

# Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física  
**Maestría en Ciencia de Datos**



## Proyección de la pirámide poblacional mexicana como ajuste de una curva de distribución

---

TESIS para obtener el GRADO de  
MAESTRÍA EN CIENCIA DE DATOS

Tesis presentada por: **Daniela Flores Ruiz**

Director de Tesis: **Dr. Fernando I. Becerra López**

Tlaquepaque, Jalisco, Noviembre, 2021



# **Instituto Tecnológico y de Estudios Superiores de Occidente**

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

## **Departamento de Matemáticas y Física Formato de aprobación de la Maestría en Ciencia de Datos**

*Título de la tesis:*

### **Proyección de la pirámide poblacional mexicana como ajuste de una curva de distribución**

*Autor:*

**Daniela Flores Ruiz**

Tesis aprobada para completar los requisitos de grado para la Maestría en Ciencia de Datos.

---

Director de Tesis, **Dr. Fernando I. Becerra López**

---

Revisor de Tesis, **Dr. Riemann Ruiz Cruz**

---

Revisor de Tesis, **Dr. Irving Joel Llamosas Rosas**

---

Asesor Académico, **M. Juan Carlos Martínez Alvarado**

Tlaquepaque, Jalisco, Noviembre, 2021



## Dedicatoria

Con inmensa gratitud y amor dedico este trabajo:

A mi padre Gustavo Alberto Flores Arellano. Por creer en mi y ser una motivación constante de superación. Por sus consejos, enseñanzas, su apoyo incondicional y por brindarme todo lo necesario para permitir que me enfocara en mi desarrollo humano y profesional durante mi maestría.

A mi madre Rosina Ruiz Ruelas, Por ser ejemplo de perseverancia y superación. Por escucharme, darme su consejo y hacerme sentir tranquila y amada en momentos difíciles. Por formarme como una persona autosuficiente y por brindarme su apoyo y amor incondicional.

A mis hermanas Margarita Flores Ruiz y Rosina Flores Ruiz. Quienes se aseguraron de mantenerme motivada y no dejarme renunciar a mis objetivos y metas a pesar del estrés o cansancio. Por estar presentes en todo el proceso y ser pilar de mi crecimiento, por escucharme y brindarme ideas o consejos cuando me sentí perdida.

A la familia Gaxiola Flores. Por estar presentes en todo momento. Por adoptarme y hacerme parte de su familia. Por hacerme sentir amada y cuidada.

A mis compañeros de Maestría Cesar Sánchez y Jessica Reyes. Quienes con su apoyo incondicional durante estos tres años me motivaron y apoyaron hasta conseguir este logro en mi vida. Por haber formado el mejor equipo durante estos tres años de maestría y no dejar que me rindiera nunca. Por enseñarme, aprender conmigo y no dejarme sola nunca.

A mis amigos y seres queridos presentes y pasados. Por ser un sólido pilar emocional durante mi maestría y hacer que el camino no se sintiera tan pesado. Por ser generadores de experiencias, risas y amor. Y por permitirme tener momentos de felicidad genuina a su lado.

## Agradecimientos

A Dios. Por darme fortaleza y constancia. Por permitirme aprovechar al máximo las oportunidades y llenar mi camino de retos y aprendizajes. Por hacerme sentir amada, segura y comprendida, por escucharme y estar cerca en todo momento.

A mi Director de Tesis Fernando I. Becerra López. Quien me brindó toda su experiencia y conocimiento, por guiarme en este trabajo y mantenerme centrada en mis objetivos profesionales. Por dedicarme su tiempo cada semana durante el desarrollo de este trabajo y permitirme obtener el grado de maestra en ciencia de Datos.

A mi profesor de licenciatura Juan Antonio Tello. Por ser quien inicialmente me propuso este proyecto como trabajo final de licenciatura. Por haberme compartido su pasión por la demografía, sin la cual no hubiese sido posible el desarrollo de este trabajo.

A mis profesores de maestría. Por ser la base de mi desarrollo profesional. Por motivarme y compartir sus enseñanzas.



# Proyección de la pirámide poblacional mexicana como ajuste de una curva de distribución

Daniela Flores Ruiz

## Resumen

En esta tesis se muestra una investigación y análisis con respecto al ajuste de la pirámide poblacional mexicana a una distribución de probabilidad. Tomándose como referencia los datos de los censos y conteos poblacionales realizados por el INEGI en México durante el periodo de 1970 a 2020, se utiliza una red neuronal que toma como entrada variables económicas, ambientales, sociales e históricas para el entrenamiento y ajuste de los datos, utilizándose como salida distribuciones de probabilidad propuestas (en este caso, Gamma y Log-Normal). El modelo fue de aprendizaje supervisado y se ajustaron las variables independientes a los parámetros obtenidos después de modelar la pirámide poblacional con cada una de las distribuciones. Se obtuvo un ajuste satisfactorio que permitió comprobar que todas las variables propuestas resultaron significativas para este modelo. Además, se utiliza el modelo entrenado de la distribución Log-Normal para hacer una proyección de la silueta de la pirámide poblacional mexicana para el año 2030, el cual es comparable con los resultados obtenidos en proyecciones de población realizadas por instituciones gubernamentales como INEGI y CONAPO.

## Palabras Clave

- Pirámide Poblacional
- Red Neuronal
- Proyección
- Curva de Distribución



# Índice

	<b>Page</b>
1	Introducción . . . . . 15
2	Objetivo y Descripción de Variables . . . . . 21
2.1	Objetivo del Trabajo . . . . . 22
2.2	VARIABLES DE ÁMBITO ECONÓMICO . . . . . 22
2.2.1	Inflación . . . . . 22
2.2.2	Ingreso y Gasto Corriente Nacional sobre Total de Hogares . . . . . 23
2.3	VARIABLES DE ÁMBITO AMBIENTAL . . . . . 24
2.3.1	Temperatura Media Mensual Mundial . . . . . 24
2.3.2	Percepción de la población sobre Impacto Ambiental . . . . . 25
2.4	VARIABLES DE ÁMBITO SOCIAL . . . . . 25
2.4.1	Empoderamiento Femenino . . . . . 25
2.4.2	Liberación de la diversidad sexual . . . . . 27
2.4.3	Anticonceptivos y Aborto . . . . . 29
2.4.4	Estabilidad en pareja . . . . . 30
2.4.5	Hacinamiento en zonas urbanas . . . . . 31
2.5	VARIABLES DE ÁMBITO HISTÓRICO . . . . . 37
2.5.1	Baby Boom . . . . . 37
3	Metodología . . . . . 39
3.1	Descripción de salidas por medio del ajuste a curvas de Distribución de Probabilidad conocidas . . . . . 39
3.1.1	Distribución Gamma . . . . . 40
3.1.2	Distribución Log-Normal . . . . . 41
3.2	Método para el reconocimiento de patrones: red neuronal . . . . . 42
3.3	Cálculo del error . . . . . 44
4	Experimentación y resultados . . . . . 47
4.1	Entrenamiento y Selección de configuración para la Red Neuronal . . . . . 47
4.2	Validación de los Resultados Obtenidos . . . . . 48
4.3	Configuración y Error Minimizado . . . . . 49
4.4	Proyección del modelo . . . . . 51
4.5	Análisis de influencia de las variables de entrada . . . . . 51

5 Conclusiones.....	55
Referencias .....	57
Index.....	59

# Lista de figuras

	<b>Page</b>
1.1 Pirámides de población nacional para dos escenarios de proyección, 2020 y 2030. Obtenido del estudio titulado "Proyecciones de la población de México 2010-2050" por CONAPO. . . . .	18
1.2 Pirámides Poblacionales Mexicanas. Años: 1930, 1980 y 2020. . . . .	19
1.3 Estructuras poblacionales por edad y sexo. Escenarios 2010 y 2050. Obtenido del estudio titulado "Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales" por INEGI. . . . .	20
2.1 Relación histórica entre Ingreso y Gasto Corriente sobre Número total de Hogares en México durante 1980 - 2020	23
2.2 Proporción de la participación de las mujeres en la educación superior durante 1970 - 2020 . . . . .	27
2.3 Proporción de mujeres como parte de la población económicamente activa en relación al total de hombres trabajando durante 1970 - 2020 . . . . .	28
2.4 Proporción de mujeres como parte de la población económicamente activa en relación al total de personas censadas durante 1970 - 2020 . . . . .	28
2.5 Proporción de parejas divorciadas en relación al total de personas censadas durante 1970 - 2020 . . . . .	30
2.6 Proporción de personas solteras en relación al total de personas censadas durante 1970 - 2020 . . . . .	31
2.7 Ejemplo de hacinamiento histórico de ZMG desde 1970 hasta 2020. . . . .	34
2.8 Zonas Metropolitanas más importantes de acuerdo al censo INEGI 2020 . . . . .	35
2.9 Hacinaamiento y ajuste de curva para la Zona Metropolitana del Valle de México 1970 - 2020 . . . . .	36
2.10 Hacinaamiento y ajuste de curva para la Zona Metropolitana de Guadalajara 1970 - 2020 . . . . .	36

3.1	Pirámide Poblacional 2020 vista horizontal previa a la parametrización. . . . .	40
3.2	Pre-procesamiento de la Pirámide Poblacional 2020, Ajuste Distribución Normal. . . . .	40
3.3	Curva de función de densidad de la distribución Gamma con distintos parámetros. . . . .	41
3.4	Comparación de la curva obtenida con el ajuste gamma para los datos del censo 2020. . . . .	41
3.5	Curva de función de densidad de la distribución Log Normal con distintos parámetros. . . . .	41
3.6	Comparación de la curva obtenida con el ajuste Log Normal para los datos del censo 2020. . . . .	42
3.7	En este diagrama se muestra la estructura de los datos de entrada, el modelo de procesamiento de datos que se utilizará y las salidas. . . . .	42
3.8	Ejemplo de Red Neuronal Backpropagation . . . . .	43
3.9	Gráfica de la función de activación sigmooidal utilizada en la Red Neuronal . . . . .	43
3.10	Gráfica que muestra el comportamiento del error en un modelo con 6 neuronas ocultas y tasa de aprendizaje del 0.7. . . . .	45
3.11	Diagrama del proceso completo desde la recolección de los datos hasta la proyección de la pirámide poblacional mexicana para el año 2030. . . . .	46
4.1	Datos originales de las pirámides Poblacionales normalizadas para la comparación contra las aproximación del modelo de los años 1995, 2005 y 2015. . . . .	49
4.2	Pirámide poblacional normalizada obtenida de la Red Neuronal con aproximación Log Normal para los años de test 1995, 2005 y 2015 . . . . .	50
4.3	Pirámide poblacional obtenida de la Red Neuronal con aproximación Gamma para los años de test 1995, 2005 y 2015 . . . . .	50
4.4	Gráficas de proyección de la pirámide poblacional para 2030 utilizando la distribución Log Normal. . . . .	51
4.5	Gráficas de proyección de la pirámide poblacional para 2030 utilizando la distribución Gamma. . . . .	52

# Lista de tablas

	<b>Page</b>
2.1 Tabla para el cálculo del hacinamiento histórico de la Zona Metropolitana de Monterrey. 1970 - 2020. . . . .	33
2.2 Variables propuestas no incluidas en el modelo y su justificación. . . . .	37
2.3 Variables propuestas incluidas en el modelo y el indicador utilizado. . . . .	38
4.1 Tabla de Resultados de acuerdo a las mejores configuraciones probadas con cada una de las distribuciones. . . . .	49
4.2 Tabla de la norma de los pesos obtenidos de la red neuronal por variable utilizando la distribución Gamma. . . . .	52
4.3 Tabla de la norma de los pesos obtenidos de la red neuronal por variable utilizando la distribución Log Normal. . . . .	53



# 1 Introducción

El 25 de enero de 1983 se creó el Instituto Nacional de Estadística, Geografía e Informática (\INEGI <sup>1</sup>). Organismo público y autónomo responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica de México. Dentro de sus responsabilidades y obligaciones están las de captar y difundir información de México en cuanto a su territorio, recursos, población y economía. Se encarga también del proceso en el cual se recauda toda esta información, lo que se conoce como censos, esto con la intención de dar a conocer las características de nuestro país y ayudar a la toma de decisiones.

En palabras del Fondo de Población de las Naciones Unidas (UNFPA <sup>2</sup>), "Un *censo* implica la enumeración completa de la población de un país, territorio o zona, y debe realizarse al menos una vez cada 10 años. Genera una gran cantidad de datos, incluida la cantidad de personas, su distribución espacial, su estructura etaria y sexo, así como sus condiciones de vida y otras características socioeconómicas. Esta información es crítica para la buena gobernanza, la formulación de políticas, la planificación del desarrollo, la reducción de los riesgos, la respuesta a las crisis, los programas de bienestar social y los análisis de mercado empresarial."

Los censos son el método que ha sido utilizado desde antes de la creación del INEGI para la captación de dicha información y la descripción de la población mexicana. Se ha buscado recopilar variables que ayuden a modelar y describir de la forma más acertada posible a la población.

Las variables básicas que se incluyen en los formularios censales pretenden recolectar información sobre la Salud, Bienestar, Educación, Economía y Migración de las familias mexicanas, aunque con el paso del tiempo se han ido sumando variables a los datos recaudados durante los censos. Esto ha generado que se tenga apreciación más clara de la población mexicana y su comportamiento a través del tiempo.

Uno de los resultados de este complejo procedimiento es la *\Pirámide Poblacional* <sup>3</sup>. Consta de dos histogramas verticales, cada uno representando un género (sexo) de la población. Cada uno, a través de sus barras, representa el número de personas que vive en el país

<sup>1</sup> INEGI. Quiénes somos. [https://www.inegi.org.mx/inegi/quienes\\_somos.html](https://www.inegi.org.mx/inegi/quienes_somos.html), November 2021c

<sup>2</sup> UNFPA. Censos. <https://www.unfpa.org/es/censos>, November 2021

<sup>3</sup> Roland Pressat. *Demografía estadística*. Editorial Ariel, 1972

de cada grupo de edad quinquenal en orden ascendente, comenzando por el grupo de 0 - 4, 5 - 9 y así sucesivamente. Esta representación gráfica de la población brinda información sobre la composición de la población de un país, pero además, al hacer el análisis histórico de las pirámides poblacionales de un país específico, se aprecia el comportamiento de la población y sus afectaciones debido a hechos históricos, de salud, ideologías, entre otros.

