



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIDAD EN ESTADÍSTICA APLICADA

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

La fecundidad en México y su relación con los factores sociodemográficos

## TESINA

QUE PARA OPTAR POR EL GRADO DE:  
ESPECIALISTA EN ESTADÍSTICA APLICADA

PRESENTA:  
LEONARDO AILINES GENIS

Directora de Tesina  
M. en E. Leticia Eugenia Gracia-Medrano Valdelamar  
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

Ciudad Universitaria. Cd. De México, 23 de noviembre de 2018



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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



*A MI MADRE Y A MI PADRE,  
por ser los pilares más importantes de mi vida, y por brindarme su apoyo  
y su cariño en los momentos más difíciles*



# Agradecimientos

*A la Universidad Nacional Autónoma de México*, por darme la oportunidad una vez más de formar parte de ella, y por cada momento que he vivido en sus instalaciones, los cuales fueron únicos e irrepetibles.

*A la Mtra. Leticia Gracia-Medrano*, gracias por su apoyo constante y por impulsarme a terminar este proyecto.

*A la Dra. Silvia Ruiz-Velasco, a la Dra. Lizbeth Naranjo, al Dr. Carlos Díaz y al Mtro. Javier Santibañez*, muchas gracias por su valiosa ayuda en este proyecto tan importante en mi vida y que sin duda alguna no lo podría haber alcanzado sin su apoyo.

*A mis profesores*, gracias por brindarme las herramientas necesarias para seguir creciendo profesionalmente en el área que tanto me apasiona.



# Índice general

Capítulos	Página
<b>Introducción</b>	<b>III</b>
<b>1. La fecundidad y la transición demográfica</b>	<b>1</b>
<b>2. Marco teórico</b>	<b>5</b>
<b>3. Fuente de información y metodología</b>	<b>9</b>
3.1. La Encuesta Nacional de la Dinámica Demográfica, 2014 . . . . .	9
3.2. Modelos estadísticos para el análisis de variables de conteo . . . . .	10
3.2.1. Modelo de Regresión Poisson . . . . .	11
3.2.2. Modelo de Regresión Binomial Negativa . . . . .	11
3.2.3. Modelo Poisson Inflado por Ceros ( <i>Zero Inflated Poisson</i> ) . . . . .	13
<b>4. Validación y obtención de resultados</b>	<b>15</b>
4.1. Análisis descriptivo . . . . .	15
4.2. Selección del modelo . . . . .	24
4.3. Resultados . . . . .	30
<b>5. Conclusiones</b>	<b>33</b>
<b>Apéndice</b>	<b>35</b>
<b>A. Modelos para variables de conteo en R</b>	<b>37</b>
<b>B. Modelación en R</b>	<b>43</b>
<b>Bibliografía</b>	<b>52</b>





# Introducción

Dentro de la dinámica de las poblaciones intervienen tres componentes básicos: la mortalidad, la migración y la fecundidad. Esta última ha tenido cambios notables en México con el paso del tiempo. Durante la década de los años sesenta México presentó los niveles más altos de fecundidad en todo el siglo XX, reportando más de siete hijos en promedio por mujer, sin embargo, debido a la implementación de distintas políticas públicas diseñadas para la planificación familiar (garantizando el derecho de la población a decidir de una forma libre, responsable e informada sobre el número de hijos e hijas que desea tener), así como el incremento en la difusión de información y el acceso a los distintos métodos anticonceptivos para planificar un embarazo y reducir las infecciones de transmisión sexual, ha logrado que la fecundidad presente una trayectoria descendente.

La transición demográfica es un proceso de cambio demográfico en donde influyen aspectos sociales, económicos, culturales, etc., y que ha explicado los cambios en los niveles de fecundidad, mortalidad y migración. No obstante, algunos autores han demostrado que dichos cambios son más significativos en ciertos grupos de la población, y que, generalmente están asociados a los niveles de vulnerabilidad y, en el caso de la fecundidad, la exposición a un embarazo.

Un caso particular es la situación que se presenta en el país con relación al embarazo en adolescentes. De acuerdo con los datos de la Encuesta Nacional de la Dinámica Demográfica (ENADID), la tasa de fecundidad en mujeres de 15 a 19 años de edad ha incrementado en estos últimos años, de pasar de 69.2 hijos nacidos vivos por cada mil adolescentes en el año 2007 a 77.0 para el año 2012, motivo por el cual, en enero del 2015, el Gobierno de la República puso en marcha la Estrategia Nacional para la Prevención del Embarazo en Adolescentes (ENAPEA), con el objetivo reducir en un 50 por ciento la tasa de fecundidad adolescente para el año 2030 y erradicar el embarazo en niñas menores de 15 años.

Para cumplir los objetivos de la ENAPEA, un grupo de especialistas determinaron las líneas de acción que se necesitan llevar a cabo para reducir los embarazos en las adolescentes: incrementar las oportunidades educativas en las y los jóvenes, fomentar una educación integral en sexualidad, crear oportunidades laborales apropiadas para las y los adolescentes, incrementar el número de leyes y/o acuerdos que permitan a las y los adolescentes a decidir de una forma libre e informada sobre su sexualidad, crear servicios de salud para la atención de jóvenes que buscan información sobre salud sexual y reproductiva así como prevenir y atender la violencia sexual contra las niñas, niños y adolescentes<sup>1</sup>.

---

<sup>1</sup>Informe de los indicadores para el monitoreo y evaluación de la ENAPEA. Fecha de consulta: 16/08/2018. Disponible en: [https://www.gob.mx/cms/uploads/attachment/file/309114/Informe\\_Final\\_Indicadores\\_12032018.pdf](https://www.gob.mx/cms/uploads/attachment/file/309114/Informe_Final_Indicadores_12032018.pdf)

Es por ello que el presente trabajo de investigación tiene por objetivo determinar las posibles causas sociales, demográficas, económicas, culturales, etc. que influyen en el número de hijos e hijas que tendrá una mujer durante su etapa reproductiva, con la finalidad de identificar los grupos con mayor exposición a los cambios en los niveles de fecundidad. Cabe mencionar que este trabajo no solo se limita al estudio de las adolescentes, sino a todas las mujeres en edad fértil que podrían tener un embarazo (15 a 49 años de edad) ya que los factores sociodemográficos que pueden influir en los niveles de fecundidad estarán presentes durante toda su vida.

De esta manera, este trabajo está conformado por los siguientes capítulos, con el fin de entender la situación de la fecundidad en el país y su relación con algunos factores de riesgo:

1. **La fecundidad y la transición demográfica.** En este primer capítulo se presenta toda la información relativa a términos y conceptos relacionados con la fecundidad, asimismo se analiza la situación actual de algunos países en comparación con nuestro país, con el fin de entender la posición en la que nos encontramos hoy en día.
2. En el segundo capítulo se muestra el **marco teórico** de investigación. En él se revisan algunas investigaciones y estudios que han asociado los niveles de fecundidad con algunos factores de riesgo. Esta revisión bibliográfica es de suma importancia debido a que, con base en ella, se sustentan las hipótesis de investigación.
3. En el tercer capítulo se muestra la **f fuente de información y las distintas metodologías** para realizar esta investigación. La fuente de información es la Encuesta Nacional de la Dinámica Demográfica (ENADID) para su levantamiento del año 2014 (última encuesta levantada a la fecha) y los métodos estadísticos que se analizan son el Modelo de Regresión Poisson, el Modelo de Regresión Binomial Negativa y el Modelo Poisson inflado por ceros (*Zero Inflated Poisson*).
4. En el capítulo “**Validación y obtención de resultados**”, como primer punto se realiza un análisis descriptivo de la información de la ENADID con el fin de garantizar la validez de las hipótesis de investigación, asimismo se muestran las pruebas realizadas para la selección de un modelo óptimo que cumpla con los supuestos estadísticos y que tenga mayor capacidad predictiva. Finalmente, tras haber seleccionado el modelo óptimo, se lleva a cabo la interpretación de resultados y la validación de las hipótesis de investigación planteadas en el marco teórico.
5. El último capítulo está asignado para las **conclusiones de la investigación**. También se aborda la importancia de la validación de los supuestos de los modelos estadísticos analizados para una correcta interpretación de resultados.

# Capítulo 1

## La fecundidad y la transición demográfica

En demografía, la fecundidad se define como el número promedio de hijos tenidos en una generación de mujeres a lo largo de su vida reproductiva, si estas mujeres no están expuestas al riesgo de muerte durante esta etapa. Una unidad de medida utilizada para determinar los niveles de fecundidad es la tasa global de fecundidad. De acuerdo con la Comisión Económica para América Latina y el Caribe (CEPAL), la tasa global de fecundidad se define como el número de hijos que en promedio tendrá una mujer en una cohorte hipotética de mujeres que durante su vida fértil tuvieron a sus hijos de acuerdo a las tasas de fecundidad por edad y que no estuvieron expuestas a riesgos de mortalidad desde el nacimiento hasta el término del periodo fértil<sup>1</sup>. Esta medida de fecundidad es de suma importancia para muchos países porque forma parte de los indicadores resumen que describen la calidad de vida de su población, su estado de envejecimiento y su acceso a la información en términos de planificación familiar y uso de métodos anticonceptivos.

En los últimos años se ha presentado una caída en los niveles de fecundidad como parte de la transición demográfica que atraviesan la mayoría de los países desarrollados. La transición demográfica se refiere a los procesos demográficos identificables y que constituyen una propuesta de explicación de la dinámica demográfica y sus relaciones con los factores sociales, económicos y culturales (Zavala de Cosío, 1992). Sin embargo, en la mayoría de los países en vías de desarrollo la transición demográfica apenas comienza a presentarse, mientras que por otro lado en algunos países, las tasas de fecundidad por edad tienden a presentar ligeros aumentos con el paso de los años, especialmente en los grupos de edad más jóvenes (Chackiel, 2004).

Otro concepto relacionado con la tasa global de fecundidad es el nivel de reemplazo generacional. El Consejo Nacional de Población (CONAPO) lo define como el nivel que se necesita mantener, por un largo periodo, para asegurar que una población se reemplace a sí misma. De este modo, en una población donde todas sus mujeres sobreviven en su edad reproductiva y en donde la probabilidad de tener a una hija en cada embarazo es del 50 por ciento, el nivel de reemplazo será de 2.0 hijos por mujer, aunque se ha demostrado que en algunos países donde los niveles de mortalidad son bajos o moderados y la razón de masculinidad es de 105 hombres por cada 100 mujeres, el nivel de reemplazo generacional es de 2.1 hijos por mujer<sup>2</sup>.

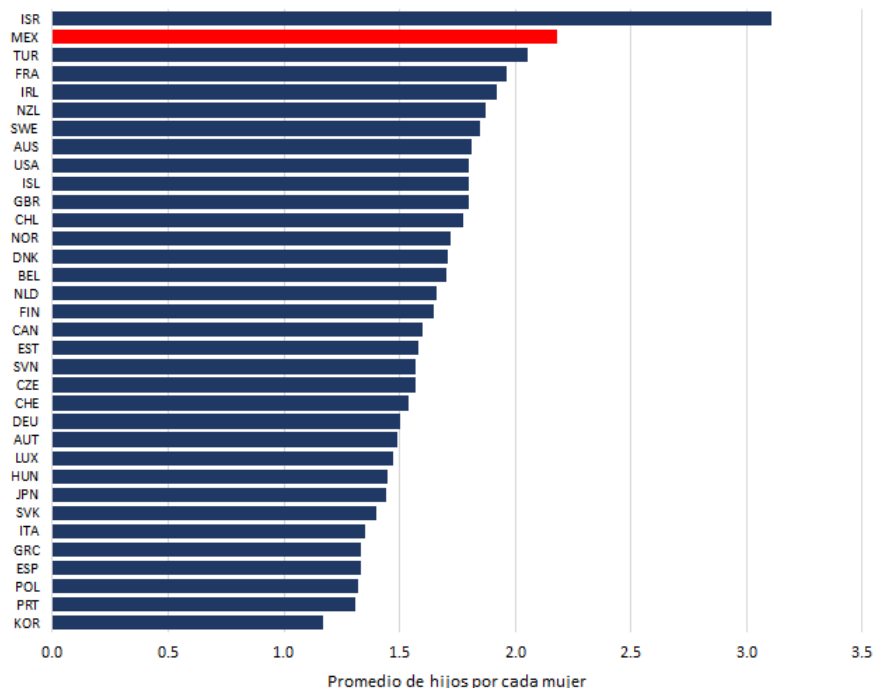
---

<sup>1</sup>Tasa Global de Fecundidad. CELADE, CEPAL. Disponible en: <https://celade.cepal.org/redatam/pryesp/sisppi/webhelp/fecundidad.htm>

<sup>2</sup>CONAPO, Glosario. Disponible en: [http://www.conapo.gob.mx/en/CONAPO/Glosario\\_CONAPO](http://www.conapo.gob.mx/en/CONAPO/Glosario_CONAPO)

De acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE), las tasas de fecundidad entre sus países miembros han presentado una reducción notable durante estas últimas décadas, pasando de 2.7 hijos en promedio por mujer para el año 1970 a 1.6 durante el nuevo siglo<sup>3</sup>, quedando por debajo del nivel de reemplazo generacional. En la Figura 1.1 se muestra el promedio de hijos por mujer en cada uno de los países miembros de la OCDE durante el año 2016.

Figura 1.1: Promedio de hijos por mujer en los países miembros de la OCDE, 2016.



**Nota:** KOR = Corea; PRT = Portugal; POL = Polonia; ESP = España; GRC = Grecia; ITA = Italia; SVK = República Eslovaca; JPN = Japón; HUN = Hungría; LUX = Luxemburgo; AUT = Austria; DEU = Alemania; CHE = Suiza; CZE = República Checa; SVN = Eslovenia; EST = Estonia; CAN = Canadá; FIN = Finlandia; NLD = Países Bajos; BEL = Bélgica; DNK = Dinamarca; NOR = Noruega; CHL = Chile; GBR = Reino Unido; ISL = Islandia; USA = Estados Unidos; AUS = Australia; SWE = Suecia; NZL = Nueva Zelanda; IRL = Irlanda; FRA = Francia; TUR = Turquía; MEX = México; ISR = Israel

**Fuente:** Estimaciones del Banco Mundial, 2016.

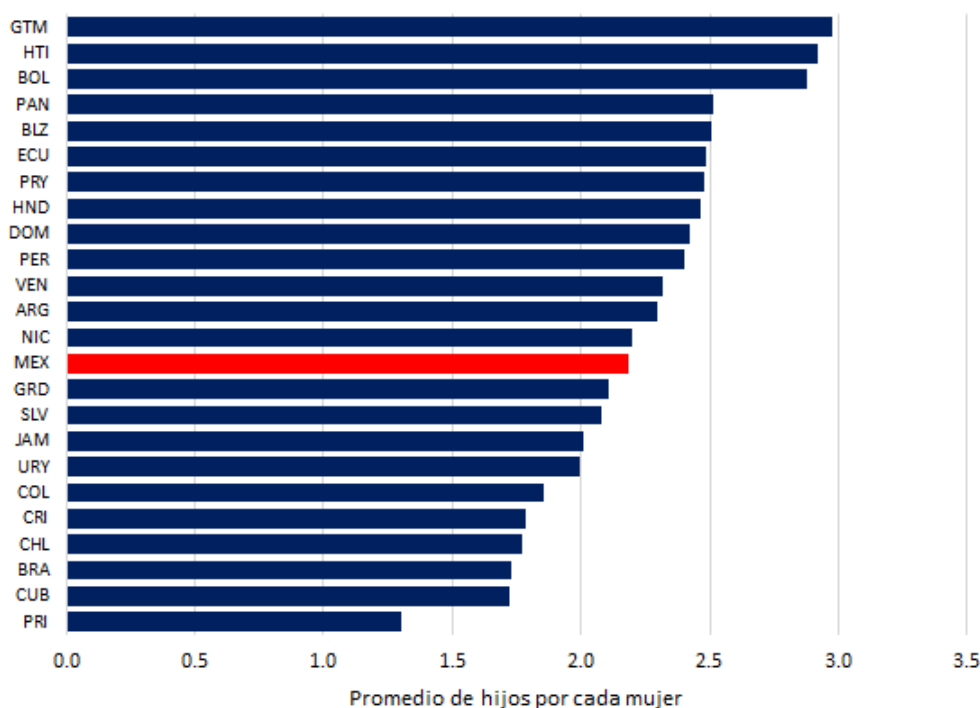
En la figura anterior se muestra que Israel ocupa el primer lugar entre todos los países miembros de la OCDE, con 3.11 hijos en promedio por mujer; en segundo lugar se ubica México con un promedio de 2.18 hijos y en tercer lugar se ubica Turquía con 2.05 hijos en promedio para cada mujer en edad fértil. Un punto importante por analizar es que estos tres países se encuentran por encima del nivel de reemplazo generacional, lo que garantiza que, con sus niveles de fecundidad, la población se pueda reemplazar a sí misma para los siguientes años, si se preservan dichos niveles.

<sup>3</sup>Tasa de natalidad en México, OCDE. Disponible en: <http://www.oecd.org/centrodemexico/medios/21tasadenatalidadenmexico.htm>

Por otro lado, los países con el promedio de hijos por mujer más bajos entre los países miembros de la OCDE se encuentran Corea, Portugal y Polonia, con un promedio de 1.17, 1.31 y 1.32, respectivamente. Este resultado indica que en estos países se presenta un retraso importante en la conformación de familias y con una alta frecuencia de falta de hijos en la familia (OCDE, 2012), el cual está directamente asociado con las condiciones de vida de la población (educación, empleo, salud, etc.).

En América Latina se han presentado grandes cambios demográficos durante el último siglo, que fueron consecuencia de algunos cambios sociales, económicos y culturales (Chackiel, 2004). Chackiel menciona que América Latina registró un descenso en los niveles de fecundidad durante la década de los años sesenta, lo que implicó una reducción significativa en la tasa de crecimiento medio anual para la población, por lo que comenzó un proceso de envejecimiento en la mayoría de los países que forman parte de la región. En la Figura 1.2 se muestra el número promedio de hijos por mujer durante el año 2016 para los países de América Latina.

Figura 1.2: Promedio de hijos por mujer en los países de América Latina, 2016.



**Nota:** PRI = Puerto Rico; CUB = Cuba; BRA = Brasil; CHL = Chile; CRI = Costa Rica; COL = Colombia; URY = Uruguay; JAM = Jamaica; SLV = El Salvador; GRD = Granada; MEX = México; NIC = Nicaragua; ARG = Argentina; VEN = Venezuela; PER = Perú; DOM = República Dominicana; HND = Honduras; PRY = Paraguay; ECU = Ecuador; BLZ = Belice; PAN = Panamá; BOL = Bolivia; HTI = Haití; GTM = Guatemala

**Fuente:** Estimaciones del Banco Mundial, 2016.

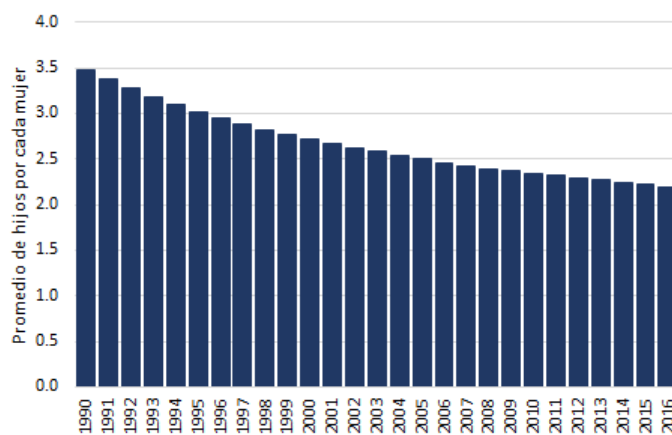
Se puede observar que durante el año 2016 los tres países con el promedio de hijos más altos en América Latina son Guatemala (2.97), Haití (2.91) y Bolivia (2.87); mientras que los países con el promedio más bajo en el mismo año son Puerto Rico (1.3), Cuba (1.72) y Brasil (1.73). Si bien algunas

investigaciones señalan que los niveles de fecundidad de América Latina han reducido notablemente en estos últimos años a tal punto de presentar niveles muy cercanos a América del Norte y Oceanía, sus niveles apuntan a que América Latina presenta importantes desigualdades dentro de sus países, registra grandes brechas en el estatus socioeconómico, en el nivel escolar, en el acceso a los servicios así como una diferencia sustancial de acuerdo a la zona o localidad de residencia de la población (Rodríguez *et al.*, 2015).

