativa.

La propensión a tener un hijo de las mujeres blancas es casi dos veces (1.8) por sobre las que tienen más de cuatro hijos y son mestizas y a medida que aumentan el número de hijos esta propensión disminuye.

A diferencia de las mujeres indígenas y negras en las que la propensión a tener un solo hijo disminuye en un 0.42 y 0,49 respectivamente en comparación con las que tienen más de cuatro hijos y son mestizas.

Las mujeres indígenas disminuyen su propensión a tener 1, 2, 3 y 4 hijos en alrededor de 0.8 en comparación con las que tienen más de 4 hijos y son mestizas.

Las mujeres casadas o unidas que tienen dos, tres y cuatro hijos tienen la misma propensión en comparación con las que tienen más de 4 hijos y son solteras.

Las mujeres divorciadas y separadas tienen mayor propensión a tener 2, 3, 4 hijos siendo la propensión mayor en las separadas, todo esto en comparación con las mujeres solteras que tienen más de 4 hijos.

Las mujeres de la región amazónica al compararlas con las que tienen más de cuatro hijos y que viven en la costa tienen menor propensión a tener 1, 2, 3 o 4 hijos.

En las mujeres más educadas, su propensión a tener 1 solo hijo es muy alta al compararla con tener 4 o más hijos y no tener ninguna instrucción.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

La metodología explicada en el capítulo 3 permitió la construcción de diferentes modelos no lineales. Por consiguiente, en este capítulo se exhiben las conclusiones y recomendaciones de los resultados obtenidos sobre el comportamiento del número de hijos que tiene una mujer ecuatoriana de entre 15 a 49 años en función de la edad, del estado marital, autodefinición étnica, educación y lugar de residencia para el período del 2015 al 2019.

4.1 Conclusiones

- Entre las mujeres que han tenido hijos durante el periodo del 2015 al 2019, y que en promedio tienen 2 ± 1 hijos, se obtuvo que el 43.90% son madres solteras, casi el 100% sabe leer y escribir, el 0.08% carecen de un nivel de instrucción, mientras que el 18% tiene una instrucción superior, el mayor porcentaje se concentra en niveles de educación que comprenden las etapas de la escuela y el colegio. Mayoritariamente viven en la zona urbana y se consideran mestizas.
- Al aplicar los diferentes modelos de regresión a los datos y en particular, luego de realizar un modelo lineal generalizado Poisson para estudiar el comportamiento de los datos se observó infradispersión, lo cual ocurre porque los conteos de la variable de respuesta empiezan desde uno. Por ende, se empleó un modelo de regresión Zero-Truncado.
- El modelo de regresión Poisson Zero-Truncado muestra que el número de hijos que tiene una mujer durante su vida reproductiva se ve influenciado por

su edad, etnia, estado civil, nivel de instrucción, si sabe leer y escribir, región y área donde reside.

- De este mismo modelo se tiene que las mujeres indígenas, afrodescendientes y montubias, en ese orden tienen en promedio mayor probabilidad de tener más hijos que las mujeres mestizas resultados que también se confirman en los modelos de regresión logística aplicados.
- La media del número de hijos de mujeres cuyo nivel más alto de instrucción es el centro de alfabetización es similar a la de las mujeres que no tienen ningún nivel de instrucción. Lo cual tiene sentido porque el centro de alfabetización es un nivel muy bajo de instrucción, de hecho, es un programa dedicado a las personas con analfabetismo.
- El modelo de regresión logística binaria indica que las mismas variables predictoras del modelo de regresión Poisson son significativas. Esto es, edad, etnia, estado civil, nivel de instrucción, si sabe leer y escribir, región y área donde reside.
- El análisis de devianza demuestra que el modelo logístico representa correctamente a los datos.
- Del análisis de odds ratio se obtuvo que las mujeres blancas tienen mayor propensión de tener dos hijos o menos que las mestizas. Lo contrario a las mujeres de las demás razas que tienen una propensión más baja, significando esto, que las otras etnias tienen mayor propensión a tener más de dos hijos en comparación con las mestizas.
- Las mujeres ecuatorianas solteras tienen mayor propensión de tener dos hijos o menos que las que tienen otro estado civil como casada o unida, divorciada, etc.
- Las mujeres que poseen cualquier nivel de instrucción tienen mayor propensión de tener dos hijos o menos a comparación de las que no tienen ninguno.
- Las mujeres que no saben leer ni escribir tienen mayor propensión de tener 3 hijos o más durante su vida reproductiva.