Gracias al estudio de las diferentes pirámides poblacionales alrededor del mundo y sus registros históricos, se ha podido construir una clasificación que explica fenómenos comunes. Las pirámides se clasifican usualmente en tres tipos:

- **Pirámide progresiva o de Pagoda:** De forma más o menos triangular, típica de países subdesarrollados con altas tasas de natalidad y una esperanza de vida corta. Ejemplos actuales: Bolivia, Etiopía, Samoa.
- **Pirámide estacionaria o estancada:** Propias de países en vías de desarrollo, en los que se ha empezado a controlar la mortalidad, pero la natalidad sigue siendo superior a la de los países desarrollados. Ejemplos actuales: Brasil, Sudáfrica, Tailandia.
- **Pirámide regresiva o de bulbo:** La parte inferior de la pirámide es más estrecha que las zonas intermedias-superiores. Esto se debe a que la natalidad ha descendido y la población comienza a estar envejecida. Es típica de países desarrollados. Ejemplos actuales: Nueva Zelanda, Taiwán, Estados Unidos.

La **Tasa de Crecimiento Demográfico** es un indicador social que representa la tasa media anual de cambio del tamaño de la población durante un período determinado. Su finalidad es medir la velocidad a la que cambia el tamaño de la población. A pesar de ser un indicador relativamente simple, su implicación en el desarrollo sostenible de un país es sumamente importante ya que "un crecimiento demográfico rápido puede restar capacidad a los países para afrontar una amplia gama de problemas de importancia económica, social y ambiental, especialmente cuando se produce en un entorno de pobreza y falta de acceso a los recursos, o de pautas de producción y consumo insostenibles, o en zonas vulnerables desde un punto de vista ecológico." como lo indica la División de Desarrollo Sustentable del Departamento de Asuntos Económicos y Sociales de la Organización de las Naciones Unidas. (UNDESA)

La **Tasa de Mortalidad** es la proporción de personas que mueren por una causa concreta en un período en una población. Estas mediciones son importantes ya que, al tener certeza de las causas de su incremento o disminución, permiten la creación de políticas públicas y mejoramiento de los sistemas de salud de un país. Además,

la relación directa con la pirámide poblacional es evidente, ya que una alta tasa de mortalidad posiblemente reflejará una pirámide poblacional progresiva, y por el contrario, una tasa baja podría reflejar una pirámide poblacional regresiva.

La **Migración** es otro de los factores considerados en la mayoría de los estudios y proyecciones demográficas. Puede afectar el crecimiento y la estructura por sexo y edades de una población. Ya que los migrantes suelen ser personas jóvenes en etapas productivas, las regiones de destino rejuvenecen e indirectamente incrementan su tasa de natalidad, y por lo contrario, las regiones de origen envejecen y disminuyen su natalidad.

La proporción de hombres y mujeres también se ve afectada de forma similar, las regiones de destino y origen presentan cambios en la estructura, tamaño y dinámica de los hogares. Diversos estudios realizados en contextos de origen de los migrantes han documentado que la migración promueve la formación de hogares con jefatura femenina, así como de hogares ampliados o extensos; además, ha contribuido a la creación de formas inéditas de organización y convivencia, al igual que al surgimiento de nuevas pautas de derechos y obligaciones.

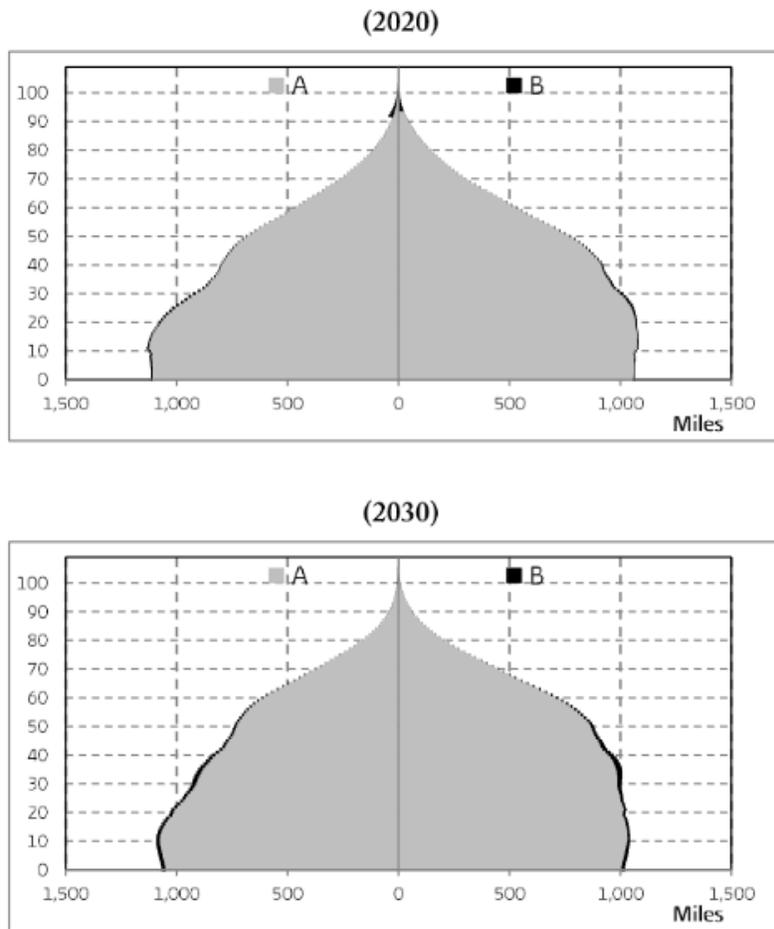
### *El caso de México*

En el año de 1900, en México vivían 13'607,259 personas. Durante 2020 se llevó a cabo el último censo poblacional<sup>4</sup>, el cuál confirmó que en México habitaban 126'014,024 personas. Este crecimiento no ha sido lineal y al analizar las pirámides poblacionales mexicanas se pueden ver cambios importantes dados por hechos históricos (como, por ejemplo, la Segunda Guerra Mundial).

En las gráficas 1.2 se aprecian tres pirámides poblacionales mexicanas a través del tiempo. La pirámide poblacional de 1930 puede ser clasificada como una pirámide progresiva, dado que su base es ancha y la parte de arriba es muy angosta. Analizando la pirámide de 1980 podemos ver representado el incremento exponencial de nacimientos generado por los tiempos de paz post segunda guerra mundial. Por último, la pirámide de 2020, en donde podemos ver que su base ya no es amplia, y se aprecia que los quinquenios intermedios aumentan su tamaño. Ésta tercera pirámide puede ser clasificada como una pirámide regresiva. Esta transición se alcanza a apreciar perfectamente al hacer la comparación de las pirámides poblacionales desde 1970 hasta 2020.

En los últimos años se han detectado algunas variables que, debido a cambios culturales, sociales y económicos, pueden ser relacionados con los cambios en la pirámide poblacional. Entre ellos están, por ejemplificar algunos, la participación de las mujeres en el ámbito laboral,

<sup>4</sup> INEGI. Comunicado de prensa núm 24/21. [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ResultCenso2020\\_Nal.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ResultCenso2020_Nal.pdf), 2021a



A - Proyección de población; B - Hipótesis alternativa.

Figure 1.1: Pirámides de población nacional para dos escenarios de proyección, 2020 y 2030. Obtenido del estudio titulado "Proyecciones de la población de México 2010-2050" por CONAPO.

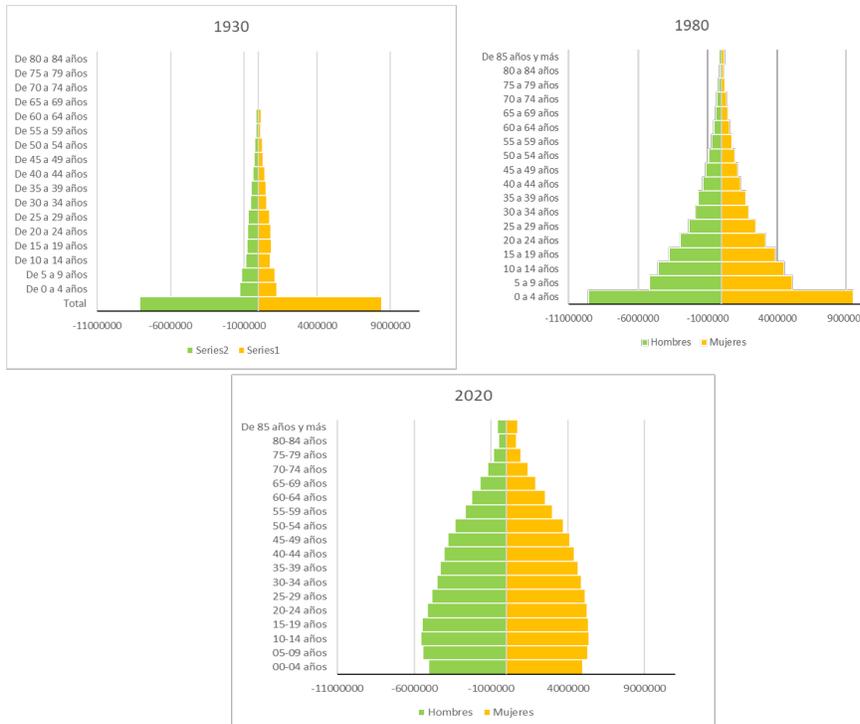


Figure 1.2: Pirámides Poblacionales Mexicanas. Años: 1930, 1980 y 2020.

la apertura a una sexualidad libre y el costo de la vida.

Con el paso de los años, México ha pasado de tener una pirámide progresiva a una estacionaria. Esto puede deberse a diferentes factores, como los previamente mencionados. Existen diversos estudios demográficos que se enfocan en hacer proyecciones de población en México, mismos que utilizan como principales indicadores de esta transición piramidal las tasas de crecimiento demográfico, mortalidad y migración.

En Diciembre del 2012, el Consejo Nacional de Población Publicó un estudio titulado "Proyecciones de la población de México 2010-2050"<sup>5</sup>, en el cuál muestra la metodología que se siguió para generar proyecciones de pirámides poblacionales hasta 2050. En este caso las variables utilizadas para dicho estudio fueron Fecundidad, Mortalidad, Migración Internacional y Migración Interna. Pueden apreciarse algunos de los resultados en la figura 1.1 en donde se muestran las proyecciones para 2020 y 2030.

Por otro lado, en 2014, el INEGI presentó un artículo llamado "Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales"<sup>6</sup>. En el cuál también se pretende proyectar pirámides poblacionales mexicanas utilizando procesos estocásticos. En este caso, se utilizaron variables muy parecidas: Mortalidad, Fecundidad, Migración Neta Internacional, Población Total. En la figura 1.3 podemos

<sup>5</sup> Mtro. Leopoldo Núñez Fernández Mtro. Javier Suárez Morales Lic. Roberto López Esquinca Lic. Antonio Amerlinck Assereto Mtra. Miroslava Godínez Trejo Mat. Vicente Medina López Mtro. Rafael López Vega Mtro. Sergio I. Velarde. Act. Carlos Fuentes Villalba, Mtro. Juan Enrique García López. *Proyecciones de la población de México 2010 - 2050. Taller Interinstitucional para la Conciliación Demográfica y las Proyecciones de Población*, 1, 2012

<sup>6</sup> Víctor Manuel García Guerrero. Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales. *18REALIDAD, DATOS Y ESPACIO, REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA*, 1, 2014

ver los resultados obtenidos por INEGI, Para el año 2050 se proyecta que la población mexicana se duplique, tomando como referencia los datos de 2010.

**Estructura por edades y sexo pronosticada al 2010 (izquierda) y al 2050 (derecha), México**

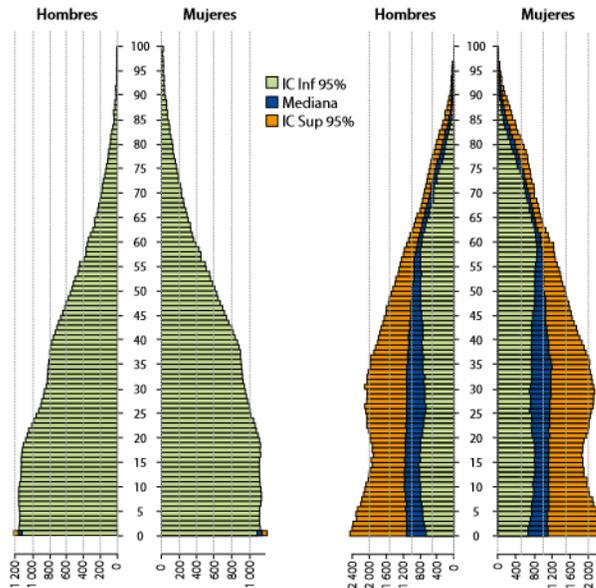


Figure 1.3: Estructuras poblacionales por edad y sexo. Escenarios 2010 y 2050. Obtenido del estudio titulado "Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales" por INEGI.

Los modelos propuestos en los últimos años han permitido, hasta ahora, hacer una correcta planeación y prevención de algunos factores de riesgo poblacional relacionados con el crecimiento acelerado de la población.

## 2 *Objetivo y Descripción de Variables*

### Contents

---

2.1	Objetivo del Trabajo . . . . .	22
2.2	Variables de ámbito económico . . . . .	22
2.2.1	Inflación . . . . .	22
2.2.2	Ingreso y Gasto Corriente Nacional sobre Total de Hogares . . . . .	23
2.3	Variables de ámbito ambiental . . . . .	24
2.3.1	Temperatura Media Mensual Mundial	24
2.3.2	Percepción de la población sobre Impacto Ambiental . . . . .	25
2.4	Variables de ámbito social . . . . .	25
2.4.1	Empoderamiento Femenino . . . . .	25
2.4.2	Liberación de la diversidad sexual . .	27
2.4.3	Anticonceptivos y Aborto . . . . .	29
2.4.4	Estabilidad en pareja . . . . .	30
2.4.5	Hacinamiento en zonas urbanas . . .	31
2.5	Variables de ámbito histórico . . . . .	37
2.5.1	Baby Boom . . . . .	37

---

A continuación se presenta el objetivo del trabajo y se describen las variables se consideraron podrían ser significativas para la explicación del fenómeno de cambio en la pirámide poblacional. Es importante mencionar que se presentan de acuerdo a una clasificación general (económico, social, ambiental e histórico). En cada una, se describirán los siguientes puntos:

- Tipo de dato.
- Fuente de consulta.
- Descripción de la variable.
- Intención en el estudio.