Los bajos niveles en la tasa global de fecundidad en algunos países de América Latina se pueden explicar por los cambios sociales y económicos en torno a la mujer. Cuba es un caso particular, en donde el rol de la mujer marcó una gran diferencia en los niveles de fecundidad del país. Durante la mitad del siglo pasado Cuba reportó un promedio de 4.1 hijos por mujer (Chackiel, 2004), pero con la implementación de políticas públicas con relación al control de nacimientos y como uno de los países en América Latina que cuenta con una ley de despenalización del aborto (Alfonso, 2004), ha tenido como consecuencia una reducción notable en sus niveles de fecundidad, a tal punto de reportar 1.7 hijos por mujer para el año 2016.

Entre tanto, México comenzó a tener un avance importante en el proceso de la transición demográfica durante el siglo XX (Morales *et al.*, 2016). Sin embargo, no fue sino hasta la mitad de los años sesenta que comenzó a presentar un descenso importante en los niveles de fecundidad debido a la difusión gradual de políticas públicas y programas de planificación familiar, así como en los cambios sustanciales en la calidad de vida de la población (Morales *et al.*, 2016). En la Figura 1.3 se muestra la evolución del promedio de hijos por mujer en México desde el año 1990 hasta 2016.

Figura 1.3: Promedio de hijos por mujer en México, 1990-2016.



**Fuente:** Estimaciones del Banco Mundial, 1990-2016.

De acuerdo a la figura anterior, se muestra una reducción importante en el promedio de hijos por mujer a causa de los cambios derivados de la transición demográfica en el país, al pasar de 3.4 hijos en promedio por mujer en el año 1990 a 2.1 hijos en el año 2016, muy cercano al nivel de reemplazo generacional. Este resultado sugiere que México pasará a ser, en los siguientes años, una población envejecida como resultado de un crecimiento lento de la población. Por tal motivo es necesario analizar los factores sociales, demográficos, económicos, culturales, etc. que han impactado directamente en los cambios la fecundidad en el país.

# Capítulo 2

## Marco teórico

A lo largo de los años, distintos investigadores y especialistas en el tema han buscado la relación que existe entre la tasa de fecundidad y algunos factores sociodemográficos, tales como los niveles de pobreza, el nivel de escolaridad de la madre y del padre, la condición de habla de lengua indígena, el lugar de residencia habitual, el acceso a los servicios de salud donde la población pueda recibir la información y los medios necesarios para una planificación familiar, la construcción social de género, las reducidas oportunidades económicas y laborales, entre otros. Es por ello que es necesario ubicar al embarazo como un proceso de cambio social y cultural que repercute en distintos grados de acuerdo al país y al contexto social en el que se ubica (Safe Passages of Adulthood, 2001).

Algunos investigadores mencionan que la edad es un factor importante que puede influir en el número de hijos que tiene una mujer. Welti (2012) menciona que el patrón de la fecundidad por edad sirve para derivar factores de ajuste aplicables a la fecundidad, que originalmente emanan solo del promedio de hijos nacidos vivos. En su trabajo muestra que el promedio de hijos nacidos vivos aumenta conforme incrementa la edad de la madre. Así, el promedio de hijos nacidos vivos en las mujeres de 20 a 24 años es de 0.80, entre las mujeres de 35 a 39 años de edad este promedio incrementa a 3.40 y finalmente para las mujeres de 45 a 49 años el promedio de hijos es de 4.75 por mujer (Welti, 2000). También menciona que el promedio de hijos nacidos vivos es más alto si la mujer fue madre durante la adolescencia.

El Instituto Nacional de Estadística y Geografía (INEGI) señala que la probabilidad de tener un mayor número de hijos está relacionada con la edad de la mujer (INEGI, 2017) ya que, de acuerdo con la Encuesta Intercensal 2015, entre las madres de 12 y 19 años de edad el 82.2% tienen un solo hijo; en las madres de 20 a 29 años el porcentaje con un solo hijo disminuye a 46.5% y aumenta a 34.9% la proporción de mujeres con dos hijos; y finalmente, el 51.9% de las madres de 30 a 49 años de edad tienen tres o más hijos.

El nivel de escolaridad de la población también influye en el número de hijos e hijas dentro de la familia. Asistir a una institución educativa es fundamental para disminuir la tasa de fecundidad (Stupp y Cáceres, 2001) ya que en este lugar se puede recibir una educación integral en sexualidad, así como pláticas y orientaciones a los jóvenes sobre las distintas formas para prevenir un embarazo o una infección de transmisión sexual a través del uso correcto de los métodos anticonceptivos.



Este último resultado indica que la escolaridad tiene un alto impacto sobre el conocimiento y uso correcto de los métodos anticonceptivos (y como consecuencia en la reducción en la tasa de fecundidad) ya que las mujeres que cuentan con un mayor nivel de educación tendrán nociones más concretas en cuanto a algunas definiciones y conceptos relacionados con el tema. Además se sabe que dentro de las instituciones educativas se brindan pláticas y talleres sobre salud sexual y reproductiva, lo que implica que las y los jóvenes estudiantes tengan un mejor conocimiento del tema.

Sin embargo, algunos autores comentan que cuando una mujer presenta un embarazo en edades tempranas, la probabilidad de que pueda asumir la maternidad y los estudios simultáneamente es muy baja, y como consecuencia ésta tiende a posponer o abandonar sus estudios como resultado del embarazo (Llanes, 2010), lo que llega a suponer que las mujeres con menor grado de estudio presenten en promedio un mayor número de hijos en comparación con aquellas mujeres que cuentan con un nivel de estudios superior.

Otro factor importante en el incremento de la tasa de fecundidad es la situación conyugal de las mujeres (Belmonte, 2007), esto debido a que las mujeres casadas o que viven en unión libre son las que presentan mayores probabilidades de concebir. Belmonte menciona en su trabajo que en el estado de Sonora las mujeres viudas, casadas y en unión libre son el grupo que promedian el mayor número de hijos, seguido de las mujeres divorciadas y finalmente las solteras.

También, Curtis y Blanc (1997) demostraron que las mujeres casadas o que viven en unión libre tienen una mayor probabilidad de hacer uso incorrecto de los métodos anticonceptivos en comparación con las mujeres solteras, esto debido a que existe una menor regulación en el número de la descendencia por parte de las mujeres casadas o unidas, mientras que entre las mujeres solteras el uso efectivo de los métodos anticonceptivos está fuertemente relacionado con la prevención de un embarazo no planeado.

La condición de actividad económica de las mujeres es considerada como otro factor que influye en la tasa de fecundidad. Algunas investigaciones apuntan que la participación femenina en el campo laboral se relaciona con mayor énfasis a las actividades domésticas, la cual se ha considerado más como una construcción cultural (OIT, 2009). Así, bajo dicha construcción persisten dos posturas: la primera encomienda a las mujeres el cuidado de la familia, así como el cuidado de los hijos e hijas como su principal tarea o función dentro del hogar y la segunda considera su participación laboral más como complemento a los recursos generados por los hombres (OIT, 2009).

Sin embargo, algunos estudios indican que los cambios en las condiciones económicas de la población han generado transformaciones sociodemográficas de tal forma que dentro de los hogares es frecuente encontrar a dos proveedores de ingresos (padre y madre) así como en aquellos donde la mujer es la única sostenedora de la familia (OIT, 2009). Así, en los últimos años ha aumentado el número de mujeres con un empleo remunerado durante y después de la crianza de sus hijos e hijas (ONU, 2001), por lo que la tasa de empleo en las mujeres parece estar influida por el número de hijos pequeños.

Los niveles de fecundidad entre las mujeres indígenas son más elevados en comparación con el resto de la población (Valdivieso, 2012), ya que forman parte de un grupo rezagado dentro de los programas sociales enfocados a la planificación familiar y acceso a los métodos anticonceptivos. Szasz y Lerner (2010) mencionan en su trabajo que una de las causas por las que las mujeres indígenas no hayan escuchado hablar sobre los métodos anticonceptivos es debido al limitado acceso que tienen a estos, lo que permite pensar que dichas mujeres no tienen los elementos suficientes para regular su fecundidad ni de estar informadas sobre las características y consecuencias sobre el uso de métodos (Meneses *et al.*, 2017).

De acuerdo con los resultados de la Encuesta de Salud Reproductiva de 1999 (Szasz y Lerner, 2010), sólo un tercio de las mujeres hablantes de lengua indígena residentes en zonas rurales usan algún método anticonceptivo, de las cuales el 14 por ciento usan métodos tradicionales, 0.5 por ciento usan el condón masculino, 45 por ciento emplean la esterilización femenina, el 31 por ciento usan el dispositivo intrauterino y el 9.5 por ciento usan otro anticonceptivo.

Szasz y Lerner (2010) mencionan que aproximadamente la mitad de las usuarias de estos dos últimos métodos hacen uso de ellos por prescripción médica o preferencia de su pareja. Este resultado sugiere que los servicios de salud reproductiva dentro de las comunidades rurales y en áreas en donde la presencia de las mujeres que hablan alguna lengua indígena es muy alta, no alientan la información ni la libre decisión sobre el tamaño idóneo de la descendencia.

Otro factor que puede influir en la incidencia del embarazo es la localidad de residencia habitual de las mujeres. En las comunidades rurales se piensa que el embarazo está ligado a cuestiones culturales e ideológicas en las que se desenvuelven las mujeres, y que además se ha transmitido de generación en generación que su función es ser madres, por lo que el embarazo y el matrimonio forman parte de sus proyectos de vida (Tuñón, 2000).

Asimismo, de acuerdo con los resultados de la Encuesta Nacional de la Dinámica Demográfica 2014, en las zonas rurales se presentan los porcentajes más bajos de uso de métodos anticonceptivos, y en el caso de las mujeres que hacen uso de los métodos anticonceptivos presentan un alto riesgo de no usarlos correctamente, esto como resultado de una orientación deficiente (Meneses *et al.*, 2017), lo que en muchas ocasiones resulta en un embarazo no planeado.

Con relación a la pobreza, se pueden encontrar dos enfoques, uno que postula una relación negativa con los niveles de fecundidad y otro con un enfoque positivo (Llovet, 1989). El primer enfoque menciona que los niveles de fecundidad más bajos se encuentran dentro de las familias más pobres, partiendo del supuesto que los hijos son considerados como un bien de consumo, por lo que las oportunidades económicas reducen cuando se tiene a más de un hijo (Becker, 1960). Por otro lado, el segundo enfoque apoya la evidencia de que a mayores niveles de pobreza se presentarán mayores niveles de fecundidad.

Algunos autores han evidenciado que en las localidades en donde las mujeres pertenecen a estratos socioeconómicos altos se presentan niveles de fecundidad más bajos, menores niveles de fecundidad adolescente, un calendario sexual y reproductivo más envejecido, una incidencia mayor en el uso de métodos anticonceptivos y comportamientos sexuales más cuidadosos para protegerse de las infecciones de transmisión sexual respecto a las mujeres de estrato socioeconómico más bajo (Di Cesare, 2007).

Además, se ha demostrado que los grupos socioeconómicamente más pobres no solo presentan la fecundidad más alta, sino que también son clasificados como un grupo vulnerable en materia de salud sexual y reproductiva (Di Cesare y Rodríguez, 2006), esto como consecuencia de un bajo acceso y disponibilidad de los métodos anticonceptivos, lo que pasa a ser un problema de necesidades insatisfechas<sup>1</sup> (Schoumaker y Tambutin, 1990).

Así, de acuerdo a la revisión literaria se plantearon las siguientes hipótesis de investigación: 1.- los niveles de fecundidad incrementan conforme incrementa la edad de las mujeres; 2.- el número de hijos disminuye conforme incrementa el nivel de escolaridad de las mujeres; 3.- el número de hijos incrementa entre las mujeres casadas o que viven en unión libre; 4.- el número de hijos en mujeres que trabajan (o buscan trabajo) es más alto en comparación con las mujeres que se dedican a los quehaceres y cuidados del hogar; 5.- el número de hijos en mujeres que hablan alguna lengua indígena es más alto en comparación con las mujeres que no hablan lengua indígena; 6.- el número de hijos incrementa en las mujeres que viven en las localidades rurales; y 7.- los grupos más pobres presentan un número mayor de hijos.

---

<sup>1</sup>La necesidad insatisfecha se define como la cantidad de mujeres que desean limitar o espaciar sus embarazos, pero no están usando ningún método anticonceptivo, en relación con el número total de mujeres.

# Capítulo 3

## Fuente de información y metodología

Como se ha mencionado al inicio de este trabajo, el objetivo principal fue determinar las posibles causas sociales, demográficas, económicas, culturales, etc. que influyen en el incremento en la tasa de fecundidad en el país, para ello se buscó un modelo estadístico que permitiera relacionar la tasa de fecundidad con diversos factores de riesgo. Para cumplir con dicho objetivo fue necesario determinar la fuente de información de la cual se obtuvieran las características anteriormente mencionadas.

### 3.1. La Encuesta Nacional de la Dinámica Demográfica, 2014

La fuente de información para esta investigación fue la Encuesta Nacional de la Dinámica Demográfica (ENADID) en su levantamiento del año 2014. Esta encuesta tiene como principal objetivo brindar de información estadística relacionada con el nivel y el comportamiento de los componentes de la dinámica demográfica: Fecundidad, Mortalidad y Migración (interna e internacional), así como de información con relación al tema de salud sexual y reproductiva en las mujeres de 15 a 54 años de edad. El marco de muestreo que se utilizó para la ENADID 2014 es el Marco Nacional de Viviendas 2012 del Instituto Nacional de Estadística y Geografía (INEGI) con base en la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2010.

El tamaño de muestra para cada entidad federativa fue de aproximadamente 3,000 viviendas, con una sobremuestra de 500 viviendas en 8 entidades federativas y de 1,000 viviendas aproximadamente en el caso de la Ciudad de México. Así, el tamaño de muestra fue de 101 mil 389 viviendas a nivel nacional, donde la unidad de observación fueron las mujeres de 15 a 54 años de edad residentes en las viviendas seleccionadas, por lo cual la encuesta tiene representatividad a nivel nacional. El esquema de muestreo de la ENADID 2014 es probabilístico, estratificado, bietápico y por conglomerados, es decir, es una encuesta bajo un esquema de diseño complejo.

Para el presente trabajo de investigación se utilizó la información correspondiente a las características de cada uno de los integrantes de la vivienda y al módulo para la mujer elegible dentro de las temáticas de fecundidad, preferencias reproductivas y nupcialidad. Para determinar el número de hijos nacidos vivos se utilizó la siguiente pregunta dentro del cuestionario para la mujer de 15 a 54 años:

- En total, ¿cuántos hijos e hijas que nacieron vivos ha tenido usted?

Y las preguntas de las cuales se tomarán las variables para modelar la tasa de fecundidad son:

- La edad de la mujer al momento de la encuesta.  
¿Cuántos años cumplidos tiene usted?
- El nivel de escolaridad.  
¿Cuál es el último año o grado que aprobó usted en la escuela?
- La situación conyugal.  
¿Actualmente usted vive con su pareja en unión libre, es separada de una unión libre, es separada de un matrimonio, es divorciada, es viuda de una unión libre, es viuda de un matrimonio, es casada o es soltera?
- La condición de actividad económica.  
¿Durante la semana pasada usted trabajó (al menos una hora)?  
Aunque ya me dijo usted que no trabajó, ¿la semana pasada ayudó en un negocio familiar, vendió o hizo algún producto para vender, ayudó en las labores del campo o en la cría de animales, a cambio de un pago realizó otro tipo de actividad, tenía un trabajo pero estuvo ausente, o no trabajó ni ayudó?  
Entonces, ¿la semana pasada buscó trabajo, se dedicó a los quehaceres del hogar, es estudiante, es pensionado(a) o jubilado(a), tiene una discapacidad permanente que le impide trabajar u otra situación?
- La condición de habla de lengua indígena.  
¿Usted habla algún dialecto o lengua indígena?
- El lugar de residencia<sup>1</sup>.
- El estrato sociodemográfico<sup>2</sup>.

### 3.2. Modelos estadísticos para el análisis de variables de conteo

Una variable de conteo se define como un número de sucesos y/o eventos que ocurren en una misma unidad de medida, en un determinado periodo de tiempo. Los modelos para variables de conteo tienen la característica de que la variable independiente es un conjunto de números enteros no negativos, que representa el total de eventos ocurridos y que puede ser explicado por un conjunto de covariables de tipo numérico (discretos o continuos) o categórico (nominal u ordinal). A continuación se muestran los principales modelos estadísticos para variables de conteo.

---

<sup>1</sup>Según el INEGI, el ámbito urbano son aquellas localidades con más de 100,000 habitantes, el complemento urbano es aquella localidad donde el número de habitantes se encuentra entre 2,500 y 99,999, y las localidades rurales son aquellas donde el número de habitantes es menor a 2,500.

<sup>2</sup>Esta estratificación considera las características sociodemográficas de los habitantes de las viviendas, así como las características físicas y el equipamiento de estas, expresadas a partir de 34 indicadores construidos con la información del Censo de Población y Vivienda 2010. En el siguiente capítulo se detalla con más profundidad la construcción de esta variable.

### 3.2.1. Modelo de Regresión Poisson

Una distribución de probabilidad que permite modelar una variable de tipo conteo es la distribución Poisson, que tiene función de probabilidad:

$$P[Y = y] = \frac{\mu^y e^{-\mu}}{y!} \quad y = 0, 1, 2, \dots$$

Donde  $\mu$  representa la media de las ocurrencias, que generalmente son expresadas como tasas. La distribución Poisson es utilizada para modelar eventos por unidad espacial o por unidad de tiempo. Esta distribución tiene como característica principal que  $E[Y] = Var[Y] = \mu$ , es decir, a medida que la media crezca (o decrezca) la varianza crece (o decrece).

En este trabajo de investigación se toma como supuesto que el número de hijos nacidos vivos sigue una distribución Poisson como resultado de la Ley de los Eventos Raros (el número total de eventos seguirá una distribución Poisson si un evento puede ocurrir en cualquier punto del tiempo o espacio bajo observación, pero la probabilidad de ocurrencia en un punto determinado es relativamente pequeña) (Cameron y Trivedi, 1998).

