



**UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MÉXICO
FACULTAD DE ECONOMÍA**



**“FACTORES ECONÓMICOS Y SOCIODEMOGRÁFICOS ASOCIADOS
CON LA PARTICIPACIÓN EN LA INFORMALIDAD LABORAL DE LA
POBLACIÓN MAYOR DE 60 AÑOS. MÉXICO 2018”**

TESIS

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADA EN ACTUARÍA

PRESENTA:

ELISA ALMAZÁN SALAZAR

ASESORA:

Dra. ANNEL HURTADO JARAMILLO

REVISORES:

Dra. LILIANA RENDÓN ROJAS

TOLUCA, ESTADO DE MÉXICO

2021

ÍNDICE

Introducción	4
Cap. I -- Seguridad económica en la vejez	8
1.1 Definición de seguridad económica en la vejez.....	9
1.2 Aspectos determinantes de la seguridad económica	11
1.3 Factores que determinan la posición económica en la vejez.....	13
1.3.1 Ciclo de vida individual y familiar.....	14
1.3.2 Antecedentes laborales.....	17
1.3.2.1 Ahorro.....	19
1.3.3 Biografías generacionales.....	20
1.3.4 Protección social.....	26
1.3.5 Pobreza en la vejez.....	29
1.4 Fuentes de la seguridad económica en la vejez.....	32
1.4.1 Participación económica	33
1.4.1.1 Envejecimiento activo	40
1.4.2 Seguridad social.....	42
1.4.3 Transferencias familiares	45
Cap. II -- Análisis de datos de respuesta binaria: Regresión logística	48
2.1 Definición de regresión logística y conceptos preliminares	49
2.2 Diferencias con el análisis de regresión lineal y análisis discriminante	52

2.3 Supuestos.....	55
2.4 Transformación <i>logit</i> y estimación de modelo logístico	60
2.5 Máxima verosimilitud	67
2.6 Interpretación de coeficientes.....	71
2.7 Selección de variables.....	73
2.8 Evaluación del modelo ajustado	78
2.8.1 Evaluación del ajuste global del modelo	78
2.8.2 Interpretación de resultados de las variables independientes.....	85
2.8.3 Análisis de residuos	87
<u>Cap. -- III Factores económicos y sociodemográficos que influyen en la participación de adultos mayores en la informalidad laboral</u>	91
3.1 Descripción de los factores seleccionados.....	92
3.2 Conceptualización y delimitación de la informalidad laboral.....	102
3.3 Aplicación del modelo a los datos	108
3.4 Análisis de resultados.....	114
<u>Conclusiones y recomendaciones</u>	131
<u>Bibliografía</u>	138

INTRODUCCIÓN

De acuerdo con Madrigal-Martínez (2010), durante la vejez aumentan las probabilidades de experimentar deterioro económico que ponen en riesgo la sobrevivencia y merman la calidad de vida de las personas mayores, además de que, lamentablemente, sus ingresos se ven relacionados con su nivel de inclusión en el sistema de protección social, lo que convierte a esta situación en un área de interés para la sociedad y el gobierno.

Los países en desarrollo se enfrentan al incremento en el nivel de participación de los adultos mayores, la oferta limitada de empleos para los grupos de edad avanzada y la insostenibilidad de los sistemas de seguridad social, aunado a una limitada cobertura universal (Garay & Montes de Oca, 2006). La estrategia que use el Estado se debe centrar en políticas públicas que contemplen aspectos de seguridad social, salud y asistencia social (Guzmán, 2002).

A nivel regional, en los próximos años, se espera un incremento sostenido de la proporción de personas de 60 años o más en América Latina y El Caribe. Entre 2000 y 2025 serán 88 millones de personas mayores y, para los siguientes 25 años el pronóstico tendrá un incremento de 86 millones (Guzmán, 2002); y si se añade que la mayoría que labora pertenece a la economía informal, con trabajos de ingresos bajos, inestables y con un posible despido en caso de situaciones adversas, el escenario es desalentador (Maldonado & Yáñez, 2014).