Una vez identificadas las variables que parecen tener una incidencia en las modificaciones a la pirámide poblacional, se pudo definir

específicamente cuáles podrían ser usadas en el modelo y cuáles no. En cada una se describen las razones por las que se tomó o rechazó del modelo propuesto y se explica el procesamiento previo que requirió cada una antes de ser incluida en el modelo.

## 2.1 *Objetivo del Trabajo*

Obtener un modelo que permita hacer una proyección de la pirámide poblacional mexicana utilizando variables económicas, sociales, ambientales e históricas, ajustando la silueta de la pirámide poblacional a una curva de distribución conocida.

Objetivos específicos del trabajo:

- Encontrar la significancia de cada una de las variables incluidas en el modelo.
- Realizar una estimación de la pirámide poblacional mexicana para el año 2030.
- Encontrar la configuración más adecuada para minimizar el error de los datos de ajuste.
- Encontrar la distribución que mejor ajuste el comportamiento de la pirámide poblacional en cada uno de los años utilizados.

## 2.2 *VARIABLES DE ÁMBITO ECONÓMICO*

### 2.2.1 *Inflación*

Se consideró inicialmente como indicador de inflación el IPC. Datos Mensuales obtenidos de INEGI<sup>1</sup> desde 1969 hasta 2020. El **\IPC** representa la medida del cambio (variación), en el precio de bienes y servicios representativos del consumo de los hogares del país conocido como canasta. Al incluir esta variable en nuestro modelo buscamos representar el encarecimiento de la vida en México como causa de la desaceleración de la natalidad. Es decir, la población en edad reproductiva está reconsiderando la posibilidad de tener familia debido a la variación en el precio de la canasta básica y la percepción de los sueldos bajos en el país.

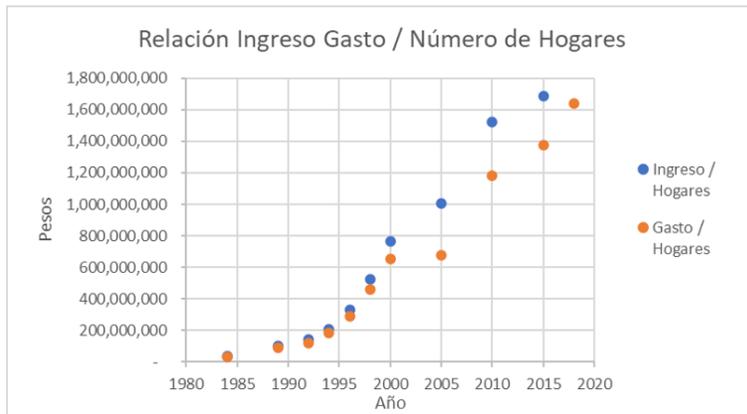
Esta variable no se tomó en cuenta dentro de nuestro modelo ya que se pensó en tomar variables que representen el impacto económico en las familias mexicanas de forma más directa.

<sup>1</sup> INEGI. Fuente de datos inegi. <https://www.inegi.org.mx/datos/>, 2021b

### 2.2.2 Ingreso y Gasto Corriente Nacional sobre Total de Hogares

Este dato se obtuvo de los censos poblacionales de INEGI<sup>2</sup>, de donde se extrajo el Ingreso y el Gasto Corriente (representado en miles de pesos) del país y se dividió cada uno por el Total de Hogares registrados durante los años de 1970 hasta 2020. El resultado es el Total del Ingreso y Gasto Corriente que corresponde a cada familia mexicana.

Se buscó representar así la economía familiar mexicana y cómo la relación Ingreso - Gasto afecta en el desarrollo de la población. Se puede ver en la gráfica 2.1 la interacción del Ingreso y el Gasto por familia y cómo ambos incrementan con el paso de los años.



<sup>2</sup> INEGI. Fuente de datos inegi. <https://www.inegi.org.mx/datos/>, 2021b

Figure 2.1: Relación histórica entre Ingreso y Gasto Corriente sobre Número total de Hogares en México durante 1980 - 2020

Ya que los datos están representados en miles de pesos, ésta variable requirió de normalización antes de ser incluida en el modelo. Además, tuvo datos faltantes en los años de 1970, 1980 y 2020, los cuales se completaron utilizando dos regresiones exponenciales.

El proceso de *normalización*<sup>3</sup> que se utilizó en este trabajo se representa en la ecuación 2.1. Representa un escalamiento simple donde el máximo valor de los datos toma el valor de 1 y el valor más bajo se toma como 0. En la ecuación  $x'$  representa el dato normalizado,  $x_i$  representa el  $i$ -ésimo dato y  $x$  representa el vector de datos completo del cual se van a obtener el valor mínimo y el valor máximo.

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2.1)$$

Para ambas regresiones se tomó el año de 1970 como año 0, 1980 como el año 10, y así sucesivamente se fueron designando los valores para el resto de los años. Esto se hizo con la intención de reducir los coeficientes resultantes de nuestras ecuaciones de regresión.

Para el Ingreso Corriente se obtuvo la ecuación 2.2, obteniendo un coeficiente de correlación  $R^2 = 98\%$  lo que indica que el ajuste es muy bueno.

<sup>3</sup> Yadolah Dodge. *The Oxford Dictionary of Statistical Terms*. Oxford University Press, 2003. ISBN 0-19-850994-4

$$y = 1219.7e^{0.0857x} \quad (2.2)$$

Para el Gasto Corriente se obtuvo la ecuación 2.3, obteniendo un coeficiente de correlación  $R^2 = 96.77\%$  lo que indica que el ajuste es también muy bueno.

$$y = 1343.5e^{0.077x} \quad (2.3)$$

Se procuró también obtener en ambas regresiones un sobreajuste ya que la intención de estos resultados es el de completar datos y no precisamente de una predicción.

## 2.3 Variables de ámbito ambiental

### 2.3.1 Temperatura Media Mensual Mundial

Los datos de Temperatura Media Mensual Mundial ( $\backslash$ TMMM) se tomaron de bases de datos oficiales de la NASA<sup>4</sup>. Los datos indican la variación de la Temperatura Promedio Mensual Mundial con referencia al mes de Julio 2018. Para hacer el cálculo de estas variaciones a nivel mundial no se hace distinción entre la temperatura de mares y tierra. Se actualiza cada mes y se tienen datos mensuales desde 1880. Para la representación de esta variable se tomó como referencia el valor de el mes de Enero de cada año o quinquenio utilizado en el modelo desde 1970 hasta 2020.

Actualmente se tiene la idea de que la población joven en México es la más consciente del medio ambiente. Se ha popularizado el reciclaje, el cuidado por el medio ambiente, los animales y el apoyo a fundaciones o movimientos por causas ecológicas, la sobre explotación de recursos y la conciencia con respecto a la sobrepoblación, la cual conlleva a un mayor consumo de los recursos disponibles, empeora las condiciones del entorno e interfiere en la satisfacción de necesidades. Parece importante hacer mención de que la población joven, es además, la que representa una gran proporción de la población en edad reproductiva del país.

Con estos datos se puede pensar que la población en edad reproductiva (siendo conscientes del impacto ambiental que genera la sobrepoblación) podría estar reconsiderando reproducirse en un futuro cercano, afectando directamente las tasas de natalidad.

Con esta variable pretendemos utilizar la temperatura media mensual global para demostrar que a la vez que aumenta la preocupación el medio ambiente, los jóvenes no parecen estar dispuestos a tener hijos, o al menos no a tener muchos.

Afortunadamente no se necesitaron técnicas de completado ya que ésta base de datos no presenta datos faltantes. Además, al representar

<sup>4</sup> 2021 GISTEMP Team. Giss surface temperature analysis (gistemp) version 4. nasa goddard institute for space studies. <https://data.giss.nasa.gov/gistemp/>, July 2021

variaciones de temperatura, los datos ya son lo suficientemente pequeños como para ser incluidos en el modelo y no fue necesario hacer normalización de los datos.

### 2.3.2 *Percepción de la población sobre Impacto Ambiental*

Esta variable busca reflejar, como su nombre lo dice, la percepción de la población. En este caso se utilizó Google Trends para encontrar un patrón de búsquedas con palabras relacionadas al cambio climático. Es decir, mientras más preocupación haya sobre el medio ambiente, más búsquedas relacionadas deberían reflejarse en la gráfica de tendencia. Esta variable no pudo ser utilizada ya que los datos estaban representados en proporción de búsquedas con relación a la fecha que más búsquedas había tenido del periodo. Por lo tanto, no pudieron adaptarse al tipo de datos que se utiliza en este modelo.

## 2.4 *Variables de ámbito social*

### 2.4.1 *Empoderamiento Femenino*

Esta variable hace referencia a la participación de las mujeres en la educación y ámbito laboral. Se sabe que la validez y participación de las mujeres a lo largo de la historia ha requerido diversos movimientos sociales. De acuerdo con los datos recaudados, fue a partir del año 2000 cuando se ve un aumento significativo en la presencia de mujeres en estudios profesionales y posgrados. Los datos del censo de 2020<sup>5</sup> nos permiten ver que sólo el 29.31% de las mujeres mexicanas censadas han tenido acceso a estudios superiores.

Por otro lado, fue a partir de los años 90's en donde se puede apreciar un incremento en el empleo de mujeres con respecto al total de mujeres censadas. Pareció importante remarcar la comparación contra los datos de hombres en México. Se midió la proporción de mujeres trabajando contra el total de hombres asalariados, lo que deja ver la equidad de oportunidades para el empleo. En 1970 podías encontrar a 23 mujeres trabajando por cada 100 hombres, a diferencia del año pasado donde se encuentran 69 mujeres por cada 100 hombres trabajando. Es importante mencionar que esta métrica no refleja el tipo de trabajo de cada sexo.

¿Qué implicaciones tienen estos porcentajes en el tamaño de la población? El **empoderamiento femenino** ha implicado grandes cambios sociales, por ejemplo, los bajos índices de natalidad, la transformación del modelo de familia, la independencia económica de las mujeres y el entendimiento de los roles no como una asignación sino como reparto de estos. Estos cambios, como cualquier otro, han enfrentado cierta resistencia en una parte de la población, sin embargo,

<sup>5</sup> INEGI. Fuente de datos inegi. <https://www.inegi.org.mx/datos/>, 2021b

con el paso del tiempo y las nuevas generaciones la aceptación ha incrementado y se ha visto una gran influencia y apoyo hacia estas nuevas ideologías de género.

La intención de incluir esta variable en el modelo es la de remarcar la importancia de la elección femenina y la ocupación en tareas fuera del hogar, la cual ha repercutido en la natalidad del país y la aportación económica de la mujer al núcleo familiar.

Éste fenómeno fue agregado al modelo, sin embargo se representó con tres distintas variables:

- Proporción de Mujeres estudiando sobre Total de Población.
- Proporción de Población Económicamente Activa (PEA) femenina sobre PEA Masculina.
- Proporción de PEA femenina sobre PEA Total.

Éstas variables se explican a detalle a continuación.

### *Educación Femenina*

Los datos originales fueron tomados de bases de datos de censos de INEGI. Muestran el número de estudiantes de sexo femenino con estudios mínimos de grado técnico, profesional o posgrado. Para complementar esta variable, se obtuvieron datos sobre el número total de mujeres estudiando durante cada año de nuestro periodo. Estos datos también fueron obtenidos de bases de datos de censos realizados por el INEGI. Con ambos datos se calculó la proporción de mujeres que llega a niveles de educación superior en México desde 1970 hasta 2020.

Ésta base de datos tiene algunos faltantes en los años de 1995 y 2015. Para hacer el completado de datos, se utilizó una regresión cuadrática que dio como resultado la ecuación 2.4 que obtuvo con un coeficiente de ajuste  $R^2 = 92.75\%$  el cuál también tiene muy buen ajuste. Estos datos no requirieron de normalización posterior.

$$y = 0.0002x^2 - 0.0028x + 0.015 \quad (2.4)$$

Una vez más se tomo el año de 1970 como año 0, 1980 como el año 10, y sucesivamente se fueron designando los valores para el resto de los años. En la gráfica 2.2 se muestra la tendencia y el ajuste de la regresión.

### *Proporción de PEA femenina sobre PEA Masculina y Proporción de PEA femenina sobre PEA total.*

Los datos que conforman éstas variables fueron tomados de los censos y conteos poblacionales de INEGI desde 1970 hasta 2020. Para

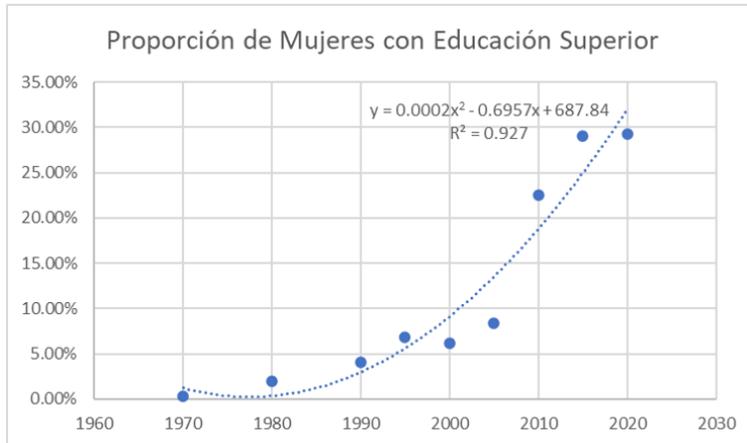


Figure 2.2: Proporción de la participación de las mujeres en la educación superior durante 1970 - 2020

esta variable se consideró únicamente la **\Población Económicamente Activa**, la cuál se refiere a los habitantes mayores a 14 años que se encuentran trabajando o, en su defecto, se encuentran dentro de la población desocupada pero que buscan activamente un trabajo.

Para la primera variable se tomaron los datos del total de Mujeres Trabajando y se dividió por el Total de hombres económicamente activos. Este cociente nos da la relación de cuántas mujeres trabajan por cada hombre activamente trabajando. Es decir, indica la proporción de Mujeres que trabajan con respecto al Total de Hombres encuestados que trabajan. Buscando representar la equidad de género en la participación laboral y de oportunidades del país.

Por otro lado, la segunda variable utiliza el Total de Mujeres Trabajando dividida por el Total de Personas Económicamente Activas en el país. Se buscó tener una representación general de la fuerza laboral femenina y su incremento o decremento a lo largo del tiempo.

El incremento de ambas variables indicarían que las mujeres dejan de depender económicamente total o parcialmente de los hombres y se vuelven parte activa de la población y proveedoras para las familias.