El modelo de Regresión Poisson es considerado como un modelo de referencia en estudios con variables de conteo (Cameron y Trivedi, 1998). Como se mencionó anteriormente, este modelo tiene como principal supuesto que  $E[Y] = Var[Y]$ , por lo que en algunas ocasiones los resultados que se obtienen suelen ser insatisfactorios si no se cumple dicho supuesto. Éste forma parte de los Modelos Lineales Generalizados, y está dado por la siguiente ecuación:

$$\ln(E[y_i | x_{i,1}, x_{i,2}, \dots, x_{i,n}]) = \ln(\mu_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_n x_{i,n}$$

Donde

$y_i$  representa el número de eventos ocurridos

$\beta_i$  son los coeficientes de regresión del modelo

$x_{i,j}$  son las variables independientes que explican el comportamiento de  $\mu_i$

A pesar de ser uno de los modelos con mayor referencia sobre el análisis en variables de conteo, este presenta muchos problemas cuando la media y la varianza de los datos analizados no son iguales, por lo que se sugiere utilizar otros modelos que consideren una subdispersión ( $E[Y] > Var[Y]$ ) o sobredispersión ( $E[Y] < Var[Y]$ ). Esta última aparece con mayor frecuencia, por lo que se analiza con mayor detalle.

### 3.2.2. Modelo de Regresión Binomial Negativa

El Modelo de Regresión Poisson es considerado como un buen modelo cuando se cumple el supuesto de  $E[Y] = Var[Y]$ , sin embargo en muchas ocasiones este supuesto no se cumple debido a las siguientes posibles causas:

- La alta variabilidad de los datos.
- Los eventos no ocurren de forma independiente.
- Los datos no se ajustan a una distribución Poisson.

- Cuando la media no está bien especificada, es decir, hay variables predictoras que no se han incluido en el modelo, o existen relaciones no lineales.

Una forma sencilla para detectar la sobredispersión es a través del cociente entre la varianza y la media de los datos analizados:

$$Z = \frac{Var[Y]}{E[Y]}$$

Bajo este criterio, se sospecha que hay sobredispersión si la varianza supera en dos veces a la media (Montero *et al.*, 2005). Otra forma de evaluar la sobredispersión de los datos es a través del Modelo de Regresión Binomial Negativa. Se dice que una variable aleatoria de conteo  $Y$  sigue una distribución binomial negativa con parámetros  $\theta \in \mathbb{Z}^+$  y  $0 < p < 1$  si tiene como función de probabilidad:

$$P[Y = y] = \binom{y + \theta - 1}{\theta - 1} p^\theta (1 - p)^y \quad y = 0, 1, 2, \dots$$

En la distribución Binomial Negativa la esperanza y la varianza están dados por:  $E[Y] = \frac{\theta(1-p)}{p}$  y  $Var[Y] = \frac{\theta(1-p)}{p^2}$ , de esta forma se establece que  $Var[Y] = \frac{1}{p}E[Y]$ , y dado a que  $0 < p < 1$  se demuestra que  $E[Y] < Var[Y]$ , lo que justifica que esta distribución permite modelar datos con sobredispersión. El modelo de regresión binomial negativa parte del supuesto de que la variable respuesta sigue una distribución Poisson, pero la media no se considera como un parámetro fijo, sino que está sigue una distribución Gamma (Hausman *et al.*, 1984):

$$\begin{aligned} y_i | x_{i,j}, \mu_i &\sim P(\mu_i) & \mu_i > 0 \\ \mu_i | x_{i,j} &\sim Ga(\tau_i, w) & \tau_i \in \mathbb{Z}^+, w > 0 \end{aligned}$$

entonces:

$$\begin{aligned} P[Y_i = y_i | x_{i,j}] &= \int_0^\infty P[Y_i = y_i | x_{i,j}, \mu_i] f(\mu_i | x_{i,j}) d\mu_i \\ &= \int_0^\infty \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \frac{1}{w^{\tau_i} \Gamma(\tau_i)} \mu_i^{\tau_i - 1} e^{-\frac{\mu_i}{w}} d\mu_i \\ &= \frac{1}{w^{\tau_i} \Gamma(\tau_i) y_i!} \int_0^\infty e^{-(1 + \frac{1}{w})\mu_i} \mu_i^{y_i + \tau_i - 1} d\mu_i \\ &= \frac{1}{w^{\tau_i} \Gamma(\tau_i) y_i!} \frac{\Gamma(y_i + \tau_i)}{(1 + \frac{1}{w})^{y_i + \tau_i}} \\ &= \frac{\Gamma(y_i + \tau_i)}{y_i! \Gamma(\tau_i)} \left( \frac{1}{1 + w} \right)^{\tau_i} \left( \frac{w}{1 + w} \right)^{y_i} \\ &= \frac{(y_i + \tau_i - 1)!}{\tau_i - 1! (y_i!)} \left( \frac{1}{1 + w} \right)^{\tau_i} \left( \frac{w}{1 + w} \right)^{y_i} \end{aligned}$$

Por lo tanto  $Y_i \sim BN(\tau_i, \frac{1}{1+w})$ . Cameron y Trivedi (1986) proponen que  $\mu_i \sim Ga(\theta_i, \frac{\mu_i}{\theta_i})$  con  $\theta_i, \mu_i \in \mathbb{Z}^+$ , de esta forma  $Y_i \sim BN(\theta_i, \frac{\theta_i}{\theta_i + \mu_i})$  con  $E[Y_i|x_{i,j}] = \mu_i$  y  $Var[Y_i|x_{i,j}] = \mu_i + \frac{1}{\theta_i}\mu_i^2 = \mu_i + \alpha_i\mu_i^2$

Así, el Modelo de Regresión Binomial Negativa estará dada por:

$$\ln(E[y_i|x_{i,1}, x_{i,2}, \dots, x_{i,n}]) = \ln(\mu_i) = \beta_0 + \beta_1x_{i,1} + \beta_2x_{i,2} + \dots + \beta_nx_{i,n}$$

La forma de evaluar si los datos presentan sobredispersión es a través de la log-verosimilitud de los modelos Poisson y Binomial Negativa. Como se vio anteriormente,  $Var[Y] = \mu + \alpha\mu^2$ , entonces si  $\alpha \rightarrow 0$ ,  $Var[Y] = \mu$ , por lo que se plantea la siguiente prueba de hipótesis:

$$H_0 : \alpha = 0 \quad vs \quad H_1 : \alpha > 0$$

Calculando la función de log-verosimilitud para ambos modelos, el estadístico de prueba es  $RV = -2[l_{Poisson} - l_{BN}]$  el cual se distribuye asintóticamente como Chi-Cuadrada con un grado de libertad (Cameron y Trivedi, 1998). Por lo tanto se rechaza la hipótesis nula, con un nivel de significancia de  $\gamma$ , si el estadístico de prueba es mayor que  $\chi_{1,1-\gamma}^2$ .

### 3.2.3. Modelo Poisson Inflado por Ceros (*Zero Inflated Poisson*)

Hay ocasiones en donde hay demasiados ceros, que propician un mal ajuste. Para solucionar este problema se parte del supuesto de que los ceros provienen de dos procesos distintos: el proceso Bernoulli y el proceso Poisson. De esta forma, el modelo se divide en un Modelo de Regresión Logístico para modelar la probabilidad de que un cero observado provenga del modelo Poisson y en un Modelo de Regresión Poisson para modelar la probabilidad de obtener el resto de los valores de la variable de conteo, incluyendo ceros. A este modelo se le conoce como *Zero Inflated Poisson* (Poisson inflado por ceros).

Como se mencionó, este modelo tiene dos componentes:

- a) Un modelo para el evento binario que define si la variable de conteo es cero o no. Es un Modelo de Regresión Logístico para la probabilidad  $\pi_i$  de que el cero observado provenga del modelo Poisson, con variables predictoras  $z_1, z_2, \dots, z_n$

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \gamma_0 + \gamma_1z_{i,1} + \gamma_2z_{i,2} + \dots + \gamma_nz_{i,n}$$

- b) Un modelo Poisson para la otra clase, que puede tomar el valor cero o los otros valores, con variables predictoras  $x_1, x_2, \dots, x_m$

$$\ln(\mu_i) = \beta_0 + \beta_1x_{i,1} + \beta_2x_{i,2} + \dots + \beta_mx_{i,m}$$

Por lo tanto, la distribución de probabilidad de una variable aleatoria de conteo  $y_i$  en un modelo *Zero Inflated Poisson* se puede escribir como:

$$P[y_i = j] = \begin{cases} \pi_i + (1 - \pi_i)e^{-\mu_i} & \text{si } j = 0 \\ (1 - \pi_i) \left( \frac{\mu_i^j e^{-\mu_i}}{j!} \right) & \text{si } j = 1, 2, \dots \end{cases}$$



Una forma de determinar si los datos tienen problemas de excesos de ceros es a través de la prueba de Vuong. La prueba de Vuong está diseñada para comparar si dos modelos ( $M_1$  y  $M_2$ ) se ajustan a los mismos datos por máxima verosimilitud. Específicamente, la hipótesis nula indica que los dos modelos se ajustan a los datos igualmente bien.

La prueba calcula, para cada observación, el logaritmo de la relación entre el valor previsto en el modelo inflado por ceros y el valor previsto por el modelo tradicional. A esta nueva variable se le calcula el siguiente estadístico de prueba:

$$V = \frac{\sqrt{n}(\bar{m})}{S_m}$$

Donde  $n$  es el número de observaciones,  $\bar{m}$  y  $S_m$  es la media y la desviación estándar típica. La estadística de prueba se distribuye como una normal estándar  $N(0, 1)$ . Valores positivos altos (superiores al cuantil  $Z_{1-\frac{\alpha}{2}}$ ) son un argumento en favor al modelo inflado por ceros; valores negativos altos (inferiores al cuantil  $Z_{\frac{\alpha}{2}}$ ) son un argumento en favor de los modelos tradicionales; mientras que un valor entre  $Z_{\frac{\alpha}{2}}$  y  $Z_{1-\frac{\alpha}{2}}$  apoya la hipótesis nula (ambos modelos se ajustan de forma similar) (Montero *et al.*, 2005).

# Capítulo 4

## Validación y obtención de resultados

Hay ocasiones en que el modelo no se llega a ajustar bien a los datos, generalmente debido a dos casos: a) el modelo no está bien especificado, es decir, que hay variables predictoras que no se han incluido al modelo o bien existen relaciones no lineales, o b) existe una sobredispersión en los datos ( $E[Y] < Var[Y]$ ). Por lo tanto, es necesario hacer un análisis previo de la información antes de emitir algún resultado que derive del modelo de regresión propuesto.

### 4.1. Análisis descriptivo

Como primer paso se analiza el comportamiento de la variable respuesta (número de hijos nacidos vivos). De acuerdo con la información de la ENADID 2014, el tamaño de muestra fue de 89,904 mujeres de 15 a 49 años de edad (se excluyen para este análisis 12 mujeres de las cuales no se pudo obtener información puntual sobre su nivel de escolaridad). En el Cuadro 4.1 se muestra la distribución de las mujeres de 15 a 49 años según el número total de hijos nacidos vivos que han tenido hasta el momento del levantamiento de la encuesta.

Cuadro 4.1: Total de mujeres de 15 a 49 años según número de hijos nacidos vivos, 2014.

Número HNV	Tamaño de muestra	Número de mujeres (estimado)	Desviación estándar	Coefficiente de variación	Límite inferior (90%)	Límite superior (90%)
0	29,276	10,662,338	91,938.04	0.86	10,511,110	10,813,560
1	15,233	5,463,209	56,840.24	1.04	5,369,715	5,556,703
2	19,842	7,179,636	68,294.61	0.95	7,067,301	7,291,971
3	15,253	5,192,915	54,661.84	1.05	5,103,004	5,282,826
4	6,023	2,037,066	34,579.80	1.70	1,980,187	2,093,945
5	2,296	769,200	20,150.51	2.62	736,055	802,345
6	1,016	341,632	13,056.97	3.82	320,155	363,109
7	486	163,783	9,430.01	5.76	148,272	179,294
8	250	92,588	7,925.47	8.56	79,552	105,624
9	112	38,949	4,509.32	11.58	31,532	46,366
10	68	28,262	4,295.21	15.20	21,197	35,327
11	30	10,585	2,133.38	20.15	7,076	14,094
12	15	6,086	1,956.38	32.15	2,868	9,304
13	3	847	515.49	60.86	0	1,695
14	1	252	n.d.	n.d.	n.d.	n.d.

Fuente: Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como puede observarse en el Cuadro 4.1, el número de mujeres disminuye conforme incrementa el número de hijos nacidos vivos, también se puede observar que el coeficiente de variación para las mujeres que tienen 10 hijos o más es mayor al 15 %, por lo que la inferencia sobre estas mujeres podría ser errónea debido a la falta de precisión estadística, lo cual se puede ser consecuencia de que dichos eventos suelen ser poco frecuentes.

También se puede observar que el 33.33 % de las mujeres de 15 a 49 años de edad declararon no haber tenido a ningún hijo nacido vivo hasta el momento de la encuesta, el 22.45 % declaró haber tenido dos hijos, el 17.08 % mencionó que solo han tenido a un hijo, el 16.23 % tuvieron tres hijos, y finalmente el 10.91 % mencionó que tuvieron cuatro hijos o más. Estos resultados muestran que, en cuanto a preferencias reproductivas, las mujeres prefieren tener dos hijos en vez de uno solo, sin embargo, aún prevalece un porcentaje significativo de mujeres que no tienen hijos.

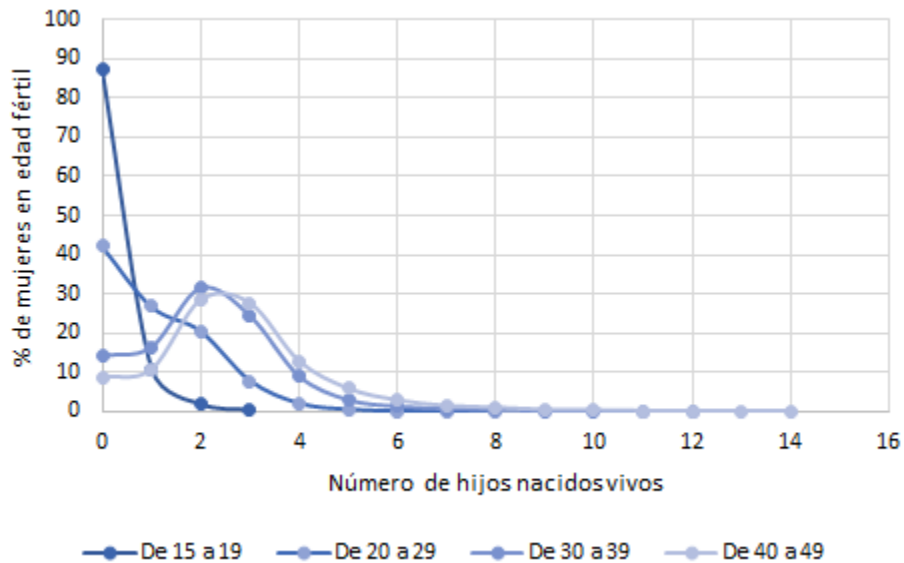
En el segundo capítulo se revisó la relación de algunos factores de riesgo con el número de hijos que tienen las mujeres, por ello antes de construir un modelo que nos permita entender dichas relaciones es necesario analizar los datos más recientes sobre el tema. Para ello se realizará un análisis descriptivo sobre las características anteriormente mencionadas en las mujeres de 15 a 49 años de edad según el número de hijos nacidos vivos, con el fin de entender algunas relaciones o características que puedan influir en la incidencia del embarazo, así como proporcionar evidencia a favor de las hipótesis de investigación.

Uno de los factores analizados es la edad de la mujer al momento de la encuesta. Como anteriormente se mencionó, el INEGI señala que la probabilidad de tener un mayor número de hijos está relacionada con la edad de la mujer, y la razón es que las mujeres más jóvenes están comenzando su etapa reproductiva así como el inicio de su actividad sexual, por lo que el número de hijos en este grupo de edad es bajo en comparación con las mujeres de mayor edad.

La edad de la madre al momento del nacimiento del primer hijo es fundamental para determinar el número de hijos que tienen las mujeres, ya que, según Welti (2000), el promedio de hijos es más alto en las mujeres que fueron madres durante la adolescencia, aunque también influye la percepción de las mujeres sobre el número idóneo de hijos que desea tener y el tiempo que desea esperar entre un nacimiento y otro. En la Figura 4.1 se observa el porcentaje de mujeres en edad fértil según número de hijos nacidos vivos y la edad que tiene al momento de la encuesta.

De acuerdo a la ENADID 2014, el 87.29 % de las mujeres de 15 a 19 años declararon no haber tenido ningún hijo nacido vivo. Este porcentaje disminuye a 42.30 % entre las mujeres de 20 a 29 años, es decir, aproximadamente dos de cada cinco mujeres en este grupo de edad aún no han tenido a su primer hijo. Entre las mujeres de 30 a 39 años el 14.09 % mencionaron en la encuesta que no han tenido a ningún hijo durante toda su vida; y finalmente el 8.69 % de las mujeres de 40 a 49 años de edad declararon no tener hijos. En la Figura 4.1 se puede observar que el porcentaje de mujeres por número de hijos nacidos vivos comienza a crecer conforme incrementa la edad de las mujeres, por tal motivo se sustenta la hipótesis de que a mayor edad mayor será el número de hijos por mujer.

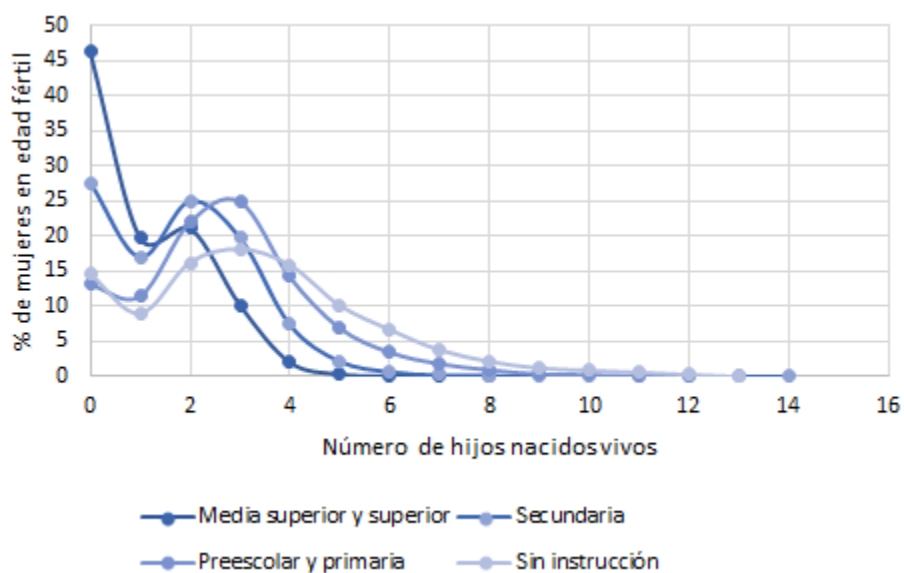
Figura 4.1: Porcentaje de mujeres en edad fértil según número de HNV y grupos de edad, 2014.



**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

También, como han mencionado algunos autores, la educación es primordial para determinar el número de hijos, así como un factor de suma importancia para la prevención un embarazo no planeado y las infecciones de transmisión sexual, ya que dentro de las instancias educativas se pueden adquirir conocimientos sobre salud sexual y reproductiva. En la Figura 4.2 se muestra el porcentaje de mujeres en edad fértil por número de hijos que han tenido según su nivel de escolaridad al momento de la encuesta.