En México la situación no es distinta; para 2050, 17 de cada 100 personas serán adultos mayores, lo que representará 24.9 millones de habitantes del país. En líneas generales, incrementos en el sector de adultos mayores representa un desafío, no sólo en el aspecto demográfico, sino en materia de seguridad social, puesto que habrá una

mayor demanda de servicios de salud, recursos humanos, así como de movilidad y accesibilidad (CONAPO, 2020).

Bajo este contexto, las personas de edad se encuentran expuestas a vulnerabilidades ligadas a la salud, seguridad social, educación y, por supuesto, aspectos laborales. Es evidente un relego en ámbitos de trabajo por la preferencia del mercado a la Población Económicamente Activa (PEA) más joven, lo que los obliga a terminar en la informalidad como un medio de percibir ingresos y sentirse partícipes en la economía familiar y de las comunidades a las que pertenecen.

Es de suponerse que la informalidad laboral es un fenómeno que afecta no sólo a personas de edad, sino a aquellos extractos poblacionales que suelen ser excluidos del mercado formal por buscar mayor flexibilidad o peor aún, por no contar con habilidades académicas requeridas. Aunque pareciera ser una solución para la inclusión económica de forma independiente para aquellos que buscan satisfacer sus necesidades básicas, el precio es la calidad de vida e inestabilidad en ingresos.

Teniendo en consideración estas perspectivas, se tiene como objetivo general en la presente investigación identificar los factores asociados con la informalidad laboral de la población mexicana mayor de 60 años, específicamente en las áreas demográficas, económicas y sociales a los que se enfrenta este grupo etario. Asimismo, se planea exponer un modelo estadístico que estime la probabilidad de que un adulto mayor se encuentre trabajando en la informalidad.

Si bien ya existen trabajos de índole académico que versan sobre la participación de la población económicamente activa de edades avanzadas, son escasos lo que se centran en la informalidad. Según estadísticas sobre el día internacional de las personas

de edad, una gran parte de la población ocupada de 60 años o más labora por cuenta propia (49.6%), lo anterior no hace más que poner de relieve un fenómeno social que merece ser investigado (INEGI, 2018).

Para cumplir con los objetivos planteados, se divide el presente trabajo de tesis en tres capítulos, así como un apartado de conclusiones, con la finalidad de sustentar a través de éstos los argumentos para dar respuesta a las preguntas de investigación planteadas y poder sentar bases de apoyo para dar un dictamen sobre la hipótesis de que existen factores sociodemográficos y económicos que inciden en la participación en la informalidad laboral en la población mexicana mayor de 60 años.

Es por lo que en el primer capítulo se parte de una base teórica relativa a las tres fuentes que otorgan seguridad económica en la senectud, además de que se explican las causales que determinan la ausencia o presencia en el mercado laboral en el grupo poblacional de interés; en el mismo sentido, se pretende introducir a la seguridad económica como una meta, bajo los principios de solidaridad y justicia, para lograr una sociedad para todas las edades.

Lo anterior para lograr un acercamiento conceptual de la situación económica de las personas mayores, en términos monetarios y generacionales; así como examinar las fuentes que otorgan seguridad económica en edades avanzadas y al mismo tiempo, las causas que influyen en la participación laboral. Se toman en cuenta conceptos de la teoría del ciclo vital y del envejecimiento para conformar los factores que se relacionan con la seguridad económica en el grupo etario de interés.

El capítulo II comprende una descripción sobre la técnica multivariante de la regresión logística, enmarcándola siempre como una herramienta bastante versátil para

cuantificar fenómenos predictivos de carácter categórico; *grosso modo*, este método pretende calcular una probabilidad de ocurrencia y cómo fluctúa dentro una variable de una escala definida, todo en función de un conjunto de variables independientes.

Simultáneamente, se explica la diferencia que tiene respecto al análisis discriminante y regresión lineal múltiple en aras de demostrar que el modelo *logit* es ideal en la estimación de la probabilidad de que un adulto mayor se encuentre trabajando, así como su poder clasificatorio; de la misma forma, se exponen los supuestos estadísticos en los que se basa la construcción del modelo y las formas de evaluar su ajuste a nivel global como en sus coeficientes. Además, se incluyen métodos de selección de variables para encontrar el modelo más parsimonioso que represente mejor el fenómeno.