- Es menos propenso que las mujeres que habitan en la Amazonía tengan dos hijos o menos, referenciándolas con las de la Costa. Ellas tienen mayor propensión a tener más de dos hijos.
- Las mujeres del área rural tienen mayor propensión de tener tres hijos o más que las del área urbana.
- Para comparar modelos basándose en los criterios AIC o BIC, los modelos deben tener el mismo tipo de variable de respuesta. En este estudio la variable de respuesta no es la misma porque se aplicó una transformación para el modelo logístico.
- De la aplicación de regresión logística Multinomial se aprecia que las mujeres negras, indígenas, montubias o de otra etnia tienen una menor propensión a tener uno o dos hijos tomado como referencia tener 4 hijos o más y ser mestiza.
- Para las mujeres más educadas, su propensión a tener 1 solo hijo es muy alta al compararla con tener 4 o más hijos y no tener ninguna instrucción.
- Destaca la región Sierra con propensión alta para tener 1 hijo o dos por sobre tener 4 o más y vivir en la Costa.
- Todos los modelos se ajustan correctamente a los datos, por lo que la implementación de cualquiera de ellos es correcta.

4.2 Recomendaciones

- Las mujeres ecuatorianas en edad fértil en el periodo considerado muestra que el número de hijos que se tiene se polariza a tener más o menos hijos según la etnia, los estudios y el estatus marital, mostrando realidades sociales para los diferentes grupos y diferentes elecciones, solapando por ejemplo a indígenas y negras, solteras y menos educadas con un mayor número de hijos. Esto da muestra de los problemas sociales a resolver que tiene Ecuador en este sentido.
- Se deben emplear bases de datos de calidad, provenientes de fuentes confiables, para poder tener un análisis más representativo de la población.

Siempre es posible mejorar los registros administrativos de las instituciones públicas.

- En términos del análisis estadístico realizado: es recomendable hacer un análisis descriptivo de los datos para comprender el comportamiento de las variables, sobre todo para la variable de respuesta porque de esa forma se puede intuir el tipo de modelo emplear.
- Si al realizar un modelo Poisson los datos presentan sobredispersión, se puede probar un modelo quasipoisson para reducir la varianza de los datos.
- Si al realizar un modelo Poisson los datos presentan infradispersión, se puede probar un modelo pospoisson para que se ajuste mejor a los datos.