Figura 4.2: Porcentaje de mujeres en edad fértil según número de HNV y nivel de escolaridad, 2014.

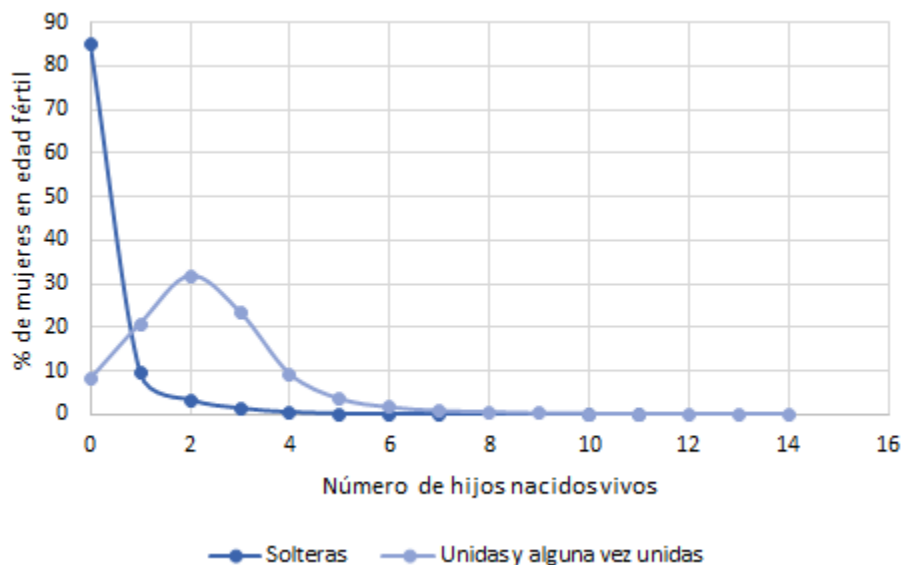


**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Se puede observar que el porcentaje más alto de mujeres sin hijos corresponde a las mujeres con un nivel de educación de media superior y superior (46.3%), por el contrario, el grupo con el porcentaje más bajo corresponde a las mujeres con sólo preescolar y primaria, seguidas de las mujeres sin instrucción (13.1% y 14.6%, respectivamente). Un punto importante por analizar es el comportamiento del tamaño de la descendencia según la escolaridad de la madre, ya que como se muestra en la figura 4.2 las mujeres con un mayor nivel de escolaridad prefieren familias pequeñas, mientras que conforme reduce el nivel escolar el tamaño de la familia tiende a incrementar, lo que sustenta la hipótesis: a menor nivel de educación es mayor el número de hijos que tienen las mujeres.

Respecto a la situación conyugal de las mujeres, el INEGI la clasifica en los siguientes grupos: solteras (aquellas que al momento de la encuesta declararon que no formaban parte de alguna unión o matrimonio), unidas (las mujeres que viven con su pareja en unión libre o matrimonio) y alguna vez unidas (viudas, separadas y divorciadas). Para fines de esta investigación las mujeres unidas y alguna vez unidas formaron un solo grupo, ya que algunos estudios mencionan que son el grupo con mayor probabilidad de concebir (o haber concebido) a un hijo. En la Figura 4.3 se muestra el porcentaje de mujeres en edad fértil por número de hijos nacidos vivos según su situación conyugal durante el año 2014.

Figura 4.3: Porcentaje de mujeres en edad fértil según número de HNV y situación conyugal, 2014.

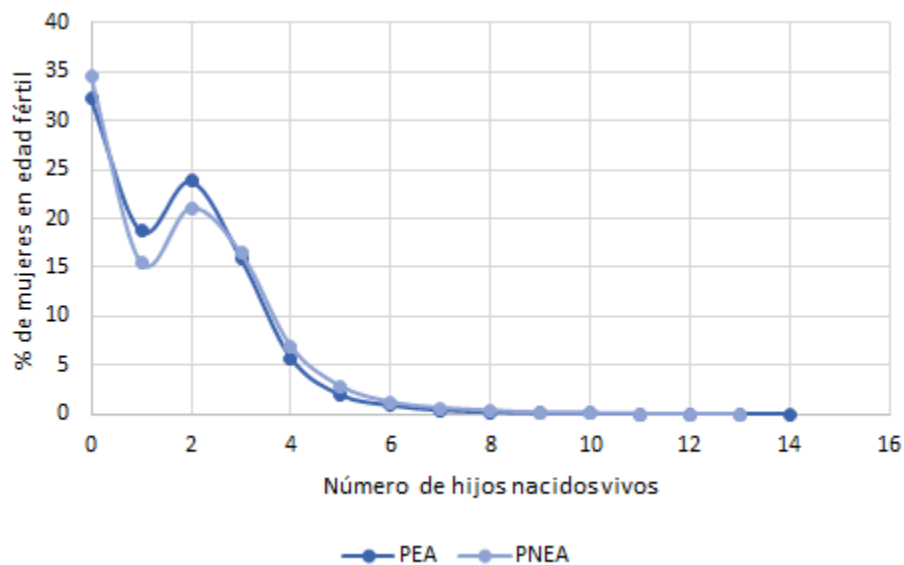


**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como se puede apreciar en la Figura 4.3, el porcentaje de mujeres solteras sin hijos es de aproximadamente 85 por ciento, mientras que el porcentaje de mujeres unidas y alguna vez unidas sin hijos es del 8.1 por ciento. También las mujeres unidas y alguna vez unidas tienden a tener más hijos que las mujeres solteras, por lo que este resultado sustenta la hipótesis de que las mujeres que forman (o formaron) parte de una unión o matrimonio están más alentadas a tener uno o más hijos, en comparación con las mujeres solteras, ya que estas últimas buscan reducir las probabilidades de un embarazo no planeado.

Otro factor relacionado con los niveles de fecundidad es la condición de actividad económica de las mujeres. De acuerdo con el INEGI, a través de la Encuesta Nacional de Ocupación y Empleo (ENOE), la Población Económicamente Activa (PEA) se define como la proporción de la población de 15 años y más que indicaron haber realizado alguna actividad económica o que se enfocó en la búsqueda de empleo durante la semana previa al levantamiento de la encuesta, mientras que la Población No Económicamente Activa (PNEA) se define como la proporción de la población de 15 años y más que al momento de la encuesta indicó que no realizó ninguna actividad económica o que simplemente no buscó empleo durante la semana anterior a la encuesta. En la Figura 4.4 se muestra el comportamiento en el número de hijos de acuerdo a la condición de actividad económica de las mujeres de 15 a 49 años de edad durante el año 2014.

Figura 4.4: Porcentaje de mujeres en edad fértil según número de HNV y condición de actividad económica, 2014.

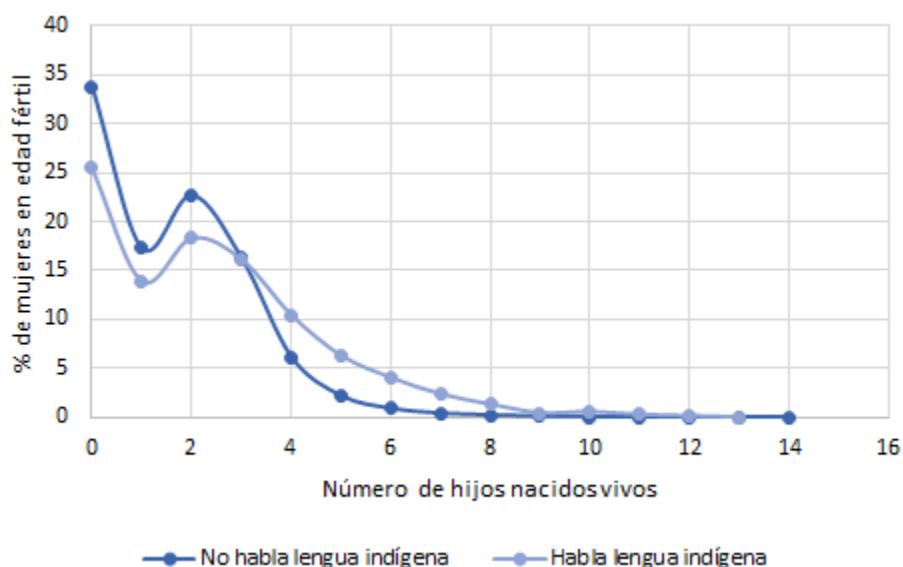


**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

De acuerdo con estos resultados, el porcentaje de mujeres en edad fértil económicamente activas sin hijos es de aproximadamente 32 por ciento, mientras que para las mujeres no económicamente activas éste porcentaje incrementa en 34.4 por ciento. Un patrón importante en esta figura son los niveles reportados para uno o dos hijos, en donde se puede apreciar que el porcentaje es más alto en mujeres económicamente activas, lo que reitera la idea que dentro de los hogares “pequeños” la participación económica de la mujer es de suma importancia en el sustento del hogar (en particular el sustento de los hijos); sin embargo, conforme incrementa el número de hijos ambos comportamientos tienden a ser idénticos. Una posible explicación es que dentro de algunos de estos hogares todavía prevalece la construcción cultural de que la función principal de las mujeres es el cuidado del hogar y de los hijos.

De acuerdo a la revisión realizada, la condición de habla de lengua indígena de las mujeres es un factor que influye en el número de hijos nacidos vivos. Se ha demostrado que este grupo es considerado como altamente vulnerable ante un embarazo no planeado debido a las barreras que tienen para acceder a la información, ya que ésta en su mayoría no se encuentra adaptada para cada una de las lenguas y/o dialectos, además de las reducidas oportunidades para adquirir un método anticonceptivo como medio de planificación familiar. Este hecho se debe a que principalmente en dichas comunidades están muy presentes los usos y costumbres (culturales, ideológicas, religiosas, etc.). En la Figura 4.5 se muestra el porcentaje de mujeres de 15 a 49 años de edad según número de hijos nacidos vivos y condición de habla de lengua indígena para el año 2014.

Figura 4.5: Porcentaje de mujeres en edad fértil según número de HNV y condición de habla de lengua indígena, 2014.

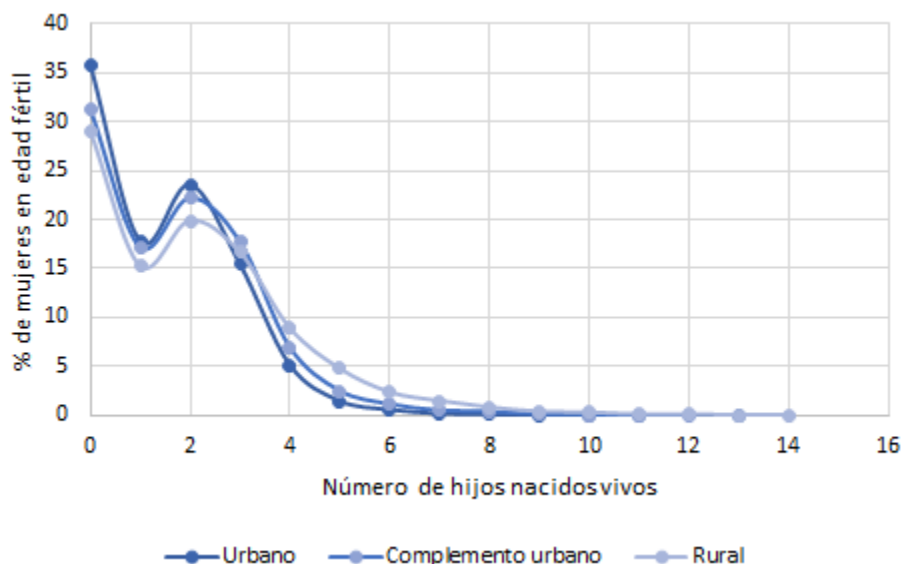


**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como se puede observar en la figura anterior, existe una brecha significativa entre las mujeres sin hijos que hablan alguna lengua indígena y las que no hablan lengua indígena, de aproximadamente 10 puntos porcentuales. También se puede percibir que las mujeres que no hablan lengua indígena tienden a tener menos hijos en comparación con las mujeres que hablan algún dialecto, y cuya diferencia es más notable a partir del cuarto hijo. Con estos resultados se sustenta la hipótesis de que las mujeres en edad fértil que hablan alguna lengua indígena son más susceptibles a tener un mayor número de hijos.

El lugar de residencia es otro factor que influye en la incidencia del embarazo. Según Tuñón (2000), dentro de algunas comunidades rurales se tiene la creencia de que la función principal de las mujeres es ser madres, por lo cual este hecho se convierte en su proyecto de vida a largo plazo. Asimismo el uso de métodos anticonceptivos como medio de planificación familiar es bajo por el limitado acceso a estos. En la Figura 4.6 se muestra el porcentaje de mujeres en edad fértil según número de hijos nacidos vivos y lugar de residencia habitual durante el año 2014.

Figura 4.6: Porcentaje de mujeres en edad fértil según número de HNV y lugar de residencia, 2014.



**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Los resultados indican que no hay diferencias sustanciales entre las localidades urbanas y aquellas consideradas como complemento urbano. Sin embargo, las diferencias tienden a ser más notables en las localidades rurales, donde el porcentaje de mujeres en edad fértil sin hijos es de aproximadamente 29 por ciento, en comparación con las mujeres residentes en localidades urbanas, donde este porcentaje es de 35.8 por ciento. Conforme tiende a incrementar el número de hijos, el porcentaje de mujeres residentes en localidades rurales es más alto en comparación con las mujeres que viven en las áreas urbanas, sustentando la hipótesis de que el promedio de hijos nacidos vivos en las localidades rurales es más alto en comparación con las localidades urbanas.

Finalmente, dentro de este análisis, se examinará el comportamiento sobre el número de hijos de acuerdo al estrato sociodemográfico. Para la ENADID 2014, el INEGI determinó cada estratificación a partir de las características sociodemográficas de los habitantes de las viviendas seleccionadas, así como de las características físicas y el equipamiento de las mismas, expresadas por medio de 34 indicadores, para lo cual se emplearon métodos estadísticos multivariados para la construcción de cuatro estratos (bajo, medio bajo, medio alto y alto)<sup>1</sup>:

1. Población que tiene derecho a recibir servicios médicos en alguna institución de salud pública o privada, excepto Seguro Popular
2. Población derechohabiente a los servicios de salud.
3. Población de 3 a 14 años de edad que asiste a la escuela.
4. Población de 15 a 24 años de edad que asiste a la escuela.
5. Población de 8 a 14 años de edad que sabe leer y escribir.

<sup>1</sup><http://www.beta.inegi.org.mx/contenidos/proyectos/enchogares/especiales/enadid/2014/doc/dm.enadid14.pdf>



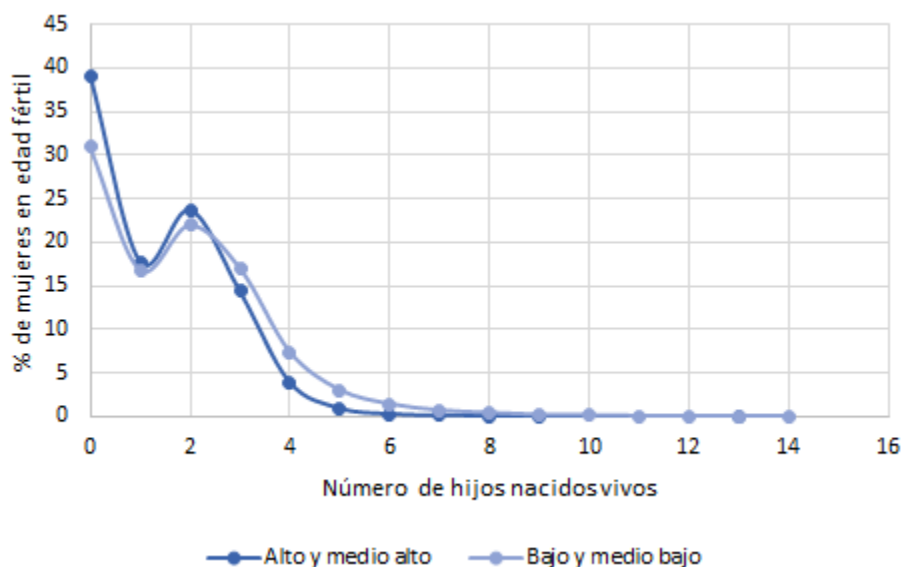
6. Población de 15 años y más que aprobó algún grado de escolaridad diferente al nivel preescolar.
7. Población de 15 años y más que tiene como máxima escolaridad seis grados aprobados en primaria.
8. Población de 15 años y más que tiene como máxima escolaridad tres grados aprobados en secundaria.
9. Grado promedio de escolaridad de la población.
10. Población de 12 años y más que trabajó; tenía trabajo, pero no trabajó; o buscó trabajo en la semana de referencia.
11. Población femenina de 12 años y más que trabajó; tenía trabajo pero no trabajó; o buscó trabajo en la semana de referencia.
12. Tasa de ocupación.
13. Población no ocupada de 12 a 17 años entre la población total de este rango.
14. Población ocupada de 18 años y más entre la población total ocupada.
15. Viviendas particulares habitadas sin hacinamiento.
16. Viviendas particulares habitadas con piso de cemento o firme, madera, mosaico u otro material.
17. Viviendas particulares habitadas que usan para dormir entre 2 y 25 cuartos.
18. Viviendas particulares habitadas que tienen más de un cuarto.
19. Viviendas particulares habitadas que tienen entre 3 y 25 cuartos.
20. Viviendas particulares habitadas que disponen de luz eléctrica.
21. Viviendas particulares habitadas que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de ella, pero dentro del terreno.
22. Viviendas particulares habitadas que tienen excusado, retrete, sanitario, letrina u hoyo negro.
23. Viviendas particulares habitadas que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.
24. Viviendas particulares habitadas que disponen de drenaje conectado a la red pública.
25. Viviendas particulares habitadas que disponen de excusado con descarga directa de agua.
26. Viviendas particulares habitadas que disponen de luz eléctrica, agua entubada dentro o fuera de la vivienda, pero dentro del terreno, así como drenaje.
27. Viviendas particulares habitadas que no se encuentran en situación de hacinamiento a nivel manzana.
28. Viviendas particulares habitadas que disponen de televisor.

29. Viviendas particulares habitadas que disponen de automóvil o camioneta.
30. Viviendas particulares habitadas que disponen de teléfono celular.
31. Viviendas particulares habitadas que disponen de teléfono celular y teléfono fijo.
32. Viviendas particulares habitadas que disponen de radio, televisor, refrigerador y lavadora.
33. Viviendas particulares habitadas que disponen de radio y televisor.
34. Viviendas particulares habitadas que disponen de todos los bienes.

El Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) define a una persona en situación de pobreza cuando tiene al menos una carencia social (en los seis indicadores de rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos de la vivienda y acceso a la alimentación) y su ingreso es insuficiente para adquirir los bienes y servicios que requiere para satisfacer sus necesidades alimentarias y no alimentarias<sup>2</sup>. De esta forma, el estrato sociodemográfico puede ser una variable que determina algunas características de los niveles de pobreza dentro de una región.