Para el capítulo III, se menciona la fuente de los datos, y se presenta una descripción estadística de la población de interés de acuerdo con características sociodemográficas y económicas que inciden en la participación del trabajo informal de los adultos mayores a través de la aplicación del modelo *logit*. Se cierra el capítulo con la descripción de las variables, se muestran los resultados del modelo, además de la interpretación de los resultados.

Finalmente, es en el capítulo IV donde se escriben las conclusiones, proposiciones y recomendaciones, haciendo énfasis en las condiciones laborales, de bienestar y salud. Además, se plantean posibles recomendaciones en materia de envejecimiento activo, que se alineen con lo establecido por la Organización Internacional del Trabajo (OIT).

CAPÍTULO I: SEGURIDAD ECONÓMICA EN LA VEJEZ

“Una vejez cómoda es la recompensa de una juventud bien empleada. En lugar de perspectivas tristes y melancólicas sobre la decadencia, nos da esperanza de una juventud eterna en un mundo mejor”.

Maurice Chevalier

En el presente capítulo se aborda la base teórica que sustenta las tres fuentes que otorgan seguridad económica en la vejez, para comprender, a través de una perspectiva integral, las determinantes que desembocan en la ausencia o presencia en el mercado laboral de la población mayor, a la vez que se propone a la seguridad económica como la herramienta que sirve de puente para incluir a aquellos que, por cuestión de edad o alguna otra característica, son olvidados.

La vejez se puede comprender como el fenómeno de índole social, complejo, multidimensional e intergeneracional que implica la culminación del ciclo vital de las personas y, por lo tanto, la pobreza, dependencia, enfermedad o discapacidad no pueden tener cabida en esta etapa. Un adulto mayor debe vivir en un estado de bienestar físico, psicológico y social, pero estos aspectos se ven determinados por sus condiciones de salud, situación económica y apoyo social (Aranibar, 2001).

El nivel de seguridad económica que tengan los adultos mayores es el resultado del diseño de los sistemas de protección social de hace tres o cuatro décadas y por supuesto, de las características del mercado laboral. Un análisis prospectivo en esta materia debe basarse en indicadores de cobertura de los sistemas de pensiones de los trabajadores y de las condiciones actuales que tengan las personas para acceder a jubilaciones y pensiones (CEPAL/CELADE, 2003).

1.1 Definición de seguridad económica en la vejez

Desde un punto de vista macroeconómico, la seguridad es la garantía y protección que la sociedad le concede a sus ciudadanos en aras de conservar sus derechos y propiedades. Trasladado al contexto de interés, la seguridad económica en la vejez es un derecho intrínseco en los principios de la protección social, junto con la atención básica a la salud y servicios sociales; resulta ser una dimensión clave para el envejecimiento activo y políticas sobre la vejez (Ivonet-Munder & Estrada-Hernández, 2017).

Entonces, la seguridad económica se concibe como la capacidad de acceder y usar independientemente una cantidad suficiente de recursos económicos regulares para asegurar una buena calidad de vida en la senectud; sin embargo, no sólo se queda en el plano material, sino que también influye de forma positiva en su bienestar subjetivo y autoestima, pues los coloca como ciudadanos con plenos derechos, capaces de tomar decisiones y desempeñar roles significativos (OISS, 2008).

En este sentido, la seguridad económica es el opuesto de la pobreza, ya que logra condiciones de envejecimiento dignas y seguras; aunque esto resulta ser más que una aspiración, tanto para quienes han estado contribuyendo toda su vida a su sistema de seguridad social como para quienes no (Huenchuan & Guzmán, 2007). Forma parte del bienestar y alude a una situación donde los individuos están casi seguros de satisfacer sus necesidades y deseos vitales (Nava-Bolaños *et al.*, 2016).

La seguridad económica en la vejez tiene mucho que ver con la calidad del mercado laboral y la fortaleza de los sistemas de pensiones, entonces hace sentido un análisis de las historias laborales de las personas, convirtiéndose en una herramienta

valiosa para la toma de decisiones dentro de las políticas públicas, sobre todo en lo que comprende los sistemas de pensiones para que se asegure estabilidad económica para la población mayor (Rubalcava-Montero, 2019).