BIBLIOGRAFÍA

- Brann, E. (1974). A Multivariate Analysis of Interstate Variation in Fertility of Teenage Girls. *American Journal of Public Health*, 2-4.
- Buizza, C., & Villafuerte, A. (2020). *La evolución de la fecundidad en Ecuador: la necesidad de nuevas políticas públicas*. Retrieved from <http://foroeconomiaecuador.com/fee/la-evolucion-de-la-fecundidad-en-ecuador-la-necesidad-de-nuevas-politicas-publicas/>
- Canto, Y. E., Ortiz, K. J., & Ortiz, Y. J. (2018). Efecto de las preferencias de fecundidad en la cantidad de hijos. Comparación entre trabajadoras peruanas. *Scielo Perú*.
- Chackiel, J. (2004). *La transición de la fecundidad en América Latina 1950-2000*. Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252004000300002
- Chowdury, M. H., & Moloy, D. J. (2014). A statistical analysis to identify potential factors of fertility in Bangladesh. *Journal of Science and Technology*.
- Dueñas, M. A. (2010). Modelos de respuesta discreta en R y aplicación con datos reales. *Universidad de Granada*.
- Eggleston, E. (1999). Causas Determinantes de Embarazos No Planeados en el Ecuador. *Int Fam Plan Perspect*.
- FACTANK. (2019, Mayo). *Is U. S. fertility at an all-time low? Two of three measures to point yes*. Retrieved from <https://www.pewresearch.org/fact-tank/2019/05/22/u-s-fertility-rate-explained/>
- Huamaní, M. A. (2014). Identificación de un modelo explicativo de retención de clientes con riesgo de fuga para una entidad bancaria aplicando regresión logística y árboles de clasificación Cart. *Universidad Nacional Agraria La Molina*.
- Iglesias, T. (2012). Métodos de Bondad de Ajuste en Regresión Logística. *Universidad de Granada*.
- INEC. (2001). Comportamiento Reproductivo de las Mujeres Ecuatorianas. *Estudios Demográficos en Profundidad*.
- INEC. (2017). Fecundidad. *INEC*.
- Ishida, K., Stupp, P., & Ordoñez, J. (2011). Estancamiento de la disminución de la fecundidad en Ecuador. *Guttmacher*.
- Knoema. (2020). *Ecuador - Tasa total de fecundidad*. Retrieved from <https://knoema.es/atlas/Ecuador/topics/Datos-demogr%C3%A1ficos/Fecundidad/Tasa-de-fecundidad>

- Lanchimba, C., & Medina, P. (2007). Fecundidad en el Ecuador y su relación con el entorno social y evolutivo. *INEC*.
- Mashfiqul Huq, C., & Deluar, M. J. (2014). A statistical analysis to identify potential factors of fertility in Bangladesh. *Journal of Science and Technology*, 65-78.
- Montalvo, A. G., & Zurita, G. (2012). Estadística y Distribución Espacial de las mujeres en edad reproductiva en el Ecuador. *ESPOL*.
- NCSS. (n.d.). *NCSS Statistical Software*. Retrieved from Poisson Regression: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Poisson_Regression.pdf
- Ortega, M. P. (2018). Fecundidad y natalidad en la comunidad Tanguarín de la parroquia San Antonio de Imbabura de la ciudad de Ibarra durante el periodo noviembre 2016 - mayo 2017.
- Ortíz, E. (2006). Factores Demográficos y Socioeconómicos de la Fecundidad Adolescente en México, 2000.
- Pandey, R., & Kaur, C. (2015). Modelling fertility: an application of count regression models. *Chinese Journal of Population Resources and Environment*.
- Pérez, J. (2010). *Fecundidad y Fertilidad*. Retrieved from Apuntes de Demografía: <https://apuntesdedemografia.com/2010/07/04/fecundidad-y-fertilidad/>
- Poma, S., & Mercedes, C. (n.d.). *Estimación de la ocurrencia de incidencias en declaraciones de pólizas de importación*.
- The Pennsylvania State University. (2018). *Logistic Regression*. Retrieved from PennState Eberly College of Science: <https://online.stat.psu.edu/stat462/node/207/>
- The Pennsylvania State University. (2018). *Poisson Regression*. Retrieved from PennState Eberly College of Science: <https://online.stat.psu.edu/stat462/node/209/>
- The World Bank. (2019). *Fertility rate, total (births per woman)*. Retrieved from The World Bank: <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>
- The World Bank. (2021). *Fertility rate, total (births per woman)*. The World Bank.
- UCLA. (2016, August 22). *Zero-Truncated Poisson / R Data Analysis Examples*. Retrieved from UCLA: Institute for Digital Research & Education Statistical Consulting: <https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/>
- Van der Heijden, P. G., Cruyff, M., & Van Houwelingen, H. C. (2003). Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statistica Neerlandica*, 289-304.
- Vargas, G. (2001). Políticas de Población y Transición Demográfica en México. *Economía y Sociedad*.
- Venables, W., & Smith, D. (2021). An Introduction to R.