Se pudo conseguir toda la información necesaria para completar esta variable. Al ser ambas variables una proporción sobre el total de hombres o de personas en el país, no se necesitó de normalización previa a su uso en el modelo. En las gráficas 2.3 y 2.4 se pueden observar las tendencias de ambas variables a lo largo de 1970-2020.

#### 2.4.2 Liberación de la diversidad sexual

El término **\diversidad sexual** cuestiona la idea de que hay una única forma de ejercer la sexualidad y los afectos, haciendo visible la existencia de otras formas de expresarlos. Incluye también la idea de que la identidad de género de una persona puede ser independiente

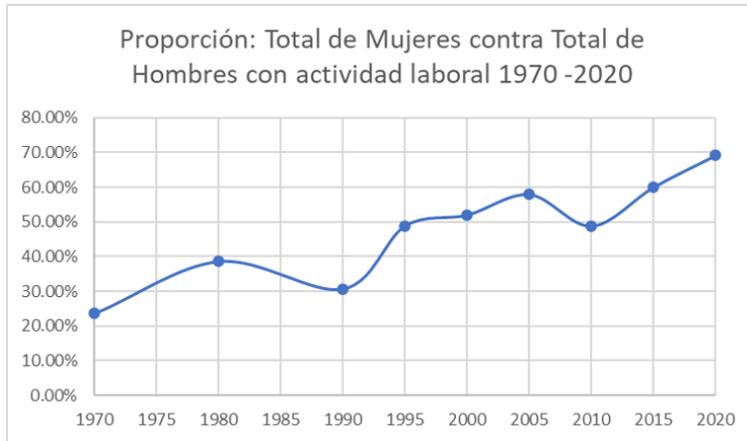


Figure 2.3: Proporción de mujeres como parte de la población económicamente activa en relación al total de hombres trabajando durante 1970 - 2020

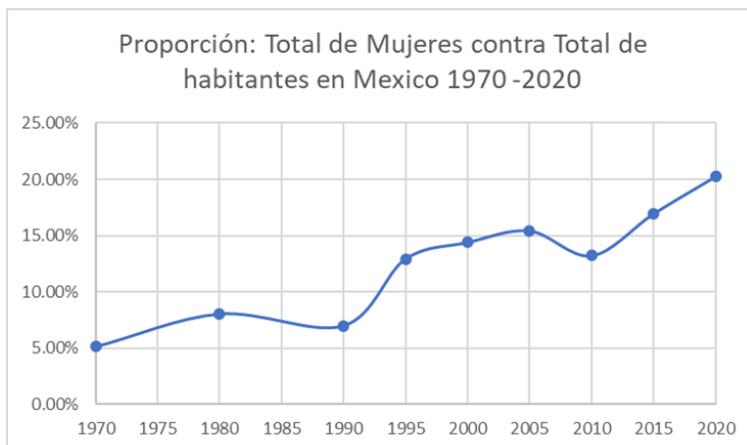


Figure 2.4: Proporción de mujeres como parte de la población económicamente activa en relación al total de personas censadas durante 1970 - 2020

del sexo con el que nació y su orientación sexual.

A pesar de que no se tiene claro en dónde comenzaron las ideologías de género, lo que se puede decir es que este enfoque busca asegurar la paridad de oportunidades, derechos y responsabilidades tanto para hombres como para mujeres, es decir, reconocer la diversidad de ambos grupos y asegurar que el sexo con el que nacieron no sea determinante en cuanto al acceso a oportunidades y derechos.

Naturalmente, con la aceptación de las ideologías de género, la percepción sobre el incremento en el número de parejas dentro de la población LGBTQ+ se ha reafirmado poco a poco. Lo que nos permite también pensar en un nuevo concepto de familia, donde no necesariamente se requieren un hombre y una mujer para la formación de dicho núcleo, sino que plantea una familia como el conjunto de personas (independientes del sexo) unidas por vínculos de matrimonio, parentesco o adopción, considerada como una comunidad natural y universal con base afectiva, que influye en la formación del individuo y tiene interés social.

Este cambio social puede ser considerado un factor en la disminución de la natalidad. Lamentablemente, no se pudo conseguir ningún dato duro que pueda representar de forma acertada la liberación de las ideologías de género, por lo que no fue utilizada.

### 2.4.3 *Anticonceptivos y Aborto*

El uso de anticonceptivos y el aborto son, naturalmente, un impacto directo en las tasas de natalidad de cualquier país. Sin embargo, no fue hasta el mes de Septiembre del 2021 que la Suprema Corte de Justicia en México declaró inconstitucional la penalización el aborto en el país. Eso sin mencionar que para el mismo mes ya existían 4 estados que permitían el aborto bajo ciertas circunstancias. Estos cambios sociales se han visto influidos por las ideas de los jóvenes con respecto a la normalización del aborto, del uso de anticonceptivos, y de la educación sexual desde niveles de educación básica.

Como se planteó en el capítulo 1, el INEGI está constantemente ampliando su cuestionario con la intención de obtener datos relevantes y de interés. Sin embargo, este tipo de datos sobre el aborto y el uso de anticonceptivos no se ven reflejados aún en estas métricas.

Por otro lado, el IMSS sí cuenta con ciertos datos sobre el uso de anticonceptivos, pero resultan sesgados, ya que sólo se contabilizan los métodos anticonceptivos semi-permanentes o definitivos (DIU, operaciones definitivas, etc...), es decir, que no se lleva un conteo o registro de los métodos anticonceptivos más usados. Con respecto al número de abortos practicados tampoco se tiene un número definitivo ya que en muchos estados de la república el aborto no es legal, así

que no se tiene una buena representación de esta métrica. No se tomó para evitar más subjetividad en el estudio. Por ambas situaciones, esta variable no pudo ser incluida en el modelo.

#### 2.4.4 Estabilidad en pareja

Dentro del cuestionario de INEGI podemos obtener el dato del estado civil de la población. En este caso, se tomaron datos de los censos de INEGI sobre Matrimonio y Divorcios.

De acuerdo a las cifras encontradas en INEGI, la proporción de divorcios registrados en México ha ido en aumento. En especial entre los años de 2015 y 2020, se tuvo un aumento de casi el 2% de la población divorciada legalmente, tomando en cuenta que estos datos solo consideran el divorcio legal y no las separaciones no oficiales de las parejas. En las figuras 2.5 y 2.6 se pueden ver las tendencias de ambas variables durante el periodo de 1970 a 2020.

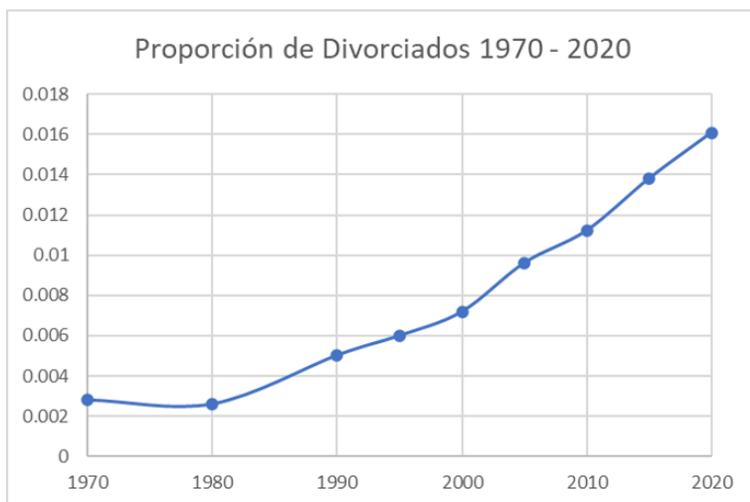


Figure 2.5: Proporción de parejas divorciadas en relación al total de personas censadas durante 1970 - 2020

La intención al incluir estas variables es la de remarcar la importancia del cambio en el concepto de familia, en el que se ve afectada la natalidad por los divorcios y la preferencia hacia la soltería. Se incluyeron dos variables en el modelo para representar este tema:

- Proporción de Divorcios
- Proporción de Solteros

Dichas variables se detalla a continuación.

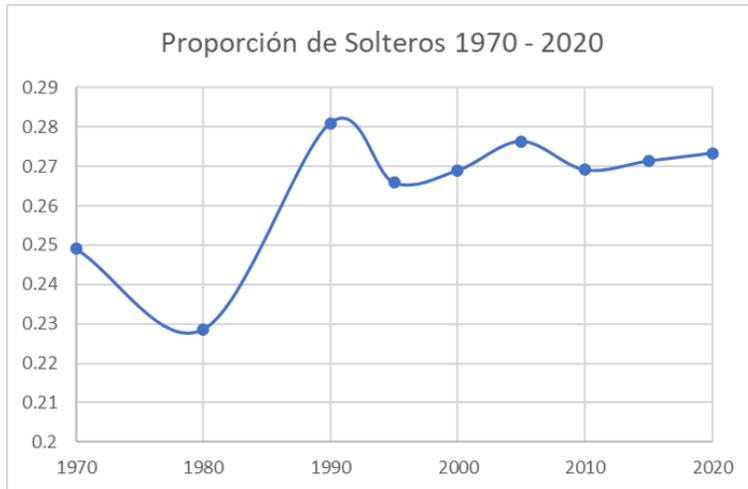


Figure 2.6: Proporción de personas solteras en relación al total de personas censadas durante 1970 - 2020

### *Proporción de Divorcios y Proporción de Solteros*

Ambas variables fueron calculadas de forma similar. Se utilizaron datos históricos de los censos de INEGI, de los cuales se tomaron datos del total de personas censadas con estado civil soltero, total de personas que concluyeron el proceso legal de divorcio y el total de habitantes el en país.

Para obtener la proporción de solteros se dividió el total de personas con estado civil solteras por el total de habitantes en México. De la misma forma, para la proporción de divorcios, se dividió el total de personas divorciadas por el total de habitantes en el país.

Afortunadamente para ambas variables se encontraron todos los datos necesarios, por lo que no fue necesario hacer completado de datos. Además dado que los datos propuestos son proporciones, tampoco requieren normalización.

#### 2.4.5 *Hacinamiento en zonas urbanas*

Como ya se había comentado en la sección 2.3, parte de la preocupación por el medio ambiente se deriva directamente de la sobrepoblación. Debido a esto, se decidió incluir una variable que representara el cambio a través del tiempo del **hacinamiento** de las principales zonas metropolitanas del país.

La Organización Mundial de la Salud (OMS)<sup>6</sup>, define el hacinamiento como la “condición donde el número de ocupantes excede la capacidad de espacio de vivienda”; el indicador: 3 personas o más viviendo en una misma habitación. Se han hecho estudios que indican que cuando una vivienda presenta hacinamiento, sus ocupantes presentan mayor riesgo de sufrir ciertas problemáticas como violencia doméstica,

<sup>6</sup> OMS. Hacinamiento y limitación de la conducta. obtenido de universitat de barcelona., 2018

desintegración familiar, bajo rendimiento escolar, entre otros. Además, la OMS ha encontrado relaciones entre el hacinamiento y la salud, donde las enfermedades respiratorias, gastrointestinales y de salud mental se reflejan en estrés, ansiedad y baja calidad del sueño.

Además, la salud se encuentra intrínsecamente relacionada con las condiciones de vida, por lo que poblaciones que se encuentran muy hacinadas se ven vulneradas en diversos aspectos. Debido a las carencias que presenta este tipo de población, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), cataloga a dicha población en el rubro de extrema pobreza.

Todos estos planteamientos nos permiten explicar un poco mejor algunas de las posibles creencias por las cuales los jóvenes hoy en día están decidiendo no reproducirse. Tomando en cuenta la calidad de vida y el impacto ambiental, estas variables se vuelven cada vez más importantes en el fenómeno que queremos describir.

Para delimitar esta variable se tomaron las 10 zonas metropolitanas más importantes de México declaradas en el censo de 2020. El Instituto Nacional para el Federalismo y el Desarrollo Municipal (INAFED<sup>7</sup>) define una *zona metropolitana* como "el conjunto de dos o más municipios o demarcaciones territoriales en los que se localiza una ciudad de 50 mil o más habitantes, cuya área urbana, funciones y actividades rebasan el límite del municipio o demarcación que originalmente la contenía, incorporando como parte de sí misma o de su área de influencia directa a municipios vecinos, predominantemente urbanos, con los que mantiene un alto grado de integración socioeconómica."

Las 10 principales zonas metropolitanas por número de habitantes, de acuerdo al censo realizado por el INEGI en 2020<sup>8</sup> son:

1. Zona metropolitana del Valle de México.
2. Zona metropolitana de Monterrey.
3. Zona metropolitana de Guadalajara.
4. Zona metropolitana de Puebla-Tlaxcala.
5. Zona metropolitana del Valle de Toluca.
6. Tijuana.
7. Zona metropolitana de León.
8. Zona metropolitana de Querétaro.
9. Zona metropolitana de Ciudad Juárez.
10. Zona metropolitana de La Laguna

<sup>7</sup> INAFED. Conoce los beneficios de pertenecer a una zona metropolitana. <https://www.gob.mx/inafed/articulos/conoce-los-beneficios-de-pertenecer-a-una-zona-met>  
June 2017

<sup>8</sup> INEGI. Fuente de datos inegi. <https://www.inegi.org.mx/datos/>, 2021b

Una vez ubicada cada zona metropolitana se buscó el hacinamiento histórico de cada una de ellas, tomando en cuenta el crecimiento de la misma y la posible inclusión de nuevos municipios a la zona metropolitana a través de los años. Sin embargo, no se encontró ningún dato concreto que tomara en cuenta todas estas especificaciones por lo que se decidió hacer una estimación propia en base a otras variables disponibles. Para lograr esto, se utilizaron tres distintos datos, la población total por municipio en cada uno de los años seleccionados para el proyecto, la extensión territorial de cada municipio y el año en que dicho municipio fue incorporado a la Zona Metropolitana correspondiente.

En la tabla 2.1 se muestra un ejemplo de los datos recopilados para la Zona Metropolitana de Monterrey (ZMM) obtenida de datos publicados por INEGI y CONAPO. Se muestran todos los municipios que son considerados parte de la zona metropolitana de acuerdo al censo de 2020. Se incluye la columna "HM" que representa el Hacinamiento Municipal de acuerdo al censo de 2020 y también en la columna "km2" se muestra la extensión territorial de cada uno de los municipios. Las columnas que representan los años estudiados nos muestran el dato del total de población de los municipios en cada uno de los años estudiados. Como se puede observar para algunos municipios no se consideraron los datos de todos los años. Esto debido a que, como se explicó anteriormente, se consideró el año en el que cada uno de los municipios pasó a conformar parte de la zona metropolitana.

| Hacinamiento de la Zona Metropolitana de Monterrey

Table 2.1: Tabla para el cálculo del hacinamiento histórico de la Zona Metropolitana de Monterrey. 1970 - 2020.