En el tercer trimestre del 2020, de cada 100 mujeres, 50 se encuentra en el sector informal, marcando una brecha importante comparándola con los hombres, quienes, a pesar de tener prácticamente la misma escolaridad, tienen una ocupación mucho menor en dicho sector; aunado a esto, 15.54% de los adultos de edad con estudios profesionales se encuentra laborando en ocupaciones no profesionales. Finalmente, 12.57% está disponible para trabajar, sin expectativas de inserción laboral (**TABLA 1.1**).

TABLA 1.1 Indicadores de género para la población 60+, 2020

	III TRIMESTRE 2020		
	Total	Hombres	Mujeres
Promedio de años de escolaridad de la población económicamente activa	6.89	6.96	6.75
Promedio de años de escolaridad del jefe(a) del hogar	6.82	7.34	5.94
Porcentaje de ocupados con estudios profesionales terminados en ocupaciones no profesionales	15.54%	16.56%	13.81%
Porcentaje de población disponible para trabajar sin expectativas de inserción laboral	12.57%	15.03%	11.16%
Tasa de ocupación en el sector informal	36.67	31.09	49.13

FUENTE: Elaboración propia con base en INEGI. ENOE, 2020. Indicadores de género. Tercer trimestre.

Independientemente de la importancia del concepto, es evidente que encierra muchas implicaciones y conlleva gran complejidad, no sólo porque pone de relieve si los adultos mayores tienen recursos, sino que además si éstos son suficientes, si los reciben de forma periódica y si tienen plena disponibilidad de los mismos; empero, para ello se tiene que definir qué se considera suficiente, regular y bajo qué condiciones se afirma que el adulto mayor dispone de los recursos (Madrigal-Martínez, 2010).

1.2 Aspectos determinantes de la seguridad económica

Condiciones que han estado cobrando especial interés en las últimas décadas para múltiples países son las concernientes a la vejez, sobre todo desde las perspectivas económicas y demográficas, en el sentido que, a pesar de que el envejecimiento es un logro social resultado del aumento en la esperanza de vida, también plantea desafíos importantes en diferentes ámbitos, siendo uno de los más destacables y que poco a poco ha cobrado auge la seguridad económica.

Los estudios académicos de la seguridad económica para la población de edad avanzada son relativamente recientes y escasos; sus primeros antecedentes en habla hispana se remontan a las aportaciones de Pérez-Ortiz (1996) quien usa el concepto de posición económica para el diagnóstico de la situación y el papel que tienen en la sociedad española los adultos mayores, a partir de la estructura de los hogares en donde residen.

Pérez-Ortiz (1996) plantea una serie de cuestiones que funcionan como parteaguas para develar las dimensiones económicas de la población mayor; muestra que la posición económica de un adulto mayor depende de factores intrínsecos relacionados entre sí, mismos que retoman Huenchuan & Guzmán (2007) (ver **FIGURA 1.1**). Enfatiza la necesidad de relativizar los comportamientos de los que son considerados viejos y, por otro lado, esperando que la posición económica sea diferente para cada persona.

Huenchuan y Guzmán (2007) logran unificar y hacer una propuesta teórica-conceptual que toma como base nociones de la economía del envejecimiento, sociología de la vejez y teoría del ciclo vital; con esto pretenden analizar la constancia y recurrencia de los ingresos e incidencia de la pobreza en edades avanzadas para el estudio de la