Las investigaciones apuntan a una relación directa entre la pobreza y la fecundidad (a mayores niveles de pobreza mayor tasa de fecundidad) ya que como lo menciona Di Cesare (2007) en los estratos socioeconómicos más altos las mujeres presentan un calendario sexual y reproductivo más envejecido y tienen una mayor incidencia en el uso de métodos anticonceptivos así como al tener relaciones sexuales con mayor cuidado con el fin de protegerse de una infección de transmisión sexual o de un embarazo no planeado. En la Figura 4.7 se muestra el porcentaje de mujeres según número de hijos y estrato sociodemográfico al que pertenecen en el año 2014.

Figura 4.7: Porcentaje de mujeres en edad fértil según número de HNV y estrato sociodemográfico, 2014.



**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

<sup>2</sup>Medición de la Pobreza. Glosario. Disponible en: <https://www.coneval.org.mx/Medicion/Paginas/Glosario.aspx>

En la figura anterior se puede observar que la brecha que existe entre las mujeres según estrato socio-demográfico y que no tienen hijos es de aproximadamente diez puntos porcentuales. Si bien la diferencia que existe entre ambos grupos en cuanto al número de hijos es mínima, se puede apreciar un aumento en el porcentaje de mujeres con cuatro o más hijos que pertenecen al estrato sociodemográfico bajo y muy bajo, lo cual concuerda con nuestra hipótesis sobre los niveles de pobreza y el promedio de hijos por mujer.

## 4.2. Selección del modelo

Dada la viabilidad de las variables a introducir en el Modelo de Regresión, podemos definir estas como:

- Edad al momento de la encuesta: 15, 16, 17,  $\dots$ , 47, 48 y 49 años.
- Nivel de escolaridad: 0-Media superior o superior, 1-Secundaria, 2-Preescolar y primaria, 3-Sin instrucción.
- Situación conyugal: 0-Soltera, 1-Unida y alguna vez unida.
- Condición de actividad: 0-Económicamente activa, 1-No económicamente activa.
- Condición de habla de lengua indígena: 0-No habla lengua indígena, 1-Habla lengua indígena.
- Lugar de residencia habitual: 0-Urbano, 1-Complemento urbano, 2-Rural.
- Estrato sociodemográfico: 0-Alto y medio alto, 1-Bajo y medio bajo.

De este modo, como un primer paso en este análisis se define un Modelo de Regresión Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 Edad_i + \sum_{j=1}^3 \beta_{j+1} Esc_{i,j} + \beta_5 SitCony_i + \beta_6 CondAct_i + \beta_7 CondHLL_i + \sum_{j=1}^2 \beta_{j+7} TLoc_{i,j} + \beta_{10} EstSoc_i$$

Suponiendo que la componente sistemática del modelo de regresión es correcta con base en el análisis descriptivo realizado, fue necesario evaluar si existe algún problema con la dispersión de los datos. Como un primer paso se analiza la media y la varianza del total de hijos nacidos vivos entre las mujeres de 15 a 49 años de edad. De acuerdo con los datos de la ENADID 2014, el promedio de hijos nacidos vivos en las mujeres en edad fértil es de 1.63 y su varianza es de 2.57. El cociente de la varianza entre la media es de 1.58, lo que podría sugerir que los datos no presentan sobredispersión, no obstante es necesario validar este resultado con pruebas más robustas.

Como se mencionó en el capítulo 3, un modelo que permite identificar los problemas de sobredispersión es el modelo de regresión con distribución Binomial Negativa. En el Cuadro 4.2 se muestran los resultados para ambos modelos (Poisson y Binomial Negativa) con el fin de determinar si existen problemas de dispersión en los datos.

Cuadro 4.2: Modelos de regresión Poisson y Binomial Negativa, 2014.

Variable	Distribución Poisson					Distribución Binomial Negativa				
	Estimate	Std. Error	z value	P-Value		Estimate	Std. Error	z value	P-Value	
Intercepto	-2.62	0.02	-162.82	0.00	***	-2.62	0.02	-162.82	0.00	***
Edad	0.04	0.00	117.18	0.00	***	0.04	0.00	117.17	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	0.28	0.01	42.39	0.00	***	0.28	0.01	42.39	0.00	***
Preescolar o primaria	0.41	0.01	54.84	0.00	***	0.41	0.01	54.84	0.00	***
Sin instrucción	0.49	0.01	32.97	0.00	***	0.49	0.01	32.97	0.00	***
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	1.78	0.01	136.74	0.00	***	1.78	0.01	136.74	0.00	***
Condición de actividad										
Económicamente activa (referencia)										
No económicamente activa	0.03	0.01	4.86	0.00	***	0.03	0.01	4.86	0.00	***
Condición de habla de lengua indígena										
No habla lengua indígena (referencia)										
Habla lengua indígena	0.08	0.01	7.74	0.00	***	0.08	0.01	7.74	0.00	***
Lugar de residencia habitual										
Urbano (referencia)										
Complemento urbano	0.03	0.01	4.84	0.00	***	0.03	0.01	4.84	0.00	***
Rural	0.10	0.01	14.18	0.00	***	0.10	0.01	14.18	0.00	***
Estrato sociodemográfico										
Alto y medio alto (referencia)										
Bajo y medio bajo	0.07	0.01	9.18	0.00	***	0.07	0.01	9.18	0.00	***
Null deviance	159,463 con 89,903 grados de libertad					159,446 con 89,903 grados de libertad				
Residual deviance	65,803 con 89,893 grados de libertad					65,795 con 89,893 grados de libertad				
$R^2$ Nagelkerke	0.6656					0.6203				
AIC	228,946					228,951				
$\theta$	-					13,741				

**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como se puede apreciar, en ambos modelos se obtuvieron coeficientes de regresión similares, es decir, ambos modelos explican de manera similar la tasa de fecundidad. Además, en ambos modelos las variables predictoras son altamente significativas, es decir, con p-values menores al 1%. Un punto importante por analizar es el valor de  $\theta$  en el modelo con distribución binomial negativa. En este modelo el valor de  $\theta$  es de 13,741<sup>3</sup>, es decir  $\alpha = \frac{1}{\theta} = 0,0001$ , por lo que planteamos la siguiente prueba de hipótesis:  $H_0 : \alpha = 0$  vs  $H_1 : \alpha > 0$ . De esta forma, el estadístico de prueba arrojó un valor de -3.2074 con un p-value de 0.5, es decir, no se rechaza  $H_0$ , por lo que se concluyó que los datos no presentaron problemas de sobredispersión. Por lo tanto, se pudo asumir que los datos cumplieron con el supuesto de  $E[Y] = Var[Y]$ .

Después fue necesario analizar si los datos tuvieron problemas de exceso de ceros. Un modelo que permite analizar el problema de exceso de ceros es el *Zero Inflated Poisson*. En este modelo se consideró que la edad, la escolaridad y la situación conyugal son las tres variables que determinan la probabilidad de que una mujer no tenga hijos nacidos vivos (ya que son las tres variables que mayor peso tienen en cuanto a la clasificación de mujeres sin hijos, esto de acuerdo al análisis descriptivo realizado). En el Cuadro 4.3 se muestran los resultados de ambos modelos (Poisson y *Zero Inflated Poisson*).

<sup>3</sup>El valor de  $\theta$  fue estimado con la función `glm.nb()` de la librería MASS en R

Cuadro 4.3: Modelos de regresión Poisson y *Zero Inflated Poisson*, 2014.

Variable	Modelo Poisson					Modelo <i>Zero Inflated Poisson</i>					
	Estimate	Std. Error	z value	P-Value		Estimate	Std. Error	z value	P-Value		
<b>Count model</b>											
Intercepto	-2.62	0.02	-162.82	0.00	***	-1.66	0.03	-51.88	0.00	***	
Edad	0.04	0.00	117.18	0.00	***	0.03	0.00	101.58	0.00	***	
Nivel de escolaridad											
Media superior y superior (referencia)											
Secundaria	0.28	0.01	42.39	0.00	***	0.26	0.01	38.56	0.00	***	
Preescolar o primaria	0.41	0.01	54.84	0.00	***	0.39	0.01	52.15	0.00	***	
Sin instrucción	0.49	0.01	32.97	0.00	***	0.49	0.02	32.45	0.00	***	
Situación conyugal											
Soltera (referencia)											
Unida y alguna vez unida	1.78	0.01	136.74	0.00	***	1.00	0.03	35.08	0.00	***	
Condición de actividad											
Económicamente activa (referencia)											
No económicamente activa	0.03	0.01	4.86	0.00	***	0.05	0.01	9.88	0.00	***	
Condición de habla de lengua indígena											
No habla lengua indígena (referencia)											
Habla lengua indígena	0.08	0.01	7.74	0.00	***	0.08	0.01	7.83	0.00	***	
Lugar de residencia habitual											
Urbano (referencia)											
Complemento urbano	0.03	0.01	4.84	0.00	***	0.03	0.01	4.65	0.00	***	
Rural	0.10	0.01	14.18	0.00	***	0.10	0.01	13.64	0.00	***	
Estrato sociodemográfico											
Alto y medio alto (referencia)											
Bajo y medio bajo	0.07	0.01	9.18	0.00	***	0.06	0.01	8.54	0.00	***	
<b>Zero-inflation model</b>											
Intercepto	-	-	-	-		6.72	0.38	17.64	0.00	***	
Edad	-	-	-	-		-0.24	0.02	-13.20	0.00	***	
Nivel de escolaridad											
Media superior y superior (referencia)											
Secundaria	-	-	-	-		-1.04	0.09	-11.82	0.00	***	
Preescolar o primaria	-	-	-	-		-0.84	0.15	-5.79	0.00	***	
Sin instrucción	-	-	-	-		1.01	0.35	2.92	0.00	**	
Situación conyugal											
Soltera (referencia)											
Unida y alguna vez unida	-	-	-	-		-18.92	270.43	-0.07	0.94		
Null deviance	159,463 con 89,903 grados de libertad							n.d.			
Residual deviance	65,803 con 89,893 grados de libertad							n.d.			
R <sup>2</sup> Nagelkerke	0.6656							0.6180			
AIC	228,946							223,847			

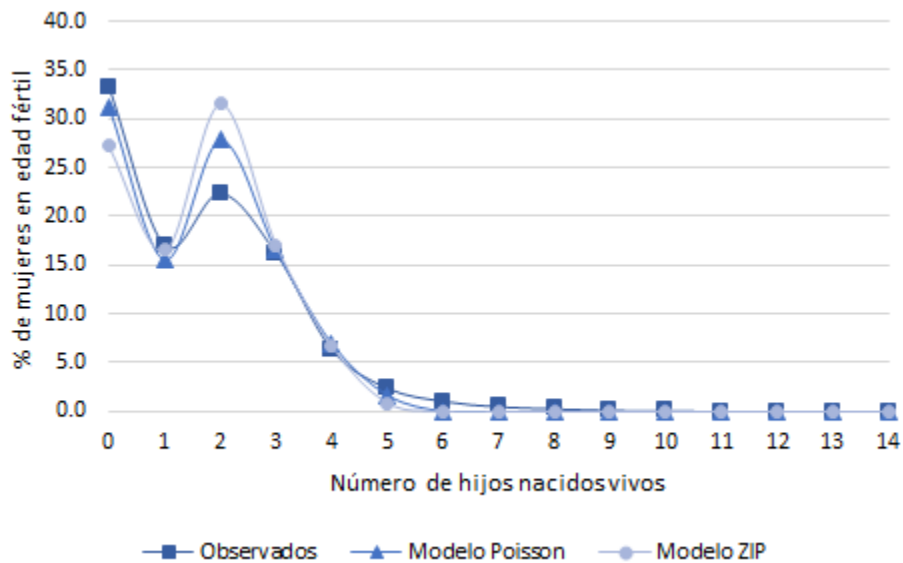
**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Al observar los resultados de ambos modelos se pudo apreciar diferencias significativas en los coeficientes de regresión, por lo que, bajo una primera inspección, ambos modelos explicaron la tasa de fecundidad de diferente forma. Es importante mencionar que las variables en los modelos de conteo (tanto Poisson como *Zero Inflated Poisson*) son altamente significativos. Sin embargo, para el modelo que explica los ceros en el *Zero Inflated Poisson* la situación conyugal de las mujeres no fue una variable relevante para determinar la probabilidad de que una mujer no tenga ningún hijo nacido vivo, aunque al considerar el esquema de diseño de la muestra si lo es<sup>4</sup>.

<sup>4</sup>Apéndice A

Para determinar si el modelo *Zero Inflated Poisson* tiene un mejor ajuste que el modelo Poisson se hizo uso de la prueba de Vuong. De esta forma el estadístico de prueba fue de  $-33.8402$  con un p-value de  $0.0000$ , por lo que la prueba favoreció al modelo Poisson; también la  $R^2$  de Nagelkerke indicó que el modelo Poisson tiene un mejor ajuste que el modelo *Zero Inflated Poisson*. En la Figura 4.8 se muestra el ajuste de ambos modelos.

Figura 4.8: Ajuste de los modelos Poisson y *Zero Inflated Poisson*, 2014.



**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

En la Figura 4.8 se observa una subestimación en el porcentaje de mujeres en edad fértil sin hijos en ambos modelos, aunque en el modelo *Zero Inflated Poisson* fue mayor la diferencia entre el valor observado y el valor estimado. Por otro lado, en cuanto a las mujeres con dos hijos se pudo observar una sobrestimación importante en el modelo ZIP a tal punto que incluso la estimación fue mayor que el porcentaje de mujeres sin hijos, el cual fue totalmente lo contrario a lo que muestran los datos, y aunque el Criterio de Información de Akaike (AIC) muestra que el modelo ZIP es mejor que el modelo Poisson (223,847 vs 228,946). Por lo tanto, de acuerdo al estadístico de Vuong, la  $R^2$  de Nagelkerke y al análisis gráfico de las estimaciones de ambos modelos, se decidió trabajar con el modelo de regresión Poisson.

Ahora bien, sabemos que la ENADID 2014 tiene un esquema de muestreo de diseño complejo, por lo que las observaciones no son independientes por completo, esto debido a que la selección de estas están sujetas a una agrupación por estratos y conglomerados, por lo tanto las observaciones que pertenecen a un mismo estrato y/o conglomerado pueden compartir las mismas características; además, las personas que fueron seleccionadas para la muestra no tienen las mismas probabilidades de selección, por lo que es necesario analizar la estimación del modelo de regresión Poisson bajo este esquema de muestreo ya que podría cambiar el peso de los coeficientes de regresión e incluso la significancia de estos. En el Cuadro 4.4 se muestran los resultados de dos modelos de regresión, el primero no considera el efecto del diseño de la muestra y el segundo asume que la muestra fue construida a partir de un diseño complejo.

Cuadro 4.4: Efecto del diseño de la muestra en el Modelo de Regresión Poisson, 2014.

Variable	Sin efecto del diseño de la muestra					Con efecto del diseño de la muestra				
	Estimate	Std. Error	z value	P-Value		Estimate	Std. Error	t value	P-Value	
Intercepto	-2.62	0.02	-162.82	0.00	***	-2.63	0.02	-119.46	0.00	***
Edad	0.04	0.00	117.18	0.00	***	0.04	0.00	102.37	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	0.28	0.01	42.39	0.00	***	0.28	0.01	39.83	0.00	***
Preescolar o primaria	0.41	0.01	54.84	0.00	***	0.42	0.01	50.97	0.00	***
Sin instrucción	0.49	0.01	32.97	0.00	***	0.51	0.02	26.53	0.00	***
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	1.78	0.01	136.74	0.00	***	1.78	0.02	77.54	0.00	***
Condición de actividad										
Económicamente activa (referencia)										
No Económicamente activa	0.03	0.01	4.86	0.00	***	0.03	0.01	5.31	0.00	***
Condición de habla de lengua indígena										
No habla lengua indígena (referencia)										
Habla lengua indígena	0.08	0.01	7.74	0.00	***	0.09	0.01	6.15	0.00	***
Lugar de residencia habitual										
Urbano (referencia)										
Complemento urbano	0.03	0.01	4.84	0.00	***	0.04	0.01	5.40	0.00	***
Rural	0.10	0.01	14.18	0.00	***	0.11	0.01	12.47	0.00	***
Estrato sociodemográfico										
Alto y medio alto (referencia)										
Bajo y medio bajo	0.07	0.01	9.18	0.00	***	0.08	0.01	9.99	0.00	***
Null deviance	159,463 con 89,903 grados de libertad					159,630 con 89,903 grados de libertad				
Residual deviance	65,803 con 89,893 grados de libertad					65,893 con 89,893 grados de libertad				
$R^2$ Nagelkerke	0.6656					0.6664				
AIC	228,946					226,534				

**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como puede observarse, la orientación de los coeficientes de regresión fue la misma, sin embargo, se llegan a observar algunas diferencias en la estimación de los coeficientes entre ambos enfoques. Un resultado importante es que el nivel de significancia de las variables regresoras en ambos enfoques fue muy alta, por lo que se considera que ambos modelos quedaron bien especificados.

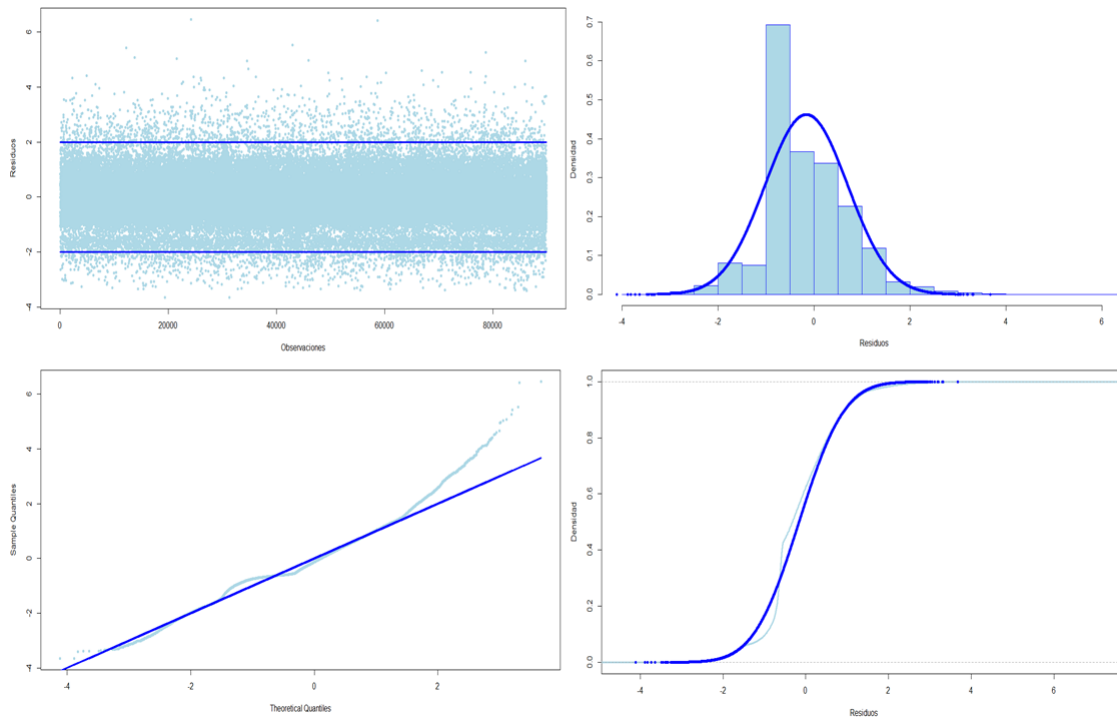
Si se observan los valores de la devianza<sup>5</sup>, este valor resultó ser menor en el modelo que no considera el efecto del diseño de la muestra (65,803 vs 65,893), aunque la diferencia entre ambos modelos es mínima por lo que se pudo concluir bajo este criterio que ambos modelos se ajustan de manera similar. Según el Criterio de Información de Akaike (AIC) éste resultó ser menor el modelo bajo el efecto del diseño de la muestra (226,534 vs 228,946), por lo tanto al considerar el efecto del diseño de la muestra se obtuvo una mayor capacidad predictiva. Finalmente la  $R^2$  de Nagelkerke fue ligeramente mayor en el segundo modelo, por lo que se concluyó que al considerar el diseño de la muestra se obtiene una mejor precisión en la estimación.