ZMM	HM	km2	1970	1980	1990	1995	2000	2005	2010	2015	2020
Apodaca	3,577.46	183.5	18,600	37,100	115,900	219,100	283,497	418,784	523,370	597,207	656,464
San Pedro Garza G	1,904.45	69.4	48,300	82,000	113,000	120,800	125,978	122,009	122,659	123,156	132,169
General Escobedo	2,323.58	207.1	10,500	37,800	98,200	176,800	233,457	299,364	357,937	425,148	481,213
Guadalupe	4,250.78	151.3	170,200	370,900	535,600	618,900	670,162	691,931	678,006	682,880	643,143
Monterrey	1,278.52	894	871,500	1,090,000	1,090,000	1,088,000	1,110,997	1,133,814	1,135,550	1,109,171	1,142,994
San Nicolás de los G	6,858.55	60.1	118,100	280,700	436,600	487,900	496,878	476,761	443,273	430,143	412,199
Juárez	1,697.35	277.80	-	13,500	28,000	50,000	66,497	144,380	256,970	333,481	471,523
El Carmen	797.54	131.00	-	-	-	-	-	-	16,092	38,306	104,478
Santa Catarina	311.14	984.5	38,100	89,500	163,900	202,100	227,026	259,896	268,955	296,954	306,322
García	384.11	1,034.08	-	-	13,100	24,000	28,974	51,658	143,668	247,370	397,205
Cadereyta Jiménez	121.80	1004.4	-	-	-	-	75,059	73,746	86,445	95,534	122,337
Salinas Victoria	65.04	1334	-	-	-	-	19,024	27,848	32,660	54,192	86,766
Santiago	61.25	763.8	-	-	-	-	36,812	37,886	40,469	42,407	46,784
Ciénega de Flores	469.26	146.5	-	-	-	-	-	-	-	42,715	68,747
Pesquería	480.08	307.5	-	-	-	-	-	-	-	87,168	147,624
Hidalgo	94.56	170.12	-	-	-	-	-	-	-	13,836	16,086

Los primeras dos variables fueron relativamente fáciles de encontrar, ya que INEGI cuenta con estos datos dentro de los datos públicos en su página web. Sin embargo, la fecha de inclusión de cierto municipio a la zona metropolitana fue más difícil de encontrar. Para este paso

una de las fuentes más importantes de información fueron las bases de datos de CONEVAL, ya que se puede encontrar una base de datos que muestra la inclusión de cada municipio para la mayoría de las zonas metropolitanas.

Se calculó la extensión territorial y el total de habitantes de cada zona metropolitana en cada uno de los años estudiados de acuerdo a la inclusión histórica progresiva de los municipios. De esta forma nos aseguramos de que los datos del hacinamiento histórico fueran los más precisos posibles. En la gráfica 2.7 puede verse para la Zona Metropolitana de Guadalajara (ZMG) la población y su incremento histórico, el aumento de su extensión territorial considerada como parte de la ZMG, de acuerdo a la inclusión de los municipios y, por último, podemos ver el hacinamiento para cada uno de los años. Si prestamos atención podemos ver que la extensión territorial durante 1970 y 1980 es la misma, es decir, no se agregó ningún municipio en esas décadas. Sin embargo, la población total aumentó de 1970 a 1980, por lo que el hacinamiento también se incrementó. Por otro lado, para 1990, tanto la extensión territorial de la zona metropolitana como la población incrementaron, por lo que se puede ver una disminución clara en el hacinamiento de la ZMG.

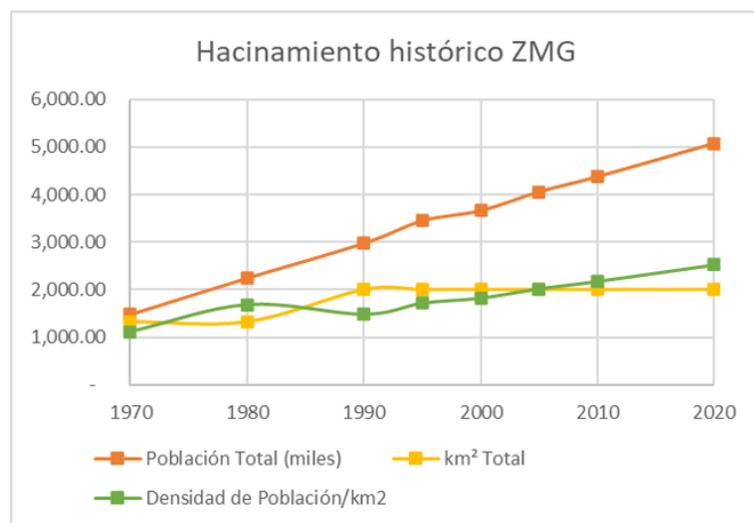


Figure 2.7: Ejemplo de hacinamiento histórico de ZMG desde 1970 hasta 2020.

Con la intención de evitar sesgo de datos al añadir municipios poco hacinados pero que por la cercanía a la zona metropolitana se incluyeron en ellas, se decidió ignorar los municipios que tuvieron un hacinamiento menor a los 500 habitantes por kilómetro cuadrado. De esta forma, municipios con poca población pero gran extensión territorial no afectarían el cálculo del hacinamiento de la zona metropolitana. En la tabla 2.1 se puede ver una línea horizontal que divide a los municipios que tienen un Hacinamiento Municipal por

debajo de los 500 habitantes por kilómetro cuadrado. Dichos municipios no fueron contemplados para el cálculo del Hacinamiento de la ZMM.

Para calcular el hacinamiento final de cada zona metropolitana se sumaron el total de habitantes y se dividió por la suma total de kilómetros cuadrados de extensión territorial. Se hizo esto para cada año en cada una de las zonas metropolitanas. Finalmente, para obtener el hacinamiento aproximado del país se promedió el hacinamiento de las diez zonas metropolitanas de cada año.



Figure 2.8: Zonas Metropolitanas más importantes de acuerdo al censo INEGI 2020

Es importante destacar que a pesar de que se tuvieron muchos datos faltantes, no excedieron el 6%, Los cuales, en la mayoría de los casos, pudieron ser completados con regresiones lineales.

Dos de las zonas metropolitanas si requirieron de regresiones más específicas. La Zona Metropolitana del Valle de México presentó datos faltantes en el año de 1995 que requirieron ser completados con una regresión cúbica. Después del ajuste dicha regresión tuvo un coeficiente  $R^2$  de 98.26%, en este caso se consideró muy buen ajuste.

Por otro lado, la Zona Metropolitana de Guadalajara tuvo datos faltantes en el año 2015, de igual forma se utilizó una regresión potencia 4 para hacer el completado de datos y obtuvo un coeficiente  $R^2$  de 97.59%, de la misma forma muy buen ajuste. En las gráficas 2.9 y 2.10 se puede ver el ajuste que se hizo a cada una de las variables.

Éstos datos también requirieron de una normalización previa a su uso en el modelo ya que los datos representaban el promedio de habitantes por kilómetro cuadrado y el rango de los datos era bastante amplio.

Como ya se había mencionado, el hacinamiento alto está relacionado con problemas de salud y calidad de vida. Además, en los estudios presentados en el capítulo 1 se proyecta que la población sea el doble de grande llegando al año de 2050, por lo que pretendemos mostrar el

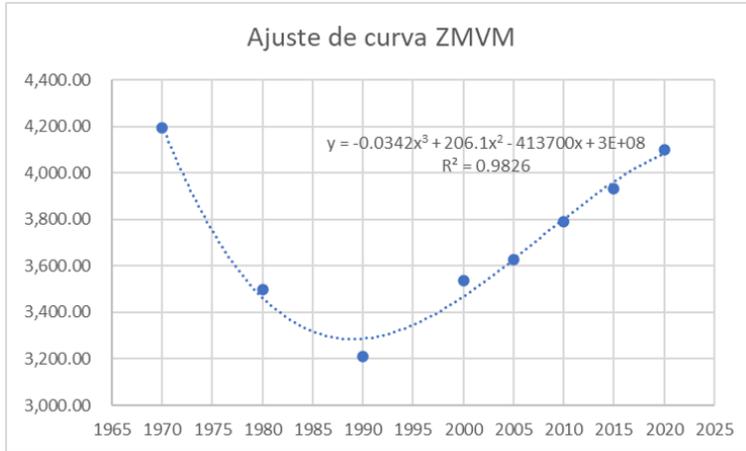


Figure 2.9: Hacinamiento y ajuste de curva para la Zona Metropolitana del Valle de México 1970 - 2020

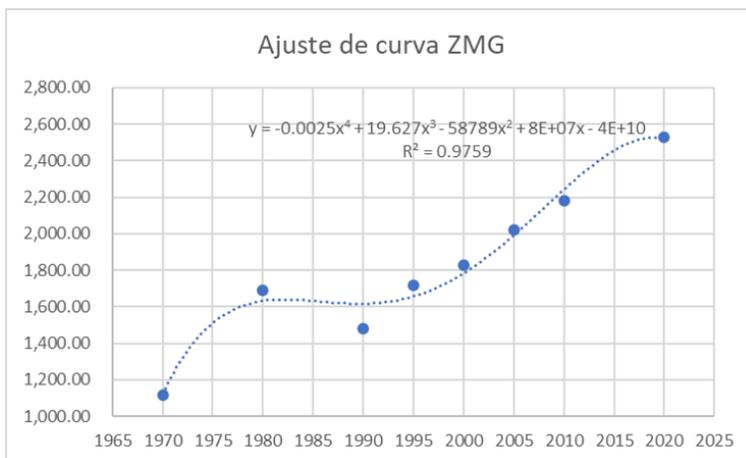


Figure 2.10: Hacinamiento y ajuste de curva para la Zona Metropolitana de Guadalajara 1970 - 2020

crecimiento acelerado que ha presentado el país en los últimos años y el impacto que este pueda tener.

## 2.5 Variables de ámbito histórico

### 2.5.1 Baby Boom

Se le conoce como **Baby Boom** a la época donde la natalidad "explotó" posterior a la Segunda Guerra Mundial, entre 1946 y 1965. Vivieron las grandes crisis y revoluciones de la historia contemporánea. Hasta el 2018, esta generación representaba aproximadamente el 10% de la población mexicana. Este hecho histórico es el que marca el inicio de la transición de la pirámide poblacional y, posteriormente, la transición hacia una pirámide regresiva en México. Si observamos la pirámide poblacional de 1970, podemos observar que la base es significativamente grande, con el paso de los años y con el crecimiento de esta generación, éste bloque comienza a ascender hacia la parte media.

Al inicio, se planteó considerar esta variable como una variable dicotómica en la cual se considerara el efecto antes y después del Baby Boom. Sin embargo, se decidió tomar en cuenta este hecho histórico únicamente como referencia para la elección del periodo de observación de nuestras variables (1970 - 2020).

En la tabla 2.2 se presentan las variables propuestas que no se incluyeron en el trabajo, así como la justificación por la que fueron removidas. Así mismo, en la tabla 2.3 se enlistan las variables que sí fueron utilizadas en el presente trabajo, así como el indicador concreto que fue utilizado.

Variables Propuestas No Incluidas en el Modelo

Variable Propuesta	Tipo	Justificación del Rechazo
Inflación	Económico	Se decidió utilizar una variable que representara una mejor forma el impacto en el poder adquisitivo de las familiar mexicanas.
Percepción de la población	Ambiental	El tipo de dato que se encontró no daba una buena representación ya que se mostraba como referencia de 100% el día con más búsquedas del periodo.
Liberación de la diversidad sexual	Social	No existen registros específicos ni exactos sobre la proporción de población que se identifica como parte de la comunidad LGBTQ+.
Anticonceptivos y Aborto.	Social	No existe un registro sobre el uso de anticonceptivos de venta libre. Debido a la reciente aceptación legal del aborto en México tampoco existen registros fidedignos.

Table 2.2: Variables propuestas no incluidas en el modelo y su justificación.

En el siguiente capítulo se mostrará la metodología elegida para revisar el efecto de las variables propuestas con la pirámide poblacional.

| Variables Propuestas Incluidas en el Modelo

Table 2.3: Variables propuestas incluidas en el modelo y el indicador utilizado.

Variable Propuesta	Tipo	Indicador
Ingreso Corriente Total sobre Total Hogares.	Económico	Monto en miles de pesos por Hogar normalizado.
Gasto Corriente Total sobre Total Hogares.	Económico	Monto en miles de pesos por Hogar normalizado.
Temperatura Media Mensual Mundial.	Ambiental	Porcentaje de variación.
Proporción de Mujeres Estudiando sobre Total de Población.	Social	Proporción porcentual de mujeres estudiando.
Proporción de PEA femenina sobre PEA Masculina.	Social	Proporción porcentual de mujeres trabajando por cada hombre trabajando
Proporción de PEA femenina sobre PEA Total.	Social	Proporción porcentual de mujeres trabajando por cada mexicano trabajando.
Proporción de Divorciados.	Social	Proporción porcentual de divorciados.
Proporción de Solteros.	Social	Proporción porcentual de mexicanos solteros.
Hacinamiento Promedio de ZM.	Social	Número de personas por $km^2$ de extensión territorial normalizado.

# 3 Metodología

## Contents

---

3.1	Descripción de salidas por medio del ajuste a curvas de Distribución de Probabilidad conocidas	39
3.1.1	Distribución Gamma . . . . .	40
3.1.2	Distribución Log-Normal . . . . .	41
3.2	Método para el reconocimiento de patrones: red neuronal . . . . .	42
3.3	Cálculo del error . . . . .	44

---

En este capítulo se describe la teoría de los métodos usados en este trabajo. Se explica el tipo de salidas que se pretende obtener y cómo fueron modeladas a distintas distribuciones de probabilidad las pirámides poblacionales.

### 3.1 Descripción de salidas por medio del ajuste a curvas de Distribución de Probabilidad conocidas

La pirámide poblacional describe, como se había planteado en el capítulo 1, la forma en que está conformada una población. Sin embargo, obtener un modelo que describa su comportamiento resulta muy complicado si se obtienen tantas salidas como quinquenios en la pirámide poblacional. En nuestro caso la pirámide poblacional esta representada por 18 grupos quinquenales, lo que implicaría tener 18 salidas en el modelo. Por lo tanto, se decidió hacer un modelo que describiera la silueta de la pirámide poblacional y no el comportamiento individual de cada barra.