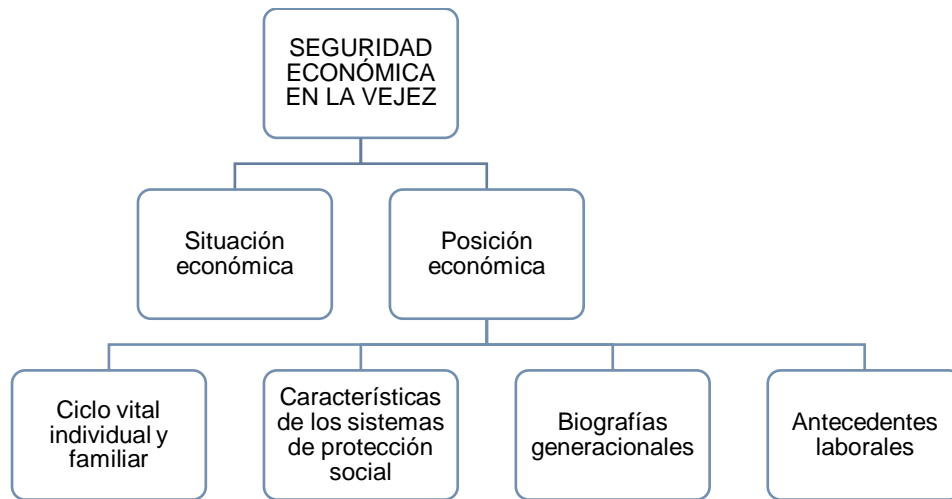
seguridad económica en adultos mayores; siendo sus aspectos determinantes: la situación y posición económica (**FIGURA 1.1**).

Logran cimentar tan bien el estudio en esta materia que artículos y trabajos de investigación de años posteriores usan como base sus aportaciones, convirtiéndolas en un referente obligado; asimismo, son generalizables incluso en regiones específicas como el Estado de México, tal como lo comprobaron Madrigal-Martínez y Millán-León en 2010, autoras que encuentran que la seguridad económica del adulto mayor está relacionada con su perfil sociodemográfico.

Retomando a la **FIGURA 1.1**, Huenchuan & Guzmán (2007), dividen en dos grandes ramas a las dimensiones que estudian a la seguridad económica en la senectud con el propósito de ahondar de forma integral y desde una perspectiva económica en la situación de los adultos mayores por ser un grupo que crece constantemente y que, además, tienen sus propias características sociodemográficas y de consumo que lo diferencian tajantemente de otros grupos de edad.

Primero, se define a la situación económica como la cantidad de bienes y servicios que se pueden adquirir, dada la disponibilidad de recursos (Nava-Bolaños *et al.*, 2016); misma viene fuertemente determinada por el poder adquisitivo, aunque sigue predominando cierta disparidad, además de ingresos bajos. En el mismo sentido, se encuentran los hábitos de consumo que el adulto mayor tiene y el apoyo familiar que puede recibir, (Sánchez-Vera, 2000) los cuales se tratan más adelante.

FIGURA 1.1 Determinantes de la seguridad económica



FUENTE: Elaboración propia con base en Huenchuan & Guzmán (2007).

Por otro lado, la posición económica es contrastable con los ingresos y bienes que tiene la población mayor respecto a otros grupos de edad y dentro del mismo (Nava-Bolaños *et al.*, 2016). De esta rama se desprenden cuatro factores que se encuentran fuertemente correlacionados y que se encuentran representados esquemáticamente en la **FIGURA 1.1**, mismos de los que se trata detalladamente en la siguiente sección.

1.3 Factores que determinan la posición económica en la vejez

Como se menciona al final de la sección anterior, la posición económica de los adultos mayores es medida a través de sus ingresos o bienes, debido a que en este punto del ciclo vital son más evidentes las desventajas que se acumularon a lo largo de sus vidas (Huenchuan & Guzmán, 2007). Para que los ingresos aseguren una buena calidad de vida en la vejez dependen, principalmente, de la edad, estado de salud, patrones de consumo previos y facilidades que proporcione el Estado (OISS, 2008).

Adentrarse en este tema no significa centrarse exclusivamente en el flujo y consistencia de los ingresos sujeta a la edad de los individuos (Pérez-Ortiz, 1996). Es por lo que, de acuerdo con Huenchuan & Guzmán (2007), la posición económica de la

población envejecida depende de cuatro factores complejamente interrelacionados y diferenciados entre sí: ciclo de vida individual y familiar, antecedentes laborales, biografías generacionales y protección social (véase **FIGURA 1.1**).

Antes de profundizar en cada uno de estos aspectos, es menester puntualizar que la posición es relativa, así como los comportamientos económicos de la población vieja; sobre todo, la conjunción de estos cuatro factores incide de formas distintas en cada individuo. Finalmente, la vida económica en la etapa final del ciclo tampoco es uniforme, pueden ocurrir mermas en el ingreso debido a situaciones propias de la vejez (Pérez-Ortiz, 1996).