Otra forma de evaluar la calidad del modelo es a través del análisis de los residuos. Estos se definen como la diferencia entre los valores observados y los valores estimados por el modelo:  $r_i = y_i - \hat{y}_i$ . En general la distribución de los residuos está sesgada y en la mayoría de los casos no tienen una distribución normal. Es así como se propone utilizar el residuo de Anscombe, el cual sustituye el valor de  $y$  por la

<sup>5</sup>La devianza es considerada como una medida de carencia de ajuste del modelo. Entre más bajo sea el valor de la devianza mejor será el ajuste del modelo.

función  $A(y) = \int_{-\infty}^y V^{-\frac{1}{3}}(t)dt$ , donde  $V(t)$  es la función varianza. Esta función busca aproximarse a una distribución normal. En la Figura 4.9 se muestra un análisis gráfico de los residuos de Anscombe para el modelo de regresión Poisson (gráfica de dispersión de los residuos, histograma, qq-plot y la función de distribución empírica).

Figura 4.9: Distribución de los residuos de Anscombe del Modelo de Regresión Poisson bajo el esquema de muestreo, 2014.



**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Como primer punto se analizaron las estadísticas descriptivas de los residuos del modelo. El valor mínimo es de -3.6506 y el máximo es de 6.4590. El primer cuantil se alcanza en -0.6609, la mediana es de -0.3239 y el tercer cuantil es de 0.3657. Estos resultados mostraron que existe un problema de simetría con los datos, es decir, existe un mayor peso en los residuos positivos.

En la gráfica de dispersión de los residuos se muestra que hay una mayor concentración de estos entre los valores -2 y 2, acumulando el 96.66 % de las observaciones, asimismo se puede observar que los residuos cumplen el supuesto de aleatoriedad y varianza constante bajo un análisis gráfico, sin embargo, este resultado debe ser reforzado bajo pruebas estadísticas más rigurosas. Otra gráfica a analizar es el histograma. El histograma de los residuos parece converger asintóticamente a una distribución normal, y aunque se muestra un sesgo importante del lado derecho, se puede asumir que los residuos en este modelo además de cumplir el supuesto de aleatoriedad y varianza constante, también cumplen aproximadamente el supuesto de normalidad.

La qq-plot (Gráfica Cuantil-Cuantil) muestra que la distribución de los residuos tiene una cola derecha muy pesada, aunque en general la qq-plot parece ajustarse de forma significativa a la qq-line (Línea



Cuantil-Cuantil), lo que podría sugerir que los datos siguen una distribución normal. Finalmente la gráfica de la función de distribución empírica muestra nuevamente el peso de la cola derecha, asimismo se muestra que el 50% de la densidad se acumula aproximadamente en el valor cero (recordando que la mediana es de -0.6609), por lo que se concluyó que los residuos Anscombe del modelo de regresión Poisson que considera el efecto del diseño de la muestra cumplen aproximadamente con el supuesto de normalidad.

Así, para cumplir con los objetivos de este trabajo de investigación se utilizaron las estimaciones del Modelo de Regresión Poisson bajo un esquema de muestreo con diseño complejo.

### 4.3. Resultados

Comúnmente la interpretación de resultados se realiza con base en la razón de tasas (rate ratio). Si consideramos una variable  $x_j = 1$  cuando hay presencia de un atributo  $j$  y  $x_j = 0$  cuando hay ausencia de este, entonces la razón de tasas para comparar la presencia del atributo en comparación la ausencia del mismo es:

$$RR = \frac{\mu_1}{\mu_0} = \frac{E[y_i|x_j = 1]}{E[y_i|x_j = 0]} = e^{\beta_j}$$

Para la interpretación de resultados en este trabajo de investigación se utilizaron las estimaciones del Modelo de Regresión Poisson bajo un esquema de muestreo con diseño complejo ya que, de acuerdo al análisis realizado, cumple con los supuestos establecidos por el modelo. Así, en el Cuadro 4.5 se muestran los resultados que explican el comportamiento de la tasa de fecundidad.

De esta forma en cuanto a la edad, la tasa de fecundidad incrementa 1.0375 unidades por cada unidad que incrementa la edad de las mujeres (con un intervalo de confianza entre 1.037 y 1.038). De esta forma quedó demostrado que a mayor edad en las mujeres mayor es el número de hijos nacidos vivos que tienen, el cual coincide con el trabajo de Welti (2000 y 2010).

También se observa que la tasa de fecundidad tiende a incrementar conforme reduce el nivel de escolaridad de las mujeres. De este modo, la tasa de fecundidad entre las mujeres que tienen secundaria representa 1.33 veces la tasa de fecundidad en las mujeres con un nivel de escolaridad de media superior o superior; la tasa de fecundidad en las mujeres que tienen sólo el preescolar o la primaria corresponde a 1.52 veces la tasa de fecundidad entre las mujeres con un nivel de estudio equivalente a media superior o superior; y finalmente la tasa de fecundidad en las mujeres que no cuentan con ningún nivel de instrucción representan 1.66 veces la tasa de fecundidad en las mujeres con preparatoria, bachillerato o un nivel de estudios superior.

Este resultado validó la hipótesis de que a mayor nivel de educación en las mujeres es menor la probabilidad de tener algún hijo, por lo que asistir a una institución educativa es fundamental para disminuir la tasa de fecundidad (Stupp y Cáceres, 2001).

En cuanto a la situación conyugal se observó que la tasa de fecundidad en mujeres unidas y alguna vez unidas es 5.95 veces la tasa de fecundidad en las mujeres solteras, con un intervalo de (5.73, 6.18) a un nivel de confianza del 90 por ciento. Este resultado valida la hipótesis de que la fecundidad incrementa en las mujeres casadas o unidas y que el promedio de hijos es más alto en este grupo en comparación con las mujeres solteras (Belmonte, 2007; Curtis y Blanc, 1997).

Para la condición de actividad económica, la tasa de fecundidad entre las mujeres que pertenecen a la población no económicamente activa es 1.03 veces la tasa de fecundidad entre las mujeres que pertenecen a la población económicamente activa (con un intervalo de confianza entre 1.02 y 1.04). Este resultado indica que la tasa de fecundidad es más alta entre las mujeres que se dedican a los quehaceres domésticos, que son estudiantes, pensionadas o jubiladas, que cuentan con un impedimento físico para trabajar u otros motivos, en comparación con aquellas que trabajan o buscan algún empleo (ONU, 2001).

Según la condición de habla de lengua indígena de las mujeres en edad fértil, la tasa de fecundidad en las mujeres que hablan alguna lengua indígena o dialecto es 1.09 veces la tasa de fecundidad entre las mujeres que no hablan lengua indígena (con un intervalo de confianza entre 1.06 y 1.11). Este resultado muestra la brecha que existe en la calidad y el acceso a los servicios de planificación familiar, ya que como mencionan algunos autores (Valdivieso, 2012; Szasz y Lerner, 2010; Meneses *et al.*, 2017), este grupo se caracteriza por ser un sector de la población vulnerable en el tema de anticoncepción, ya que dentro de estas comunidades el acceso a los anticonceptivos es limitado, y entre las mujeres que usan algún tipo de método la mayoría hacen uso de ellos por prescripción médica o preferencia de su pareja, lo que limita la oportunidad de una planificación familiar libre e informada.

De acuerdo al lugar de residencia habitual, la tasa de fecundidad en mujeres que viven en localidades consideradas como complemento urbano es 1.04 veces la tasa de fecundidad en mujeres que viven en zonas urbanas, mientras que la tasa de fecundidad en áreas rurales es 1.12 veces la tasa de fecundidad en áreas urbanas. Este resultado valida la hipótesis de que el promedio de hijos en localidades rurales es más alto que en las localidades urbanas, ya que algunos estudios mencionan que esto se debe a que en las comunidades rurales se piensa que el embarazo está ligado a cuestiones culturales e ideológicas (Tuñón, 2000).

Conforme al nivel sociodemográfico, la tasa de fecundidad entre las mujeres que viven en estratos sociodemográficos más bajos es 1.08 veces la tasa de fecundidad en los estratos sociodemográficos altos. Este resultado demuestra que a mayor pobreza mayor fecundidad, ya que como menciona Di Cesare y Rodríguez (2006), la población dentro de los estratos sociodemográficos más bajos son clasificados como un grupo vulnerable en materia de salud sexual y reproductiva a consecuencia de un bajo acceso y disponibilidad de los métodos anticonceptivos, lo que en la mayoría de las veces finaliza en un embarazo no planeado.

Finalmente se concluye que el 66.64 por ciento de la variación de los datos fue explicado por el Modelo de Regresión Poisson, por lo que en términos generales se consideró como un buen modelo.

Cuadro 4.5: *Rate Ratio* para explicar los patrones de fecundidad, 2014.

Variable	<i>Rate Ratio</i> (RR)	IC (90%)	P-Value	
Intercepto	0.0723	(0.070, 0.075)	0.0000	***
Edad	1.0375	(1.037, 1.038)	0.0000	***
Nivel de escolaridad				
Media superior y superior (referencia)				
Secundaria	1.3286	(1.313, 1.344)	0.0000	***
Preescolar o primaria	1.5241	(1.504, 1.545)	0.0000	***
Sin instrucción	1.6598	(1.609, 1.713)	0.0000	***
Situación conyugal				
Soltera (referencia)				
Unida y alguna vez unida	5.9474	(5.727, 6.176)	0.0000	***
Condición de actividad				
Económicamente activa (referencia)				
No económicamente activa	1.0307	(1.021, 1.04)	0.0000	***
Condición de habla de lengua indígena				
No habla lengua indígena (referencia)				
Habla lengua indígena	1.0890	(1.064, 1.114)	0.0000	***
Lugar de residencia habitual				
Urbano (referencia)				
Complemento urbano	1.0430	(1.030, 1.056)	0.0000	***
Rural	1.1218	(1.105, 1.139)	0.0000	***
Estrato sociodemográfico				
Alto y medio alto (referencia)				
Bajo y medio bajo	1.0822	(1.068, 1.096)	0.0000	***
Null deviance	159,630 con 89,903 grados de libertad			
Residual deviance	65,893 con 89,893 grados de libertad			
$R^2$ Nagelkerke	0.6664			
AIC	226,534			

**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

# Capítulo 5

## Conclusiones

La presente tesina tuvo como objetivo principal determinar las causas sociales, demográficas, económicas, culturales, etc. que influyen sobre el número de hijos e hijas que tendrá una mujer durante su vida reproductiva, con la finalidad de identificar a los grupos con mayor exposición a los cambios en los niveles de fecundidad, y para cumplir con dicho objetivo fue necesario analizar los distintos modelos matemáticos que permitieran explicar dichas relaciones causales, así como examinar la información que se encuentra en la Encuesta Nacional de la Dinámica Demográfica 2014. Es así como el presente trabajo de investigación llegó a los siguientes resultados:

De acuerdo al modelo estadístico, es necesario resaltar la importancia de la validación de los supuestos del modelo seleccionado. Al seleccionar un modelo de regresión Poisson para modelar una variable de tipo conteo es primordial validar el supuesto de  $E[y_i|x_i] = V[y_i|x_i]$ , ya que al no cumplirse dicho supuesto se puede llegar a una interpretación errónea de los resultados, así como en el nivel de significancia de estos. Existen modelos que permiten corregir los problemas de sobredispersión (el modelo con distribución binomial negativa) así como el exceso de ceros en la variable de conteo (*Zero Inflated Poisson* y *Zero Inflated Negative Binomial*). En este trabajo se demostró que se puede asumir el supuesto de  $E[Y] = Var[Y]$ , también que no existe un exceso de ceros en el número de hijos nacidos vivos, por lo que el Modelo de Regresión Poisson cumple con las características necesarias para alcanzar el objetivo de esta investigación.

No obstante, al considerar una encuesta construida a partir de un esquema de muestreo con diseño complejo como la fuente primaria de información para el Modelo de Regresión Poisson, es importante considerar los elementos principales del esquema de muestreo (la unidad primaria de muestreo, el estrato del diseño y los pesos muestrales) ya que se puede obtener otra visión sobre la calidad de la estimación, es decir, si el modelo refleja las características de la población objetivo o si se trata de estimaciones sesgadas. Si bien es cierto que algunos coeficientes de regresión son similares al considerar un muestreo aleatorio simple en comparación con un muestreo bajo un diseño complejo, en otros se llega a apreciar cambios sustanciales en la magnitud de los coeficientes, así como en los niveles de significancia.

En cuanto a los resultados obtenidos, se demostró que cuando incrementa la edad de las mujeres incrementa la tasa de fecundidad, validando los resultados obtenidos por Welti (2000, 2012). Es importante mencionar que en los últimos años se ha observado un incremento en la tasa de fecundidad adolescente por lo que se deben reforzar las políticas públicas dirigidas a este grupo de la población.

El nivel de escolaridad de las mujeres de 15 a 49 años de edad resultó ser una de las variables más importantes para determinar el número de hijos que tienen. Se ha demostrado que entre menor sea el nivel de escolaridad de las mujeres, mayor será la tasa de fecundidad. De acuerdo al artículo Tercero de la Constitución Política de los Estados Unidos Mexicanos: “Toda persona tiene derecho a recibir educación. El Estado -Federación, Estados, Ciudad de México y Municipios-, impartirá educación preescolar, primaria, secundaria y media superior. La educación preescolar, primaria y secundaria conforman la educación básica; ésta y la media superior serán obligatorias”, sin embargo, los resultados muestran que aún persiste el rezago educativo entre las mujeres en edad fértil y que a su vez se refleja en el número de hijos que tienen, por lo que es necesario que se siga fomentando la educación a un nivel de media superior y/o superior.

También se ha demostrado que las mujeres en edad fértil unidas y alguna vez unidas presentan una tasa de fecundidad más alta que las mujeres solteras, validando los resultados Belmonte (2007) y de Curtis y Blanc (1997) al mencionar que el promedio de hijos es más alto en las mujeres unidas y alguna vez unidas en comparación con las mujeres solteras, así como el uso efectivo de métodos anticonceptivos es más alto en las mujeres solteras, ya que estas están más alentadas en prevenir un embarazo no planeado.

Asimismo los resultados del Modelo Poisson mostraron que la tasa de fecundidad es más alta entre las mujeres que se dedican a los quehaceres domésticos, que son estudiantes, pensionadas o jubiladas, que cuentan con un impedimento físico para trabajar u otros motivos, en comparación con aquellas que trabajan o están en busca de un empleo. Cabe mencionar que, de acuerdo con las estimaciones del Modelo de Regresión Poisson, la condición de actividad económica es altamente significativa, sin embargo, esta tiende a acercarse a uno, por lo que se podría plantear la hipótesis de que esta variable ya no será relevante para determinar en el número de hijos, debido a las condiciones económicas y sociales del país, así como en el incremento de la participación de la mujer en el campo laboral.

Otro factor importante dentro de este estudio es la condición de habla de lengua indígena y el lugar de residencia habitual de las mujeres. Los resultados muestran que el promedio de hijos en las mujeres que hablan alguna lengua indígena o que viven en una localidad rural es más alto, esto debido a que esta población durante muchos años ha sido considerada como vulnerable en cuanto al acceso a los servicios de salud, esto a causa de la falta de personal capacitado para atender a la población hablante de lengua indígena, así como la disponibilidad de los establecimientos de salud en las localidades rurales (Valdivieso, 2012; Szasz y Lerner, 2010; Tuñon, 2000).

Finalmente, este trabajo permitió demostrar que las mujeres que pertenecen a un estrato sociodemográfico bajo presentan una tasa de fecundidad más alta en comparación con las mujeres que pertenecen a un estrato alto, confirmando que los grupos más pobres presentan las tasas de fecundidad más altas y que además se clasifican como un grupo vulnerable en materia de salud sexual y reproductiva (Di Cesare y Rodríguez, 2006).

Es de suma importancia mencionar que en esta tesina se analizaron algunos factores que pueden influir en los niveles de fecundidad en el país, no obstante, la fecundidad como parte de la dinámica demográfica es considerada como un fenómeno multicausal en donde pueden intervenir factores que en este trabajo no se hayan mencionado, por lo que se sugiere estudiar otras fuentes de información, así como otras metodologías que puedan brindar los elementos suficientes para explicar dicho fenómeno.