Para lograr lo anterior, se decidió girar horizontalmente la pirámide poblacional, apilando ambas barras (hombres y mujeres), comenzando desde los primeros quinquenios hasta los adultos mayores y acomodados de izquierda a derecha respectivamente. Con esto se obtuvo un histograma horizontal como se muestra en la figura 3.1, el cual representa los quinquenios de edad en su eje x y el número de personas, indistinto del sexo, en el eje y. Posteriormente, se trazó una curva que refleja la silueta de la pirámide poblacional. Uniendo el

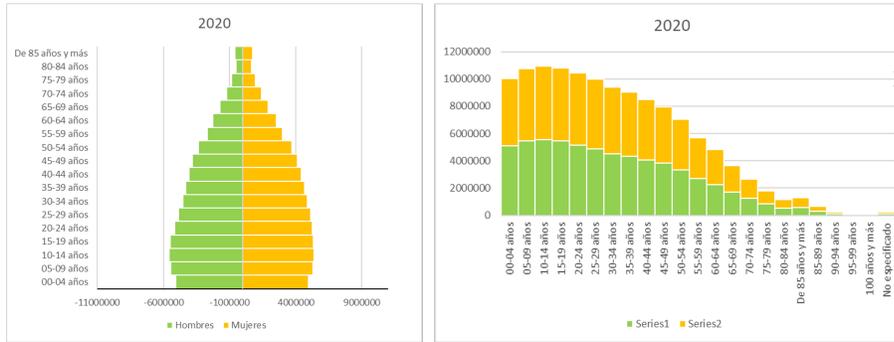


Figure 3.1: Pirámide Poblacional 2020 vista horizontal previa a la parametrización.

punto más alto de cada barra. Véase la figura 3.2 como referencia de este paso.

Este cambio se hizo con la intención de ajustar distintas curvas de distribución de probabilidad a la curva obtenida en el histograma de la pirámide poblacional. De esta manera se obtienen una silueta de la pirámide poblacional que fácilmente ajustable a las barras del histograma. La mayor ventaja es que se disminuye el número de variables de salidas del modelo. Ya que al usar una distribución de probabilidad, las variables de salida del modelo se acotan únicamente a los parámetros que describen cada una de las distribuciones.

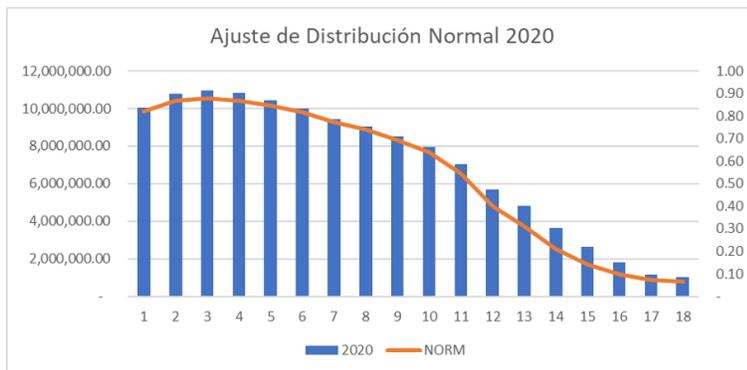


Figure 3.2: Pre-procesamiento de la Pirámide Poblacional 2020, Ajuste Distribución Normal.

Para este modelo se probaron distintas distribuciones de probabilidad buscando la que mejor se ajustara a la silueta de la curva que cae sobre el histograma, como lo muestra la figura 3.2.

### 3.1.1 Distribución Gamma

La **distribución Gamma**<sup>1</sup> también pertenece al grupo de distribuciones continuas y está conformada por dos parámetros, En nuestro caso usaremos los parámetros  $\lambda$ , que representa al parámetro inverso de escala, y  $\alpha$ , que representa el parámetro de forma. Su función de densidad está descrita en la ecuación 3.1 y la variación de sus

<sup>1</sup> NIST/SEMATECH. *e-Handbook of Statistical Methods*. 2012. DOI: <https://doi.org/10.18434/M32189>

parámetros causa modificaciones en la curva de su distribución de densidad como puede apreciarse en la figura 3.3.

$$X \sim \gamma(\alpha, \lambda) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad (3.1)$$

Una de las grandes ventajas que nos da esta distribución es que sabemos exactamente cómo está conformada la silueta de la curva de distribución. Esto resulta bastante útil al momento de hacer el ajuste con los datos de las distintas pirámides poblacionales y también nos genera buenas aproximaciones en algunos de los años utilizados en nuestro modelo. Puede verse una comparación de la parametrización gamma que se hizo para nuestros datos en la figura 3.4

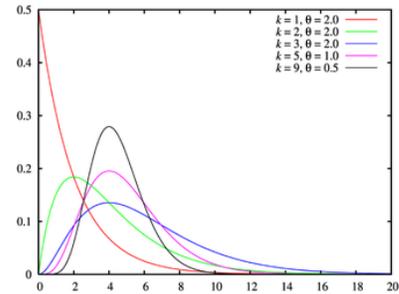


Figure 3.3: Curva de función de densidad de la distribución Gamma con distintos parámetros.

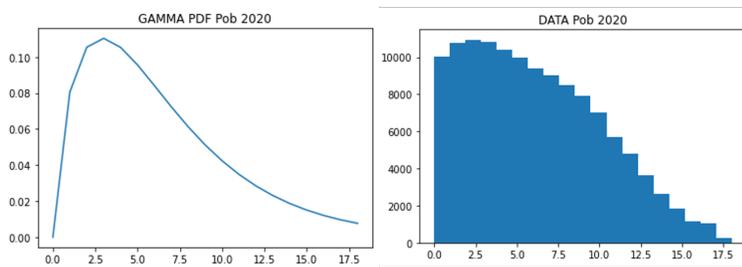


Figure 3.4: Comparación de la curva obtenida con el ajuste gamma para los datos del censo 2020.

### 3.1.2 Distribución Log-Normal

La **Distribución Normal Logarítmica** es una distribución de probabilidad continua de una variable aleatoria cuyo logaritmo está normalmente distribuido. Es decir, si  $X \sim N(\mu, \sigma)$ , entonces  $e^X \sim \text{Lognormal}(\mu_x, \sigma_x^2)$ .

Su función de densidad está dada por la ecuación 3.2. En donde podemos ver que el logaritmo natural de  $X$  sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$ .

$$\log(X) \sim N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad (3.2)$$

Además, en la figura 3.5 podemos apreciar el cambio que genera en la curva de la función de densidad los cambios en cada uno de los parámetros. Esta distribución es una de las que mejor logró ajustarse a nuestros datos, como puede apreciarse en la gráfica 3.6 donde vemos lado a lado la comparación de la silueta de la curva sobre nuestros datos de la pirámide poblacional 2020 y el ajuste logrado para esos datos con la distribución Log Normal.

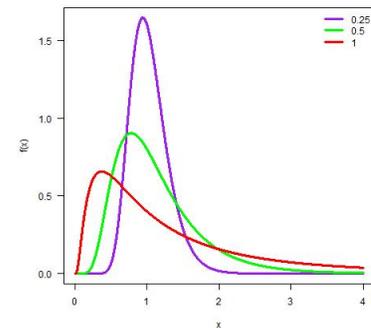


Figure 3.5: Curva de función de densidad de la distribución Log Normal con distintos parámetros.

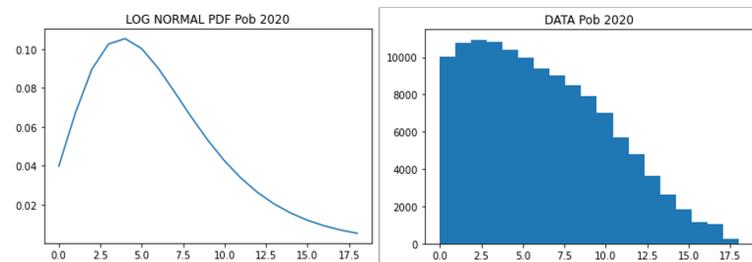


Figure 3.6: Comparación de la curva obtenida con el ajuste Log Normal para los datos del censo 2020.

### 3.2 Método para el reconocimiento de patrones: red neuronal

Existen dos tipos de aprendizaje: supervisado y no supervisado. En el caso del **\aprendizaje supervisado**, es aquel en el que se aprende a partir de salidas conocidas, por lo que se ajustan funciones a un conjunto de ejemplos de los que conocemos la relación entre la entrada y la salida. Los modelos de **\aprendizaje no supervisado** son aquellos en los que no estamos interesados en ajustar pares (entrada, salida), sino en aumentar el conocimiento estructural de los datos disponibles. El resultado de estos procesos nos lleva a dos tipos de resultados en un problema: clasificación y regresión. Cuando usamos **\clasificación**, el resultado son clases o grupos. Con clases nos referimos a categorías arbitrarias según el tipo de problema. Cuando usamos **\regresión**, el resultado es una función que vincula las entradas y las salidas del problema.

Para hacer el procesamiento de los datos de este modelo en específico se utilizó una red neuronal con Backpropagation con función de activación sigmoide. El modelo es supervisado y sus resultados son regresivos. La idea del modelo se muestra en la figura 3.7.

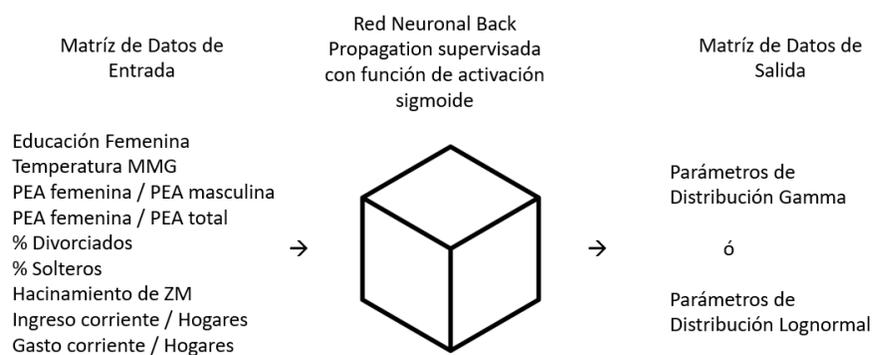


Figure 3.7: En este diagrama se muestra la estructura de los datos de entrada, el modelo de procesamiento de datos que se utilizará y las salidas.

Se utilizó una red neuronal para el desarrollo de este trabajo ya que permite trabajar un modelo complejo con muchas variables de entrada. En este caso se tienen modelos que pretenden utilizar nueve variables

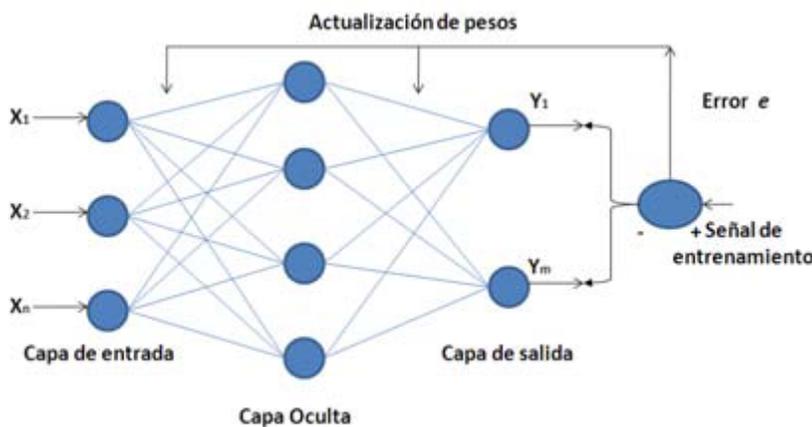
independientes y dos o tres variables dependientes, de acuerdo a la distribución de probabilidad que se utilice. Por este motivo resulta conveniente utilizar este tipo de red neuronal. Por otro lado, la red neuronal permite utilizar la probabilidad de salida como un dato continuo y acotado a un rango definido. Lo que posteriormente se utiliza para obtener proyecciones de la pirámide poblacional.

La **red neuronal Backpropagation**<sup>2</sup> calcula el error en cada neurona debido a la discrepancia de la salida obtenida con la deseada. Dicho error se calcula a la salida y se propaga o regresa a las capas de neuronas pasadas y con esto se genera el "aprendizaje" (modificación de los pesos asociados a la red). En este sentido, el error que la capa anterior recibe depende de la aportación de cada una de las neuronas ocultas, lo que permite a su vez entender cuál de ellas tuvo mayor aportación al resultado del modelo. Un ejemplo este tipo de redes neuronales con backpropagation<sup>3</sup> puede verse en la figura 3.8.

Se seleccionó la **función sigmoide**<sup>4</sup> como función de activación. Esta función transforma los valores introducidos a una escala continua (0,1), donde los valores altos tienden de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a 0. En la figura 3.9 se muestra su gráfica para referencia. La función de activación sigmoide utilizada se muestra en la ecuación 3.3.

$$\phi(x) = \frac{1}{1 + e^{-x}} \tag{3.3}$$

En este trabajo se utilizó un enfoque continuo de la salida para ajustar parámetros no discretos (de la distribución de probabilidad).



Como variables de entrada se tienen las 9 variables elegidas y de cada una se tienen 9 datos, que representa cada uno de los quinquenios o décadas censadas en México desde 1970 hasta 2020. Por lo tanto

<sup>2</sup>Hamzan Wadi. *Learn from Scratch Backpropagation Neural Networks Using Python GUI* MariaDB. Turida Publisher, 2021

<sup>3</sup>Francisco Cruz Lario Nicolay Mena O'Meara, Eduardo Vicens Salort. Metodologías de inteligencia artificial para la toma de decisiones en la red/cadena de suministro en el contexto de incertidumbre. *3rd International Conference on Industrial Engineering and Industrial Management*, pages 1296–1305, 2009

<sup>4</sup>José David Villanueva García. Redes neuronales desde cero (i) – introducción. <https://www.iartificial.net/redes-neuronales-desde-cero-i-introduccion>, October 2020

Figure 3.8: Ejemplo de Red Neuronal Backpropagation

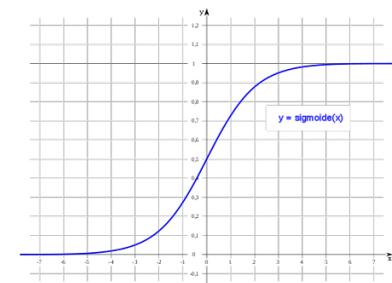


Figure 3.9: Gráfica de la función de activación sigmoide utilizada en la Red Neuronal

tenemos como entrada una matriz de 9x9 cuyos datos se encuentran en el rango continuo [0,1]. Por otro lado, se tienen como variables de salida los parámetros que dan forma a la silueta de las curvas de las funciones de distribución que fueron seleccionada. En este caso se hicieron dos modelos, cada uno con una distribución de salida distinta.