1.3.1 Ciclo de vida individual y familiar

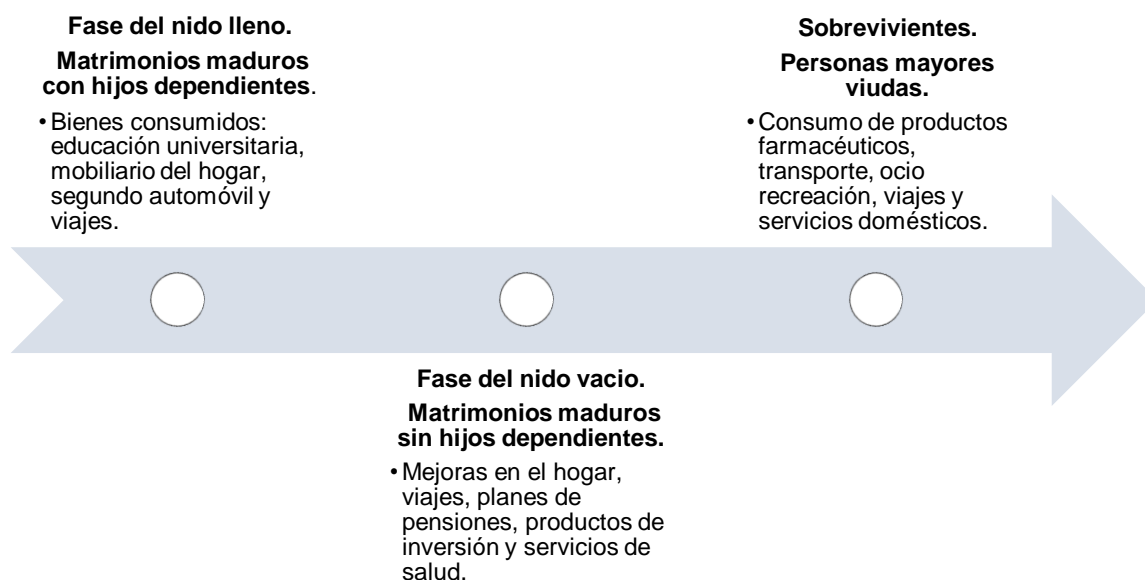
La posición económica de las personas no depende exclusivamente de la capacidad propia de generar ingresos, activos o riqueza, sino más bien es la conjunción de todos los miembros del hogar: está sujeta a la pertenencia de una unidad doméstica y de sus características; tampoco es el resultado de necesidades económicas propias o patrones de consumo, se apela más bien a la unidad familiar para la distribución de los recursos (Pérez-Ortiz, 1996).

La última etapa del ciclo vital alude a la situación en donde los individuos son mayores, así como las personas que los rodean. La inclusión de los familiares surge porque la interrelación de los ciclos de vida determina las cargas que llevan los adultos mayores y disponibilidad a ayudas que podrían recibir (Huenchuan & Guzmán, 2007). Ejemplo de ello sería el estatus de emancipación de los hijos y la consolidación de su nuevo hogar. Véase **FIGURA 1.2**.

El trabajo remunerado se consolida como el eje para garantizar las condiciones de vida, es el marcador del comportamiento, determina los ciclos vitales y las interacciones sociales. Los ciclos vitales y laborales en su conjunto establecen la forma de envejecer que va más allá de garantizar ingresos en la vejez, marca ritmos y horizontes cronológicos difusos porque la última etapa de vida se torna más imprevisible (Zuluaga-Callejas & Robledo-Marín, 2016).

La situación económica, familiar y de salud son las variables más representativas para las actitudes económicas; pero también lo son otras de índole subjetivo, como las características intrínsecas a la vejez. Anteriormente se menciona la importancia de la colectividad, aunque, contradictoriamente, a medida que se avanza en el ciclo, disminuyen las decisiones familiares conjuntas y prevalece la autonomía individual (Sánchez-Vera, 2000).

FIGURA 1.2 Ciclo Vital y consumo de los adultos mayores



FUENTE: Elaboración propia con base en (Sánchez-Vera, 2000).