# Apéndice



# Apéndice A

## Modelos para variables de conteo en R

En este primer apartado se explicará como obtener las estimaciones de los principales modelos para variables de conteo en R (Poisson, Binomial Negativa y *Zero Inflated Poisson*):

### Modelo de Regresión Poisson

```
Mod<-glm(y~x1+x2+...+xn, family = poisson, weights = NULL)
```

**Modelo de Regresión Poisson (bajo un esquema de muestreo).** En este modelo se necesita la información de la unidad primaria de muestreo, el estrato del diseño y los pesos muestrales.

```
library(survey)
Dis<-svydesign(id=~UPM_DIS, strata=~EST_DIS, weights=~W, data = datos, nest = TRUE)
options(survey.lonely.psu="adjust")
Mod<-svyglm(y~x1+x2+...+xn, design=Dis, family = poisson)
```

### Modelo de Regresión Binomial Negativa (cuando se conoce el valor de theta)

```
library(MASS)
Mod<-glm(y~x1+x2+...+xn, family = negative.binomial(theta = k), weights = NULL)
```

### Modelo de Regresión Binomial Negativa (cuando se desconoce el valor de theta)

```
library(MASS)
Mod<-glm.nb(y~x1+x2+...+xn, weights = NULL)
```

**Modelo de Regresión Binomial Negativa (bajo un esquema de muestreo).** En este modelo se necesita la información de la unidad primaria de muestreo, el estrato del diseño y los pesos muestrales.

```
library(survey)
library(MASS)
Dis<-svydesign(id=~UPM_DIS, strata=~EST_DIS, weights=~W, data = datos, nest = TRUE)
options(survey.lonely.psu="adjust")
Mod<-svyglm(y~x1+x2+...+xn, design=Dis, family = negative.binomial(theta = k))
```

### Modelo *Zero Inflated Poisson*

```
library(pscl)
Mod<-zeroinfl(y~x1+x2+...+xn | z1+z2+...+zm, dist = "poisson", link = "logit")
```



**Modelo Zero Inflated Poisson (bajo un esquema de muestreo).** En este modelo se necesita la información de la unidad primaria de muestreo, el estrato del diseño y los pesos muestrales. El método puede encontrarse en la siguiente página: [https://rpubs.com/corey\\_sparks/59698](https://rpubs.com/corey_sparks/59698)

```
library(pscl)
Mod<-zeroinfl(y~x1+x2+...+xn | z1+z2+...+zm, dist = "poisson", link = "logit",
weights=W/mean(W))

#First, I define a function to get the clustered, or robust standard errors.
#This function effectively controls for the within-strata homogeneity when
#calculating the se's for the betas.
#I stole this from: http://drewdimery.com/robust-ses-in-r/
#and http://people.su.se/~ma/clustering.pdf
#I also added a correction to use this with the hurdle and zero-inflated models
#This is how stata gets robust se's

clx2 <-
function(fm, dfcw, cluster){
  # R-codes (www.r-project.org) for computing
  # clustered-standard errors. Mahmood Arai, Jan 26, 2008.
  # The arguments of the function are:
  # fitted model, cluster1
  # You need to install libraries 'sandwich' and 'lmtest'
  # reweighting the var-cov matrix for the within model
  require(sandwich);require(lmtest)
  if(class(fm)=="zeroinfl"|class(fm)=="hurdle") {
    M <- length(unique(cluster))
    N <- length(cluster)
    K <- dim(fm$vcov)[1] #here is the rank from the zero inflated fits
    dfc <- (M/(M-1))*((N-1)/(N-K))
    uj <- apply(estfun(fm),2, function(x) tapply(x, cluster, sum));
    #fix a length problem in dfc
    vcovCL <- dfc[1]*sandwich(fm, meat=crossprod(uj)/N)*dfcw
    list(summary=coefest(fm, vcovCL))}
  else if(class(fm)!="zeroinfl"){
    M <- length(unique(cluster))
    N <- length(cluster)
    K <- fm$rank
    dfc <- (M/(M - 1)) * ((N - 1)/(N - K))
    uj <- apply(estfun(fm), 2, function(x) tapply(x, cluster, sum));
    rcse.cov <- dfc * sandwich(fm, meat = crossprod(uj)/N)
    rcse.se <- coefest(fm, rcse.cov)
    return(list( rcse.se))}

}

clx2(Mod,1,UPM_DIS)
```

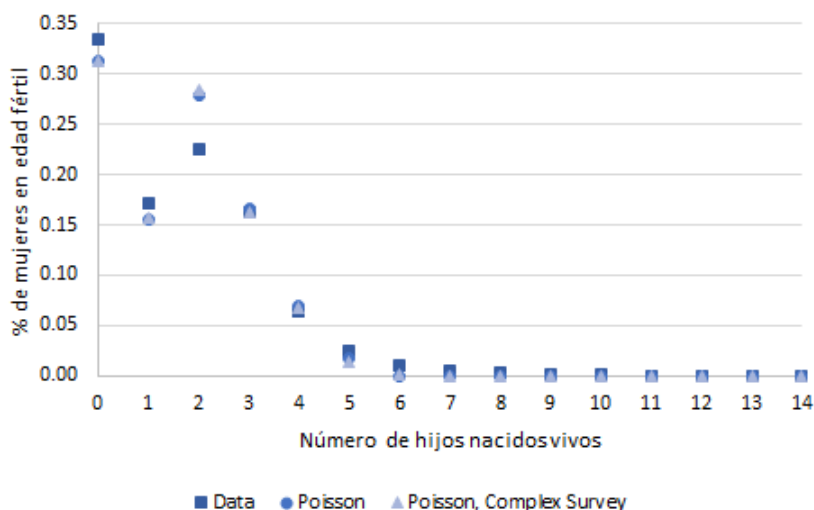
A continuación se muestra la aplicación de los modelos con los datos de la ENADID 2014 expuestos en este trabajo de investigación:

Cuadro A.1: Modelo de Regresión Poisson.

Variable	Sin efecto del diseño de la muestra					Con efecto del diseño de la muestra				
	Estimate	Std. Error	z value	P-value		Estimate	Std. Error	t value	P-value	
Intercepto	-2.62	0.02	-162.82	0.00	***	-2.63	0.02	-119.46	0.00	***
Edad	0.04	0.00	117.18	0.00	***	0.04	0.00	102.37	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	0.28	0.01	42.39	0.00	***	0.28	0.01	39.83	0.00	***
Primaria o preescolar	0.41	0.01	54.84	0.00	***	0.42	0.01	50.97	0.00	***
Sin instrucción	0.49	0.01	32.97	0.00	***	0.51	0.02	26.53	0.00	***
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	1.78	0.01	136.74	0.00	***	1.78	0.02	77.54	0.00	***
Condición de actividad										
Económicamente activa (referencia)										
No económicamente activa	0.03	0.01	4.86	0.00	***	0.03	0.01	5.31	0.00	***
Condición de habla de lengua indígena										
No habla lengua indígena (referencia)										
Habla lengua indígena	0.08	0.01	7.74	0.00	***	0.09	0.01	6.15	0.00	***
Lugar de residencia habitual										
Urbano (referencia)										
Complemento urbano	0.03	0.01	4.84	0.00	***	0.04	0.01	5.40	0.00	***
Rural	0.10	0.01	14.18	0.00	***	0.11	0.01	12.47	0.00	***
Estrato sociodemográfico										
Alto y medio alto (referencia)										
Bajo y medio bajo	0.07	0.01	9.18	0.00	***	0.08	0.01	9.99	0.00	***
Null deviance	159,463 con 89,903 grados de libertad					159,630 con 89,903 grados de libertad				
Residual deviance	65,803 con 89,893 grados de libertad					65,893 con 89,893 grados de libertad				
R <sup>2</sup> Nagelkerke	0.665576					0.666387				
AIC	228,946					226,534				

Fuente: Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Figura A.1: Ajuste del Modelo de Regresión Poisson.



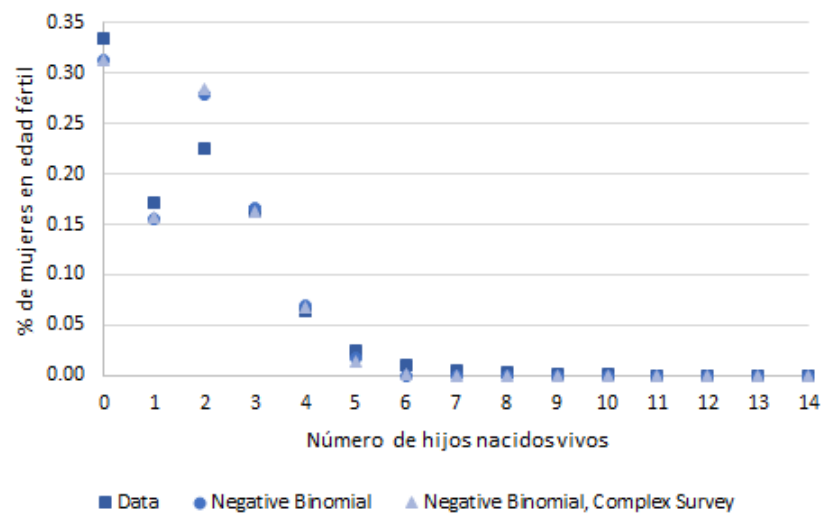
Fuente: Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Cuadro A.2: Modelo de Regresión Binomial Negativa.

Variable	Sin efecto del diseño de la muestra					Con efecto del diseño de la muestra				
	Estimate	Std. Error	z value	P-value		Estimate	Std. Error	t value	P-value	
Intercepto	-2.62	0.02	-162.82	0.00	***	-2.63	0.02	-119.46	0.00	***
Edad	0.04	0.00	117.17	0.00	***	0.04	0.00	102.37	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	0.28	0.01	42.39	0.00	***	0.28	0.01	39.83	0.00	***
Primaria o preescolar	0.41	0.01	54.84	0.00	***	0.42	0.01	50.97	0.00	***
Sin instrucción	0.49	0.01	32.97	0.00	***	0.51	0.02	26.53	0.00	***
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	1.78	0.01	136.74	0.00	***	1.78	0.02	77.54	0.00	***
Condición de actividad										
Económicamente activa (referencia)										
No económicamente activa	0.03	0.01	4.86	0.00	***	0.03	0.01	5.31	0.00	***
Condición de habla de lengua indígena										
No habla lengua indígena (referencia)										
Habla lengua indígena	0.08	0.01	7.74	0.00	***	0.09	0.01	6.14	0.00	***
Lugar de residencia habitual										
Urbano (referencia)										
Complemento urbano	0.03	0.01	4.84	0.00	***	0.04	0.01	5.40	0.00	***
Rural	0.10	0.01	14.18	0.00	***	0.11	0.01	12.47	0.00	***
Estrato sociodemográfico										
Alto y medio alto (referencia)										
Bajo y medio bajo	0.07	0.01	9.18	0.00	***	0.08	0.01	9.99	0.00	***
Null deviance	159,446 con 89,903 grados de libertad					159,613 con 89,903 grados de libertad				
Residual deviance	65,795 con 89,893 grados de libertad					65,885 con 89,893 grados de libertad				
R <sup>2</sup> Nagelkerke	0.620347					0.619673				
AIC	228,951					226,537				
Theta	13,741					13,741				

**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Figura A.2: Ajuste del Modelo de Regresión Binomial Negativa.



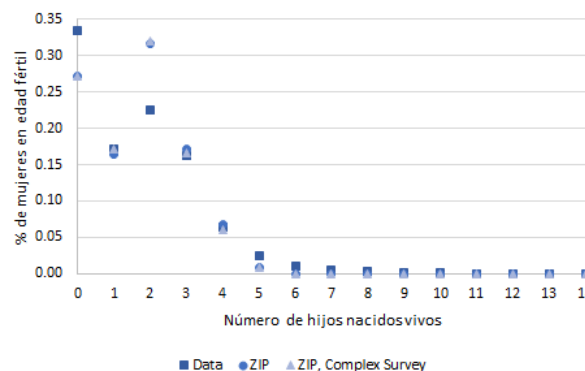
**Fuente:** Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Cuadro A.3: Modelo Zero Inflated Poisson.

Variable	Sin efecto del diseño de la muestra				Con efecto del diseño de la muestra					
	Estimate	Std. Error	z value	P-value	Estimate	Std. Error	t value	P-value		
<b>Count model</b>										
Intercepto	-1.66	0.03	-51.88	0.00	***	-1.69	0.06	-28.61	0.00	***
Edad	0.03	0.00	101.58	0.00	***	0.03	0.00	88.35	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	0.26	0.01	38.56	0.00	***	0.26	0.01	37.02	0.00	***
Primaria o preescolar	0.39	0.01	52.15	0.00	***	0.40	0.01	49.32	0.00	***
Sin instrucción	0.49	0.02	32.45	0.00	***	0.50	0.02	26.14	0.00	***
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	1.00	0.03	35.08	0.00	***	1.03	0.06	18.60	0.00	***
Condición de actividad										
Económicamente activa (referencia)										
No económicamente activa	0.05	0.01	9.88	0.00	***	0.06	0.01	10.24	0.00	***
Condición de habla de lengua indígena										
No habla lengua indígena (referencia)										
Habla lengua indígena	0.08	0.01	7.83	0.00	***	0.09	0.01	6.17	0.00	***
Lugar de residencia habitual										
Urbano (referencia)										
Complemento urbano	0.03	0.01	4.65	0.00	***	0.04	0.01	4.90	0.00	***
Rural	0.10	0.01	13.64	0.00	***	0.11	0.01	11.24	0.00	***
Estrato sociodemográfico										
Alto y medio alto (referencia)										
Bajo y medio bajo	0.06	0.01	8.54	0.00	***	0.07	0.01	9.31	0.00	***
<b>Zero-inflation model</b>										
Intercepto	6.72	0.38	17.64	0.00	***	6.88	0.78	8.86	0.00	***
Edad	-0.24	0.02	-13.20	0.00	***	-0.24	0.04	-6.50	0.00	***
Nivel de escolaridad										
Media superior y superior (referencia)										
Secundaria	-1.04	0.09	-11.82	0.00	***	-1.07	0.13	-7.91	0.00	***
Primaria o preescolar	-0.84	0.15	-5.79	0.00	***	-0.82	0.20	-4.04	0.00	***
Sin instrucción	1.01	0.35	2.92	0.00	**	0.47	0.71	0.66	0.51	
Situación conyugal										
Soltera (referencia)										
Unida y alguna vez unida	-18.92	270.43	-0.07	0.94		-18.60	2.29	-8.14	0.00	***
$R^2$ Nagelkerke			0.61803				0.618885			
AIC			223,847				221,464			

Fuente: Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.

Figura A.3: Ajuste del Modelo Zero Inflated Poisson



Fuente: Estimaciones con base en la Encuesta Nacional de la Dinámica Demográfica, 2014.



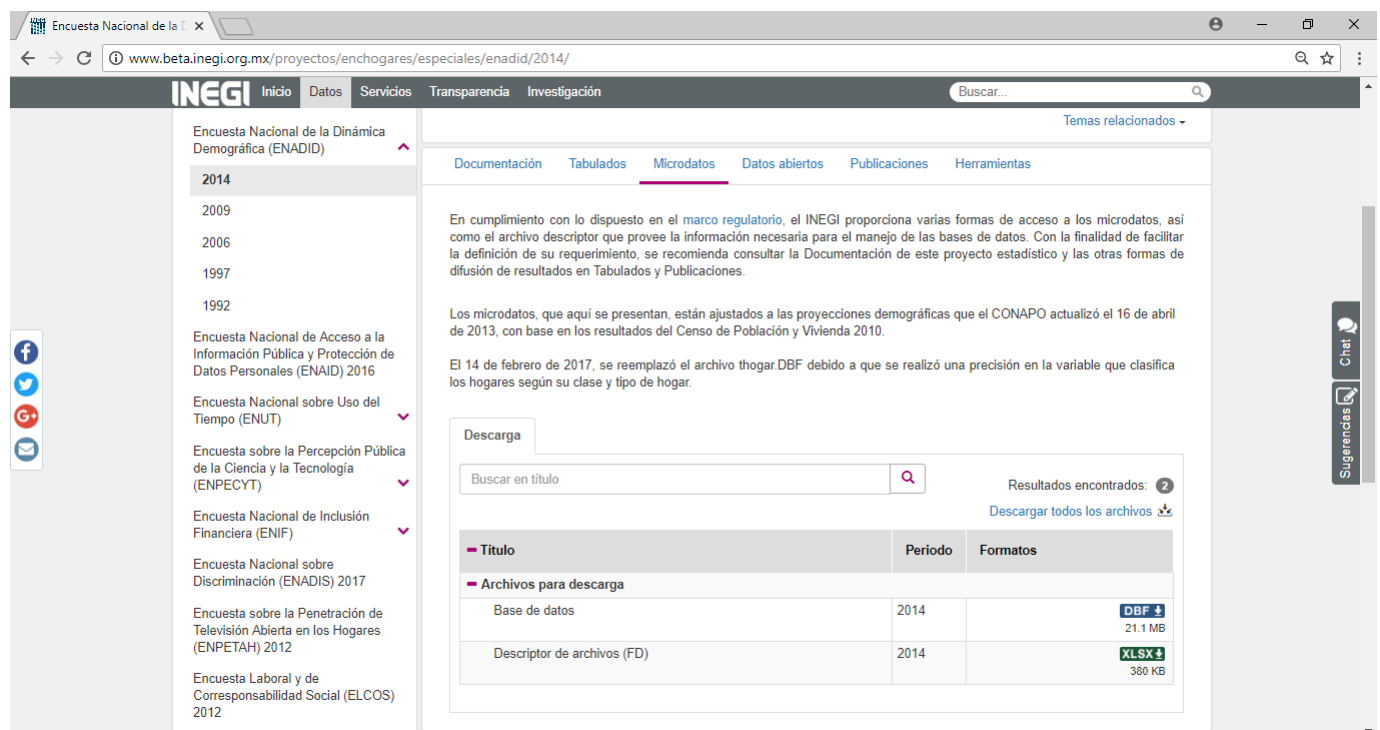
# Apéndice B

## Modelación en R

En este segundo apéndice se explicarán cada uno de los pasos para la obtención de los resultados de este trabajo de investigación. La aplicación del modelo de regresión Poisson así como el análisis de la base de datos de la Encuesta Nacional de la Dinámica Demográfica (ENADID) 2014 se realizó en R. Toda la información de la ENADID se encuentra en la siguiente página de internet:

<http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/enadid/2014/>

En esta página se puede encontrar los documentos técnicos de la encuesta (síntesis metodológica, diseño muestral, informe operativo, marco de referencia y los manuales para la entrevistadora), asimismo se incluyen los cuestionarios utilizados para recabar la información del hogar así como el de la mujer seleccionada para la encuesta. En el apartado de Microdatos se puede acceder a las bases de datos y al descriptor de archivos de la ENADID 2014.



The screenshot shows the INEGI website interface for the ENADID 2014 project. The main navigation bar includes 'Inicio', 'Datos', 'Servicios', 'Transparencia', and 'Investigación'. A search bar is located in the top right. The left sidebar lists various surveys, with 'Encuesta Nacional de la Dinámica Demográfica (ENADID)' selected. The main content area is titled 'Microdatos' and contains a 'Descarga' section. This section includes a search bar and a table of available data files for download.