El primer modelo hace la parametrización de las pirámides poblacionales con la función Gamma utilizando las funciones dadas por la librería de Python *scipy*, la cual da como parámetros de salida la inversa de  $\alpha$  y la inversa de  $\lambda$ , esto se traduce en una matriz de 9x2 que representan los parámetros para cada uno de los 9 años seleccionados.

El segundo modelo toma la parametrización con la función Log-Normal obtenida de la misma librería *scipy* en Python, la cual tiene como parámetros de salida de la red neuronal  $\sigma$ , la localización de la curva sobre el eje x, así como  $e^{\mu}$ . De la misma forma, estos datos se traducen en una matriz de 9x3, que representan los parámetros de cada uno de los 9 años seleccionados para este modelo.

### 3.3 Cálculo del error

Para el cálculo del error dentro de la red neuronal se utilizó el error convencional calculado por la norma de los errores de las neuronas de salida. Éste error es el que se minimizó para la convergencia del modelo. Establecimos un error objetivo mínimo que se definió de acuerdo al tiempo de procesamiento y los resultados obtenidos. En la figura 3.10 se puede ver la gráfica del error que se obtuvo en cada una de las iteraciones y podemos ver como baja lo suficiente como para estabilizarse muy cerca de cero. Por lo que podemos decir que efectivamente, el error se estacionó y es suficientemente pequeño como para que el modelo sea confiable. También se puede apreciar la velocidad con la que el error baja a medida que aumenta el número de iteraciones.

A pesar de que tener un error pequeño, que asegura la convergencia del modelo es un indicador de un buen modelo entrenado, se buscó utilizar una medida más acertada con respecto a la exactitud con la cual se habían hecho las predicciones de nuestro modelo. Ya que se está trabajando con curvas que simulan la silueta de la pirámide poblacional se decidió hacer una comparación relativa sobre la diferencia que tienen las curvas originales y las curvas aproximadas. Para este proceso se calculó el área bajo cada una de las curvas utilizando el **método de Simpson 1/3**<sup>5</sup> Éste método permite hacer una aproximación sobre el área de una curva teniendo como beneficio su fácil programación y cálculo. Éste método está dado por la ecuación 3.4.

$$A_{S1/3} = \frac{1}{3}(f_0 + 4f_1 + f_2) \quad (3.4)$$

<sup>5</sup> Jose Luis de la Fuente O'Connor. *Ingeniería de los Algoritmos y Métodos Numéricos*. Editorial Círculo Rojo, 2017. ISBN 978-84-9140-984-7

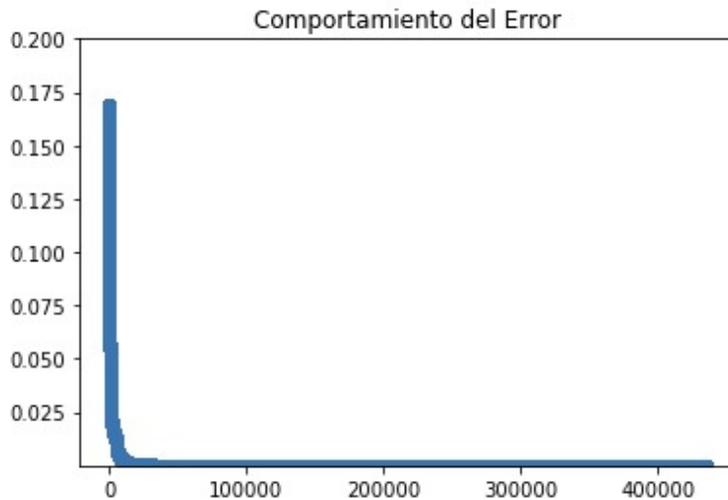


Figure 3.10: Gráfica que muestra el comportamiento del error en un modelo con 6 neuronas ocultas y tasa de aprendizaje del 0.7.

Además se utilizó también la regla del trapecio para el último par de puntos, ya que Simpson  $1/3$  trabaja en grupos de 3 puntos consecutivos y nuestras pirámides poblacionales están dadas por 18 grupos quinquenales. La *regla del trapecio* se puede ver en la ecuación 3.5.

$$A_T = \frac{h}{2}(f_0 + f_1) \quad (3.5)$$

Se decidió utilizar Simpson  $1/3$  porque está probado que otras aproximaciones más extensas como Simpson  $3/8$  provocan sobreajuste de las curvas estudiadas. Mientras que aproximaciones como la trapezoidal se quedan cortas en su ajuste.

Se utilizaron los métodos mencionados con la información de la pirámide poblacional original y con las pirámides poblacionales obtenidas de la red neuronal. Una vez calculadas ambas áreas se calculó el *error relativo* con respecto a las pirámides poblacionales originales. el error relativo se calculó como se describe en la ecuación 3.6, donde  $A_1$  representa el área bajo la curva de la curva que simula la pirámide poblacional original y  $A_2$  representa el área bajo la curva de la curva obtenida de la red neuronal.

$$ErrorRelativo = \frac{|A_1 - A_2|}{A_1} \quad (3.6)$$

En la figura 3.11 se muestra un diagrama en el cual se describe paso a paso el proceso completo desde la recopilación de los datos hasta la obtención de una proyección de la pirámide 2030.

En el siguiente capítulo se muestran los resultados de la implementación del modelo.

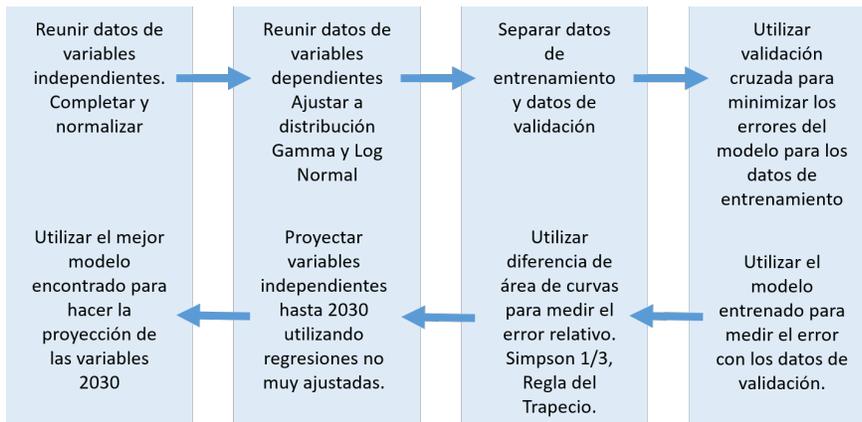


Figure 3.11: Diagrama del proceso completo desde la recolección de los datos hasta la proyección de la pirámide poblacional mexicana para el año 2030.

# 4 Experimentación y resultados

## Contents

---

4.1	Entrenamiento y Selección de configuración para la Red Neuronal . . . . .	47
4.2	Validación de los Resultados Obtenidos . . . . .	48
4.3	Configuración y Error Minimizado . . . . .	49
4.4	Proyección del modelo . . . . .	51
4.5	Análisis de influencia de las variables de entrada . . . . .	51

---

En este capítulo se describe a detalle el procesamiento que se le dio a los datos recolectados que fueron previamente completados y normalizados. El propósito es obtener los resultados de la red neuronal y poder hacer comparaciones con respecto a los datos oficiales de INEGI. Al mismo tiempo, nos parece importante hacer una proyección a futuro de la silueta de la pirámide poblacional.

### 4.1 Entrenamiento y Selección de configuración para la Red Neuronal

En el capítulo 3 se describió el pre-procesamiento de los datos (normalización y completado de datos). Esto se hizo con la intención de cuidar el rango de los datos que posteriormente entrarían al modelo, ya que tener datos con rangos muy variados se podrían generar problemas al intentar combinar los valores, los pesos de nuestra red neuronal y los resultados de salida se podrían ver afectados.

Una vez terminado el pre-procesamiento de los datos se obtuvo la matriz de nueve columnas y nueve filas, que representan las nueve variables seleccionadas y los nueve años de estudio, respectivamente. Es importante mencionar que parte del objetivo de este modelo es el de obtener la significancia de cada una de las variables de entrada en el resultado de la pirámide poblacional.

Además, se obtuvieron dos vectores de resultados utilizando las representaciones estadísticas de las silueta de las pirámides poblacionales (como se describió en la sección 3.1 utilizando las

distribuciones Gamma y Log Normal). Éstos dos vectores se normalizaron utilizando el mismo método planteado en el capítulo 2 en la ecuación 2.1 y fueron utilizados como entrada de nuestro modelo.

Posteriormente se hizo la división en datos de entrenamiento y datos de validación. Para el entrenamiento se utilizaron los años de 1970, 1980, 1990, 2000, 2010 y 2020. Y por otro lado para hacer la validación se utilizaron los años de 1995, 2005 y 2015.

Los datos de entrenamiento entraron a la red neuronal descrita en la sección 3.2 utilizando distintas configuraciones de número de neuronas y tasa de aprendizaje. Dichos valores oscilaron entre 4 y 7 neuronas y 60% y 80% en la tasa de aprendizaje. Se utilizó la técnica de Validación Cruzada para probar las posibles combinaciones de configuración con la intención obtener el modelo más simple posible pero que permitiera al modelo converger y entrenarse correctamente.

## 4.2 Validación de los Resultados Obtenidos

Una vez que el modelo terminó su entrenamiento y minimizó al error (definido como  $10^{-6}$ ), se utilizaron los datos de validación para calcular la eficiencia de nuestro modelo.

La validación de estos datos nos da como resultado un vector que contiene los parámetros normalizados de la distribución que se haya utilizado. En el caso de la distribución Gamma, se obtuvo un vector de tres filas y dos columnas, que representan los tres años de validación y los dos parámetros de la distribución Gamma. Para el caso de la distribución Log-Normal, se obtuvo un vector de tres filas y tres columnas, los cuales representan los tres años de validación y los tres parámetros de la distribución Log-Normal.

Para poder hacer la interpretación correcta de estos resultados fue necesario hacer una desnormalización de los mismos ya que introducir variables normalizadas a las distintas distribuciones no nos permitiría hacer una comparación válida con los datos originales. Una vez hecho esto se procedió a hacer el cálculo y evaluación del error.

El error se midió calculando la diferencia de áreas de curvas como se describió en el capítulo 3. En donde se aproximó el valor de la integral de cada curva utilizando el método Newton  $1/3$ . Primero se calculó el área bajo la curva de los datos originales de la pirámide poblacional para los años de 1995, 2005 y 2015. Y por otro lado se calculó el área bajo la curva de los resultados desnormalizados obtenidos de la red neuronal. Se hizo la resta respectivamente de cada año y eso se tomó como error.

### 4.3 Configuración y Error Minimizado

La configuración que mejores resultados entregó para cada una de las distribuciones fue elegida de acuerdo al tiempo de entrenamiento, convergencia de datos, sencillez del modelo y minimización del error, tal como se describe en la tabla 4.1 para cada una de las distribuciones. Podemos observar que, para la distribución Gamma, la mejor configuración se obtuvo utilizando 4 neuronas y una tasa de aprendizaje del 0.6% y el error relativo más alto obtenido de entre nuestros tres años de test fue para el año 1995 de 0.67907907. En la distribución Log Normal configuración que mejor funcionó fue 6 neuronas y 0.7% tasa de aprendizaje. Obteniendo un error máximo también en el año de 1995 de 0.2099037.

Configuración	Gamma	Log Normal
Neuronas	4	6
Tasa de aprendizaje	0.6	0.7
Error	0.67907907	0.2099037

Table 4.1: Tabla de Resultados de acuerdo a las mejores configuraciones probadas con cada una de las distribuciones.

Una vez obtenidos los errores relativos, se graficaron los datos originales y los datos obtenidos por la red neuronal. La comparación de los datos de test se pueden apreciar en las figuras 4.2, 4.1 y 4.3

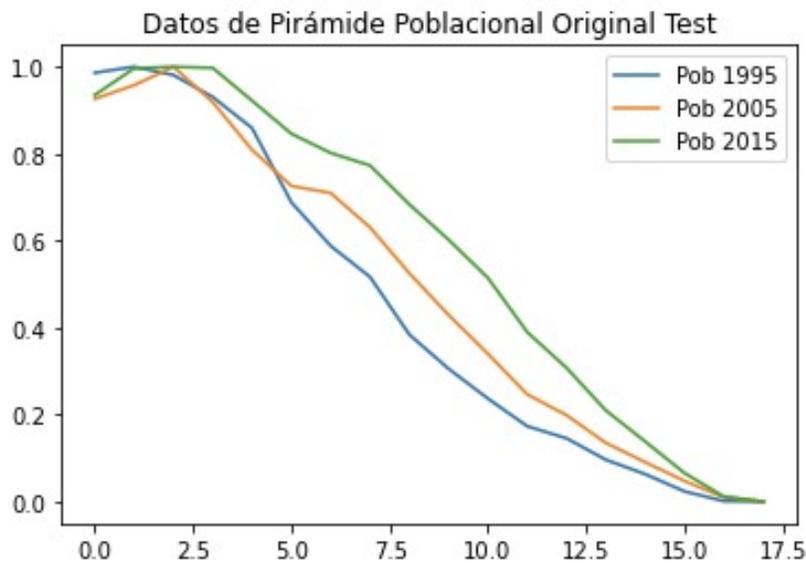


Figure 4.1: Datos originales de las pirámides Poblacionales normalizadas para la comparación contra las aproximación del modelo de los años 1995, 2005 y 2015.