Título	Periodo	Formatos
<b>Archivos para descarga</b>		
Base de datos	2014	DBF + 21.1 MB
Descriptor de archivos (FD)	2014	XLSX + 380 KB

A continuación se presenta el código en R que fue utilizado para generar los resultados del presente trabajo de investigación. Está escrito para que cualquier usuario pueda ejecutarlo sin la necesidad de cambiar alguna instrucción, además se incluyen comentarios a lo largo del código para facilitar su lectura e interpretación.

```
library(foreign)

#### Extracción de información de la ENADID 2014 ####
##Características de la mujer
Base<-read.dbf("C:/TMMujer1.dbf")
##Tamaño de muestra
Base$MUESTRA<-1
##Selección de casos
Base<-Base[,c("UPM_DIS", "EST_DIS", "P5_2", "P5_8", "DOMINIO", "P3_12", "NIV", "ESTRATO",
"COND_ACT", "P10_1_AG", "MUESTRA", "FAC_PER")]

##Etiquetas de variables
colnames(Base)[colnames(Base)=="P5_2"]<-"EDAD"
colnames(Base)[colnames(Base)=="P5_8"]<-"HIJ_NAC_VIV"
colnames(Base)[colnames(Base)=="P3_12"]<-"CHLI"
colnames(Base)[colnames(Base)=="P10_1_AG"]<-"SIT_CONY"

write.csv(Base, "C:/Base_ENADID_2014.csv", row.names = FALSE)
rm(list = ls())
gc()

#### Análisis de la información ####
Base<-read.csv("C:/Base_ENADID_2014.csv", header = TRUE)
Base<-Base[Base$NIV!=99,]

##las variables con datos faltantes son "total de hijos nacidos vivos"
##los vacios en el número de hijos nacidos vivos corresponden a mujeres que indicaron
##nunca haber estado embarazadas (P5_6), por lo que el valor es cero
Base$HIJ_NAC_VIV[is.na(Base$HIJ_NAC_VIV)]<-0

##Selección de mujeres de 15 a 49 años
Base<-Base[Base$EDAD<=49,]

##Recodificación de variables para el análisis
##Total de HNV (sólo para el análisis descriptivo)
##(0, 1, 2, 3, 4 o más)
Base$HNV_AG<-Base$HIJ_NAC_VIV
Base$HNV_AG[Base$HIJ_NAC_VIV>=4]<-4
##Grupos de edad (sólo para el análisis descriptivo)
##(1-De 15 a 19, 2-De 20 a 30, 3-De 30 a 39, 4-De 40 a 49)
Base$EDADG<-4
Base$EDADG[Base$EDAD<=39]<-3
```

```

Base$EDADG[Base$EDAD<=29]<-2
Base$EDADG[Base$EDAD<=19]<-1
##Escaridad (0-Media superior o más, 1-Secundaria, 2-Preescolar o primaria,
##3-Sin estudios)
Base$ESC<-0
Base$ESC[Base$NIV==3]<-1
Base$ESC[Base$NIV==1 | Base$NIV==2]<-2
Base$ESC[Base$NIV==0]<-3
Base$NIV<-NULL
##Situación conyugal (0-Solteras, 1-Unidas y exunidas)
Base$SIT_CONY_1<-1
Base$SIT_CONY_1[Base$SIT_CONY==3]<-0
Base$SIT_CONY<-Base$SIT_CONY_1
Base$SIT_CONY_1<-NULL
##Condición de actividad (0-PEA, 1-PNEA)
Base$COND_ACT<-Base$COND_ACT-1
##Condición de habla de lengua indígena (0-No, 1-Si)
Base$CHLI_1<-0
Base$CHLI_1[Base$CHLI==1]<-1
Base$CHLI<-Base$CHLI_1
Base$CHLI_1<-NULL
##Lugar de residencia (0-Urbano alto, 1-Complemento urbano, 2-Rural)
Base$DOMINIO<-as.character(Base$DOMINIO)
Base$DOMINIO_1<-0
Base$DOMINIO_1[Base$DOMINIO=="C"]<-1
Base$DOMINIO_1[Base$DOMINIO=="R"]<-2
Base$DOMINIO<-Base$DOMINIO_1
Base$DOMINIO_1<-NULL
##Tipo de estrato socioeconómico (0-Alto y medio alto, 1-Bajo y medio bajo)
Base$ESTRATO_1<-0
Base$ESTRATO_1[Base$ESTRATO==1 | Base$ESTRATO==2]<-1
Base$ESTRATO<-Base$ESTRATO_1
Base$ESTRATO_1<-NULL

##variables categóricas
Base$HNV_AG<-as.factor(Base$HNV_AG)
Base$EDADG<-as.factor(Base$EDADG)
Base$ESC<-as.factor(Base$ESC)
Base$SIT_CONY<-as.factor(Base$SIT_CONY)
Base$COND_ACT<-as.factor(Base$COND_ACT)
Base$CHLI<-as.factor(Base$CHLI)
Base$DOMINIO<-as.factor(Base$DOMINIO)
Base$ESTRATO<-as.factor(Base$ESTRATO)

##Se define el esquema de diseño de la muestra de la ENADID 2014
library(survey)
Diseño<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base, nest=TRUE)
options(survey.lonely.psu="adjust")

```



```

##-----Estadísticas descriptivas
##----Número total de mujeres en edad fértil según número de hijos
B1<-svyby(~MUESTRA, by=~HIJ_NAC_VIV, design = Diseño, svytotal)
##Coeficiente de variación
B1$CV<-cv(B1)*100
##Intervalos de confianza (90%)
B1$INT_INF<-confint(B1, level = 0.9)[,1]
B1$INT_SUP<-confint(B1, level = 0.9)[,2]
colnames(B1)[colnames(B1)=="MUESTRA"]<-"FAC_PER"
colnames(B1)[colnames(B1)=="se"]<-"SD"
##Tamaño de muestra
B2<-aggregate(MUESTRA~HIJ_NAC_VIV, data = Base, FUN = sum)
B1<-merge(B2,B1, by=c("HIJ_NAC_VIV"), all.x = TRUE)
B1

##----Porcentaje de mujeres en edad fértil según número de hijos y edad
##al momento de la encuesta
x<-sort(unique(Base$HNV_AG))
Base_final<-NULL
for(i in 2:5){
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~EDADG, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]
  ##Total de HNV
  B1$HNV<-x[i]
  Base_final<-rbind(Base_final,B1)
  rm(Base_1,Dis,B1)
}

##----Porcentaje de mujeres en edad fértil según número de hijos y escolaridad
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)

```

```

options(survey.lonely.psu="adjust")
##Estadísticas
B1<-svyby(~HNV, denominator=~MUESTRA, by=~ESC, design = Dis, svyratio)
B1$INT_INF<-confint(B1, level = 0.9)[,1]
B1$INT_SUP<-confint(B1, level = 0.9)[,2]
##Total de HNV
B1$HNV<-x[i]
Base_final<-rbind(Base_final,B1)
rm(Base_1,Dis,B1)
}

##----Porcentaje de mujeres en edad fértil según número de hijos y situación conyugal
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~SIT_CONY, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]
  ##Total de HNV
  B1$HNV<-x[i]
  Base_final<-rbind(Base_final,B1)
  rm(Base_1,Dis,B1)
}

##----Porcentaje de mujeres en edad fértil según número de hijos y condición de
##actividad
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~COND_ACT, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]

```

```
##Total de HNV
B1$HNV<-x[i]
Base_final<-rbind(Base_final,B1)
rm(Base_1,Dis,B1)
}

##----Porcentaje de mujeres en edad fértil según número de hijos y lengua indígena
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~CHLI, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]
  ##Total de HNV
  B1$HNV<-x[i]
  Base_final<-rbind(Base_final,B1)
  rm(Base_1,Dis,B1)
}

##----Porcentaje de mujeres en edad fértil según número de hijos y lugar de residencia
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~DOMINIO, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]
  ##Total de HNV
  B1$HNV<-x[i]
  Base_final<-rbind(Base_final,B1)
  rm(Base_1,Dis,B1)
}
```

```

##----Porcentaje de mujeres en edad fértil según número de hijos y estrato
##sociodemográfico
Base_final<-NULL
for(i in 1:5){
  print(x[i])
  Base_1<-Base
  ##Selección del total de HNV
  Base_1$HNV<-0
  Base_1$HNV[Base$HNV_AG==x[i]]<-1
  ##Diseño
  Dis<-svydesign(id=~UPM_DIS,strata=~EST_DIS, weights=~FAC_PER, data=Base_1, nest=TRUE)
  options(survey.lonely.psu="adjust")
  ##Estadísticas
  B1<-svyby(~HNV, denominator=~MUESTRA, by=~ESTRATO, design = Dis, svyratio)
  B1$INT_INF<-confint(B1, level = 0.9)[,1]
  B1$INT_SUP<-confint(B1, level = 0.9)[,2]
  ##Total de HNV
  B1$HNV<-x[i]
  Base_final<-rbind(Base_final,B1)
  rm(Base_1,Dis,B1)
}
rm(Base_final,i,x)

##-----Selección del modelo
##Inferencia sobre el número de hijos nacidos vivos
##Media
Datos<-as.data.frame(svymean(~HIJ_NAC_VIV, design=Diseño))
Datos$LIM_INF<-confint(svymean(~HIJ_NAC_VIV, design=Diseño), level = 0.9)[,1]
Datos$LIM_SUP<-confint(svymean(~HIJ_NAC_VIV, design=Diseño), level = 0.9)[,2]
##Varianza
Datos<-as.data.frame(svyvar(~HIJ_NAC_VIV, design=Diseño))
Datos$LIM_INF<-confint(svyvar(~HIJ_NAC_VIV, design=Diseño), level = 0.9)[,1]
Datos$LIM_SUP<-confint(svyvar(~HIJ_NAC_VIV, design=Diseño), level = 0.9)[,2]

library(rcompanion)
library(MASS)
library(pscl)

##----Modelo de regresión Poisson
Mod1<-glm(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+DOMINIO+ESTRATO, data=Base,
family=poisson)
summary(Mod1)
nagelkerke(Mod1)

##----Modelo de regresión binomial negativa
Mod2<-glm.nb(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+DOMINIO+ESTRATO, data=Base)
summary(Mod2)
nagelkerke(Mod2)

```

```
#Prueba de hipótesis sobre (1/Theta)
odTest(Mod2)

##-----Zero Inflated Poisson
Mod3<-zeroinfl(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+DOMINIO+ESTRATO |
EDAD+ESC+SIT_CONY, data=Base, dist = "poisson", link = "logit")
summary(Mod3)
nagelkerke(Mod3)
AIC(Mod3)
##Prueba de hipótesis sobre el modelo
vuong(Mod1, Mod3)

##Análisis gráfico de los modelos
#1) Ajuste del modelo Poisson
Base$ESTIM<-round(fitted(Mod1),0)
B1<-aggregate(FAC_PER~HIJ_NAC_VIV, data = Base, FUN = sum)
B1$PORC_1<-B1$FAC_PER/sum(B1$FAC_PER)
B1$FAC_PER<-NULL
B2<-aggregate(FAC_PER~ESTIM, data = Base, FUN = sum)
B2$PORC_2<-B2$FAC_PER/sum(B2$FAC_PER)
B2$FAC_PER<-NULL
colnames(B2)[colnames(B2)=="ESTIM"]<-"HIJ_NAC_VIV"
B1<-merge(B1,B2, by=c("HIJ_NAC_VIV"), all.x=TRUE)
rm(B2)
B1[is.na(B1)]<-0
plot(B1$HIJ_NAC_VIV, B1$PORC_1, col = "blue")
points(B1$HIJ_NAC_VIV, B1$PORC_2, col = "red", pch = "x")

#2) Ajuste del modelo Zero Inflated Poisson
Base$ESTIM<-round(fitted(Mod3),0)
B1<-aggregate(FAC_PER~HIJ_NAC_VIV, data = Base, FUN = sum)
B1$PORC_1<-B1$FAC_PER/sum(B1$FAC_PER)
B1$FAC_PER<-NULL
B2<-aggregate(FAC_PER~ESTIM, data = Base, FUN = sum)
B2$PORC_2<-B2$FAC_PER/sum(B2$FAC_PER)
B2$FAC_PER<-NULL
colnames(B2)[colnames(B2)=="ESTIM"]<-"HIJ_NAC_VIV"
B1<-merge(B1,B2, by=c("HIJ_NAC_VIV"), all.x=TRUE)
rm(B2)
B1[is.na(B1)]<-0
plot(B1$HIJ_NAC_VIV, B1$PORC_1, col = "blue")
points(B1$HIJ_NAC_VIV, B1$PORC_2, col = "red", pch = "x")

##-----Modelo de regresión Poisson (bajo el esquema de muestreo)
Mod4<-svyglm(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+DOMINIO+ESTRATO,
design=Diseño, family=poisson)
summary(Mod4)
##-devianza
```

```
Mod4$null.deviance
Mod4$df.null
Mod4$deviance
Mod4$df.residual
##-pseudo r2
nagelkerke(glm(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+DOMINIO+ESTRATO,
data = Base, family=poisson, weights = FAC_PER/mean(FAC_PER)))
##-AIC
Mod4$aic

##-----Análisis de residuos (modelo con diseño complejo)
library(surveillance)
Base$RESIDUOS<-anscombe.residuals(glm(HIJ_NAC_VIV~EDAD+ESC+SIT_CONY+COND_ACT+CHLI+
DOMINIO+ESTRATO, data = Base, family=poisson, weights = FAC_PER/mean(FAC_PER)))
summary(Base$RESIDUOS)
mean(Base$RESIDUOS)
sd(Base$RESIDUOS)
#Gráfica de dispersión
plot(Base$RESIDUOS, col = "lightblue", ylab = "Residuos", xlab = "Observaciones",
pch = 20)
lines(seq(1,dim(Base)[1]), rep(-2,dim(Base)[1]), col = "blue", lwd = 3)
lines(seq(1,dim(Base)[1]), rep(2,dim(Base)[1]), col = "blue", lwd = 3)
#Histograma
hist(Base$RESIDUOS, freq= FALSE, col = "lightblue", border = "blue", xlab =
"Residuos", ylab = "Densidad", main = "")
x<-rnorm(dim(Base)[1], mean = mean(Base$RESIDUOS), sd = sd(Base$RESIDUOS))
points(x,dnorm(x, mean = mean(x), sd = sd(x)), col = "blue", pch = 20)
#QQ plot
qqplot(x, Base$RESIDUOS, col = "lightblue", pch = 20, main = "", xlab = "Theoretical
Quantiles", ylab = "Sample Quantiles")
lines(x,x,col = "blue", lwd = 3)
##Función de distribución empírica
ECDF<-ecdf(Base$RESIDUOS)
plot(ECDF, col="lightblue", lwd=3, xlab="Residuos", ylab="Densidad", main="")
points(x, pnorm(x, mean = mean(x), sd = sd(x)), col = "blue", pch = 20)
```



# Bibliografía

- [1] Alfonso, Juan Carlos (2004), “*Cuba: de la primera a la segunda transición demográfica. El descenso de la fecundidad*”, en *La fecundidad en América Latina y el Caribe: ¿transición o revolución?*, CEPAL, Serie Seminarios y Conferencias núm. 36, Santiago de Chile.
- [2] Becker, Gary S. (1960). “*An Economic Analysis of Fertility*”. *Demographic and Economic Change in Developing Countries*. Princeton University Press.
- [3] Belmonte, Gregorio (2007). “*Construcción de masculinidades y su incidencia en el embarazo adolescente del Valle del Yaqui 2000-2004*”. Tesis de Doctorado. Instituto de Investigaciones Histórico Sociales. Universidad Veracruzana.
- [4] Cameron, A.C. y Trivedi, P.K. (1998). “*Regression Analysis of Count Data*”. *Econometric Society Monographs*. Cambridge University Press.
- [5] Chackiel, Juan (2004). “*La transición de la fecundidad en América Latina 1950-2000*”. *Papeles de población*, 10(41), pág. 9-58.
- [6] Curtis, Sian L. y Blanc, Ann K. (1997). “*Determinants of contraceptive failure, switching and discontinuation: An analysis of dhs contraceptive histories*”, en *dhs Analytical Reports*, num. 6, Calverton, Maryland.
- [7] Di Cesare, Mariachiara; Rodríguez, Jorge (2006). “*Análisis micro de los determinantes de la fecundidad adolescente: en el caso de Brasil y Colombia*”. *Papeles de Población*, Año 12, No. 18, abril-junio.
- [8] Di Cesare, Mariachiara (2007). “*Patrones emergentes en la fecundidad y la salud reproductiva y sus vínculos con la pobreza en América Latina y el Caribe. Serie población y desarrollo*”. Centro Latinoamericano y Caribeño de la Demografía (CELADE) – División de Población del CEPAL.
- [9] Hausman, Jerry; Bronwyn, Hall y Zvi, Griliches (1984). “*Econometric Models for Count Data with an Application to the Patents*”, *Econometrica*, Vol. 52. No.4, Jul. 1984, pp. 909-938.
- [10] INEGI [Instituto Nacional de Estadística y Geografía] (2017). Boletín “*Estadísticas a propósito del día de la madre (10 de mayo)*”. Aguascalientes, AGS. Disponible en: [http://www.inegi.org.mx/saladeprensa/aproposito/2017/madre2017\\_Nal.pdf](http://www.inegi.org.mx/saladeprensa/aproposito/2017/madre2017_Nal.pdf)
- [11] Llanes, Nathaly (2010). “*La maternidad adolescente y su efecto en la salida de la escuela entre mujeres mexicanas: replanteamientos y consideraciones*”. Facultad Latinoamericana de Ciencias Sociales.
- [12] Llovet, Juan José (1989). “*Relación entre fecundidad e ingresos: revisión de algunas interpretaciones*”. *Estudios Demográficos y Urbanos*, Vol. 4 No. 1, enero-abril.
- [13] Meneses, Eloina; Ailines, Leonardo; Hernández, María Felipa (2017). “*Factores que influyen en el conocimiento funcional de métodos anticonceptivos*”. CONAPO. *La Situación demográfica de México 2017*, México.



- [14] Montero, R; Jiménez, JD; Martín, JJ (2005). “*Estimación de un índice de necesidad sanitaria regional en España mediante modelos de regresión de recuento con filtro*”. Departamento de Economía Aplicada. Universidad de Granada.
- [15] Morales, Ángel Jair; Silva, Eliud; y Ordorica-Mellado, Manuel. (2016). “*Pronósticos de la fecundidad en México: una aplicación de modelos multivariados de series de tiempo*”. Papeles de población, 22(89), pág. 99-131.
- [16] OCDE [Organización para la Cooperación y el Desarrollo Económico] (2012). “*Haciendo lo mejor para las familias*”. Disponible en: [https://read.oecd-ilibrary.org/social-issues-migration-health/haciendo-lo-mejor-para-las-familias\\_9789264190474-es#page11](https://read.oecd-ilibrary.org/social-issues-migration-health/haciendo-lo-mejor-para-las-familias_9789264190474-es#page11)
- [17] OIT [Organización Internacional del Trabajo] (2009). “*Trabajo y Familia: hacia nuevas formas de conciliación con corresponsabilidad social*”. Programa de las Naciones Unidas para el Desarrollo (PNUD). Disponible en: [http://www.ilo.org/wcmsp5/groups/public/—dgreports/—gender/documents/publication/wcms\\_111376.pdf](http://www.ilo.org/wcmsp5/groups/public/—dgreports/—gender/documents/publication/wcms_111376.pdf)
- [18] ONU [Organización de las Naciones Unidas] (2001). “*Situación de la mujer en el mundo, 2000: tendencias y estadísticas*”. Publicación de las Naciones Unidas. Disponible en: [https://unstats.un.org/unsd/publication/SeriesK/seriesk\\_16s.pdf](https://unstats.un.org/unsd/publication/SeriesK/seriesk_16s.pdf)
- [19] Rodríguez, Grisell; Molina, Matilde de la C; y Quintana, Livia. (2015). “*Fecundidad. Estudio y comportamiento. América Latina y Cuba*”. Revista Novedades en Población vol. 1, núm. 21, pág. 1.
- [20] Safe Passages of Adulthood, (2001). “*Dynamic contextual analysis of young peoples sexual health. Faculty of Social Sciences*”. University of Southampton.
- [21] Schoumaker, Bruno y Dominique Tabutin (1990). “*Relations entre pauvreté et fécondité dans les pays du Sud. Connaissances, méthodologie et illustrations*”. Document de Travail, No. 2, Département des sciences de la population et du développement, Louvain-la-Neuve, Université Catholique de Luovain.
- [22] Stupp, Paul; Cáceres, José Mario (2001). “*The relationship between age at completion of schooling and age at first birth in El Salvador*”. Ponencia presentada en la XXIV Conferencia, General de Población de la IUSSP, del 17 al 23 de agosto, Salvador, Bahía.
- [23] Szasz, Ivonne y Lerner, Susana (2010). “*Los grandes problemas de México. Salud Reproductiva y Desigualdades en la Población*”, en Brígida García y Manuel Ordorica (coords.), Los grandes problemas de México, Tomo I. Población, El Colegio de México.
- [24] Tuñón, Esperanza (2000). “*Género y sexualidad adolescente. La búsqueda de un conocimiento huidizo*”. Colegio de México.
- [25] Valdiviezo, Noé (2012). “*Factores asociados al embarazo en adolescentes indígenas de la región de los valles centrales de Oaxaca, 2014*”. El Colegio de la Frontera Norte. Tijuana, B.C., México.
- [26] Welti, Carlos (2000). “*Análisis demográfico de la fecundidad adolescente en México*”. Papeles de Población, vol. 6, núm. 26, Universidad Autónoma del Estado de México, pág. 43-87.
- [27] Welti, Carlos (2012). “*Análisis de la fecundidad en México con los datos del Censo de Población y Vivienda 2010*”. Papeles de Población, vol. 18, núm. 73, Universidad Autónoma del Estado de México, pág. 45-76.
- [28] Zavala de Cosío, María Eugenia (1992). “*La transición demográfica en América Latina y en Europa*”, Notas de Población, vol. 20, núm. 56, pág. 11-32.