Datos de Pirámide Poblacional Ajustada con Red Neuronal Test Log Normal

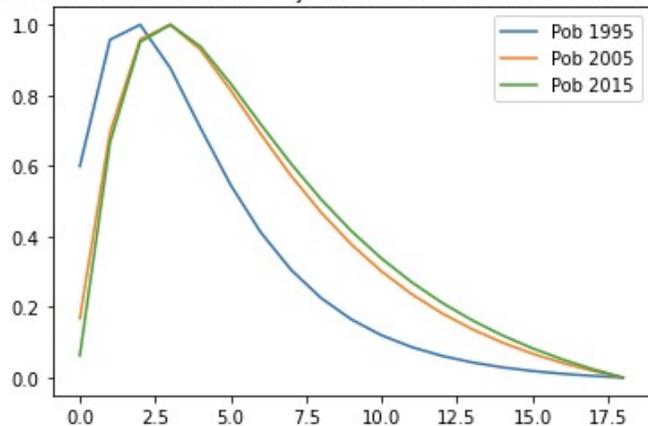


Figure 4.2: Pirámide poblacional normalizada obtenida de la Red Neuronal con aproximación Log Normal para los años de test 1995, 2005 y 2015

Datos de Pirámide Poblacional Ajustada con Red Neuronal Test Gamma

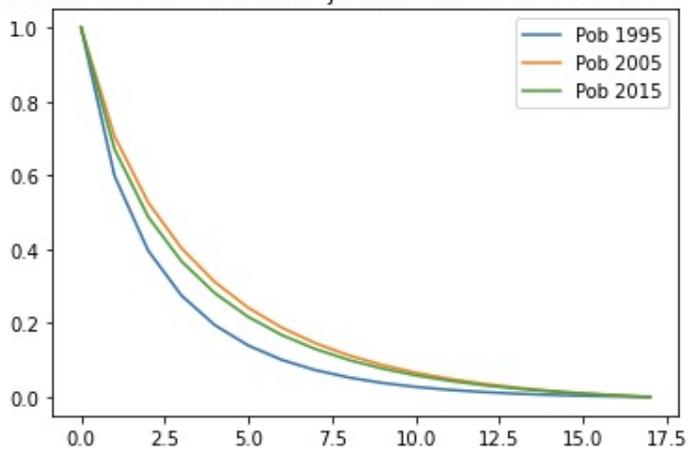


Figure 4.3: Pirámide poblacional obtenida de la Red Neuronal con aproximación Gamma para los años de test 1995, 2005 y 2015

#### 4.4 Proyección del modelo

Una vez obtenido los resultados de la tabla 4.1 y habiendo minimizado el error para cada configuración, estamos seguros de que tenemos una red neuronal entrenada que nos da resultados bastante precisos. Nos pareció interesante utilizar los datos que ya teníamos nuestras nueve variables de entrada para hacer una proyección hasta 2030. Esto se hizo con la intención de meterlas a la red neuronal previamente entrenada y obtener una idea de cómo se vería la pirámide poblacional en el año 2030.

Se hicieron extrapolaciones al año 2030 utilizando las regresiones con menor ajuste a los datos de las mencionadas en el capítulo 2, esto para evitar sobre ajuste en esta proyección. Las extrapolaciones fueron posibles ya que las tendencias de dichas variables fueron bastante lineales y simples de proyectar.

Se normalizaron los datos de las nueve variables proyectadas para 2030 y se introdujeron a la red neuronal. En las figuras 4.4 y 4.5 se muestra el resultado obtenido después de la desnormalización de las salidas de la Red Neuronal y la gráfica que dichos parámetros generan para cada una de las distribuciones probadas.

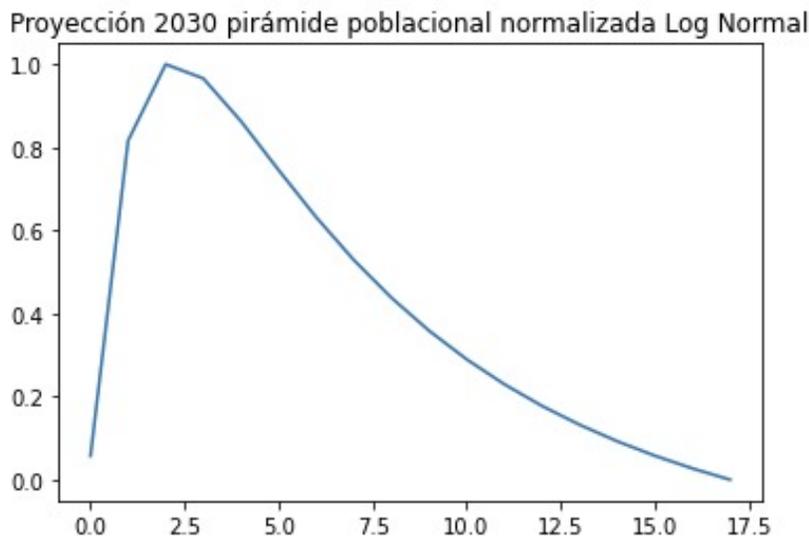


Figure 4.4: Gráficas de proyección de la pirámide poblacional para 2030 utilizando la distribución Log Normal.

#### 4.5 Análisis de influencia de las variables de entrada

Para concluir con este trabajo, se revisaron los pesos obtenidos después del entrenamiento de la red neuronal con cada una de las distribuciones propuestas. Se obtuvo la norma de cada uno de los pesos obtenidos

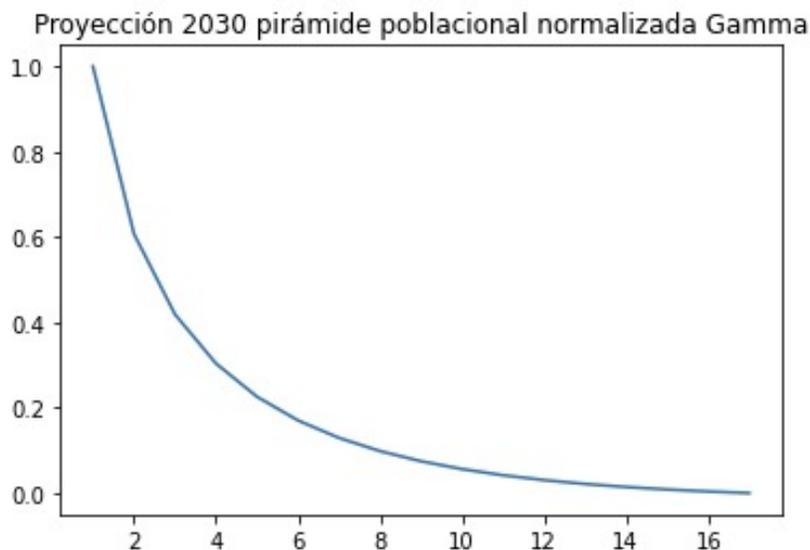


Figure 4.5: Gráficas de proyección de la pirámide poblacional para 2030 utilizando la distribución Gamma.

para cada variable. Los resultados se muestran en las tablas 4.2 y 4.3 y nos permite conocer la influencia o significancia que tuvo cada variable en el modelo. Es decir, cuál de las variables sociales, ambientales o económicas tienen más peso en la variación de la forma de la pirámide poblacional mexicana.

Ajustada con Distribución Gamma 4 Neuronas y 0.6 Tasa de Aprendizaje. Variable	Norma de los pesos obtenido durante el entrenamiento
Ingreso Corriente Total sobre Total Hogares	1.25454373
Gasto Corriente Total sobre Total Hogares	1.09250174
Temperatura Media Mensual Mundial	1.08551093
Proporción de Mujeres Estudiando sobre Total de Población.	1.03883261
Proporción de PEA femenina sobre PEA Masculina.	1.18034533
Proporción de PEA femenina sobre PEA Total.	0.95281612
Proporción de Divorciados.	1.108177472
Proporción de Solteros.	1.12191169
Hacinamiento Promedio de ZM.	0.67444722

Table 4.2: Tabla de la norma de los pesos obtenidos de la red neuronal por variable utilizando la distribución Gamma.

En la tabla 4.2 se puede apreciar como, para la distribución Gamma, la norma de los pesos de la mayoría de las variables son mayores a 1, excepto por las variables de Proporción de PEA femenina sobre PEA Total y Hacinamiento Promedio de ZM, las cuales obtuvieron 0.95 y 0.67 respectivamente. Por otro lado en la tabla 4.3 se muestra que, en la distribución Log Normal, la norma de los pesos para todas las variables es superior a 1, lo que nos dice que todas las variables son significativas para la proyección de la pirámide poblacional.

Ajustada con Distribución Log Normal 6 Neuronas y 0.6 Tasa de Aprendizaje. Variable	Norma del Peso Obtenido durante el Entrenamiento
Ingreso Corriente Total sobre Total Hogares	1.94434013
Gasto Corriente Total sobre Total Hogares	1.07926721
Temperatura Media Mensual Mundial	1.53300159
Proporción de Mujeres Estudiando sobre Total de Población.	1.38187821
Proporción de PEA femenina sobre PEA Masculina.	1.34160093
Proporción de PEA femenina sobre PEA Total.	1.41819286
Proporción de Divorciados.	1.50817954
Proporción de Solteros.	1.38057596
Hacinamiento Promedio de ZM.	1.10314218

Table 4.3: Tabla de la norma de los pesos obtenidos de la red neuronal por variable utilizando la distribución Log Normal.



## 5 Conclusiones

La distribución Log-Normal mostró un mejor ajuste a los datos de la pirámide poblacional, tomando en cuenta que el error en los datos de test fue del un 20.99% comparado con el error de la distribución Gamma, el cual fue de 67.90%

Los datos proyectados para 2030 representados en las gráficas 4.5 y 4.4, muestran una tendencia parecida a la expectativa y proyecciones que INEGI y CONPO muestran en sus publicaciones. Sin embargo, cabe remarcar que la pirámide proyectada por la distribución Log-Normal nos parece la más cercana a la realidad de acuerdo a las proyecciones de dichas instituciones. La pirámide proyectada por la distribución Gamma no se apega tanto, al menos no a simple vista.

A futuro, parece interesante trabajar para conseguir los datos de algunas de las variables que no pudieron ser incluidas en el modelo. Las variables que consideramos más interesantes son “Libertad sexual” y “Anticonceptivos y Aborto”, ya que aparentemente son viables de ser obtenidas.

Además, se pudo notar que la distribución Gamma hace una aproximación bastante certera de los primeros años (antes del año 2000), y la distribución Log-Normal hace una muy buena aproximación de las pirámides una vez que la pirámide comienza su transición hacia una pirámide estacionaria (en los años después del 2000). Sería conveniente hacer una mezcla de ambas distribuciones y observar si los errores disminuyen considerablemente.

Por último, este modelo se ajustó minimizando el error de la red neuronal hasta  $10^{-6}$ . Esto se vio sujeto a la capacidad de procesamiento y tiempo disponibles. En adelante, sería sumamente interesante utilizar procesadores más avanzados para minimizar el error y alcanzar un mejor ajuste.



## Referencias

Mtro. Leopoldo Núñez Fernández Mtro. Javier Suárez Morales Lic. Roberto López Esquinca Lic. Antonio Amerlinck Assereto Mtra. Miroslava Godínez Trejo Mat. Vicente Medina López Mtro. Rafael López Vega Mtro. Sergio I. Velarde. Act. Carlos Fuentes Villalba, Mtro. Juan Enrique García López. *Proyecciones de la población de México 2010 - 2050. Taller Interinstitucional para la Conciliación Demográfica y las Proyecciones de Población*, 1, 2012.

Jose Luis de la Fuente O'Connor. *Ingeniería de los Algoritmos y Métodos Numéricos*. Editorial Círculo Rojo, 2017. ISBN 978-84-9140-984-7.

Yadolah Dodge. *The Oxford Dictionary of Statistical Terms*. Oxford University Press, 2003. ISBN 0-19-850994-4.

José David Villanueva García. Redes neuronales desde cero (i) – introducción. <https://www.iartificial.net/redes-neuronales-desde-cero-i-introduccion>, October 2020.

2021 GISTEMP Team. Giss surface temperature analysis (gistemp) version 4. nasa goddard institute for space studies. <https://data.giss.nasa.gov/gistemp/>, July 2021.

Víctor Manuel García Guerrero. Pronóstico estocástico de la población mexicana utilizando modelos de datos funcionales. *18REALIDAD, DATOS Y ESPACIO, REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA*, 1, 2014.

INAFED. Conoce los beneficios de pertenecer a una zona metropolitana. <https://www.gob.mx/inafed/articulos/conoce-los-beneficios-de-pertenecer-a-una-zona-metropolitana>, June 2017.

INEGI. Comunicado de prensa núm 24/21. [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ResultCenso2020\\_Nal.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ResultCenso2020_Nal.pdf), 2021a.

INEGI. Fuente de datos inegi. <https://www.inegi.org.mx/datos/>, 2021b.

INEGI. Quiénes somos. [https://www.inegi.org.mx/inegi/quienes\\_somos.html](https://www.inegi.org.mx/inegi/quienes_somos.html), November 2021c.

Francisco Cruz Lario Nicolay Mena O'Meara, Eduardo Vicens Salort. Metodologías de inteligencia artificial para la toma de decisiones en la red/cadena de suministro en el contexto de incertidumbre. *3rd International Conference on Industrial Engineering and Industrial Management*, pages 1296–1305, 2009.

NIST/SEMATECH. *e-Handbook of Statistical Methods*. 2012. DOI: <https://doi.org/10.18434/M32189>.

OMS. Hacinamiento y limitación de la conducta. obtenido de universitat de barcelona., 2018.

Roland Pressat. *Demografía estadística*. Editorial Ariel, 1972.

UNFPA. Censos. <https://www.unfpa.org/es/censos>, November 2021.

Hamzan Wadi. *Learn from Scratch Backpropagation Neural Networks Using Python GUI MariaDB*. Turida Publisher, 2021.

# Index

\aprendizaje no supervisado, 42  
\aprendizaje supervisado, 42  
  
\Baby Boom, 37  
  
\censo, 15  
\clasificación, 42  
  
\distribución Gamma, 40  
\Distribución Normal Logarítmica, 41  
\diversidad sexual, 27  
  
\empoderamiento femenino, 25  
  
\error relativo, 45  
  
\función sigmoide, 43  
  
\hacinamiento, 31  
  
\INEGI, 15  
\IPC, 22  
  
\Migración, 17  
\método de Simpson 1/3, 44  
  
\normalización, 23  
  
\Pirámide Poblacional, 15  
\Población Económicamente Activa, 27  
  
\red neuronal Backpropagation, 43  
\regla del trapecio, 45  
\regresión, 42  
  
\Tasa de Crecimiento Demográfico, 16  
\Tasa de Mortalidad, 16  
\TMMM, 24  
  
\zona metropolitana, 32