Para la población mayor y para los que han alcanzado la edad legal vigente de jubilación, el factor de la edad está polarizado por otras circunstancias ajenas y que

deben considerar para comprender la posición económica (Pérez-Ortiz, 1996). Posiblemente, la tenencia de tiempo libre para el consumo es la característica que unifica al grupo poblacional mayor, y en un segundo plano queda la propiedad de una casa y hábitos de consumo controlados (Sánchez-Vera, 2000).

El retiro laboral supone un cambio en hábitos y comportamientos a medida que el adulto mayor se va amoldando a su tiempo libre y posibles actividades futuras. Pasa de ser un individuo activo en la sociedad, a un sujeto que recibe servicios específicos de acuerdo con su edad y va recluyendo en un cierto tipo de consumo, en el mismo sentido está situado en una red de intercambio no recíproco, lo que lo pone en una posición de dependencia (Aranibar, 2001).

Más allá del consumo o de una perspectiva económica, otra característica de interés para el desarrollo de la población mayor es la salud y es que, gracias al incremento en la esperanza de vida y a los avances científicos han disminuido las muertes por enfermedades virales y bacterianas; pero han ido en aumento los decesos por enfermedades crónico-degenerativas, las que implican inversiones cuantiosas por periodos prolongados (AMAFORE, 2015).

Sería un error percibir a la vejez como una etapa llena de pérdidas, porque también hay plenitud; sin embargo, esto depende de los recursos y oportunidades individuales y generacionales a las que se hayan expuesto a lo largo de su vida, sujeta a su condición y posición en la sociedad. Esto no hace más que agregar al sexo, clase social y etnia como factores conjugables con la edad, mismos que condicionan el acceso y goce de los ya mencionados recursos y oportunidades (Huenchuan, 2005b).

1.3.2 Antecedentes laborales

Las trayectorias laborales refieren a cotizaciones o aportaciones que se hayan acumulado al jubilarse y a la empleabilidad en la vejez (Huenchuan & Guzmán, 2007). Se busca determinar los movimientos en el mercado de trabajo para valorar los cambios que se producen e identificar los rasgos que los definen; por ejemplo, trayectorias que se hayan hecho en la formalidad, con estabilidad y contribuciones al sistema pensionario puede acrecentar las probabilidades de tener seguridad económica en la vejez (Rubalcava-Montero, 2019).

Es necesario entonces comprender como la historia o trayectoria laboral conforma la seguridad económica en la vejez, específicamente las condiciones laborales en las que se desarrolló el actual adulto de edad, si se empleó bajo contrato, si tuvo acceso a los beneficios de la seguridad social como aportaciones constantes a un sistema provisional o, por otro lado, cotizaciones acumuladas para la jubilación; además, se puede considerar su capacidad de emplearse en la vejez (Rubalcava-Montero, 2019).

Por tanto, el nivel socioeconómico de los adultos mayores se puede asociar con la seguridad económica y si existe una carencia en los ingresos por pensión, puede implicar la búsqueda de un empleo; hecho que se presenta con más frecuencia en los hombres, quienes por cuestiones generacionales son los proveedores y que, por lo tanto, se ven en la necesidad de continuar trabajando, a pesar de estar en la edad de retiro¹ (Garay & Montes de Oca, 2006).

¹ De acuerdo con ENOE 2020, en México, un porcentaje de la población adulta mayor de 60 años permanece en el mercado laboral, siendo 69.37% hombres y 30.63% mujeres (**TABLA 1.2**).

Por otro lado, al ser el salario la retribución del trabajo y, a su vez, como gran parte de las pensiones contributivas dependen del salario final, los ingresos de las féminas mayores son más bajos en este rubro ya que, al tener mayor esperanza de vida, tiene que ser distribuida equitativamente durante sus años de retiro, además, se añade el hecho que ellas tienen trayectorias más cortas, interrumpidas, con bajas, nulas o poco valoradas remuneraciones (Aranibar, 2001; Huenchuan, 2005b; AMAFORE, 2015).

Entonces, queda implícita la trascendencia de conocer las variables estratificadoras, sobre todo de aquellas donde el género sea determinante para la posición económica en función de la trayectoria laboral de las personas; previamente se estipula la importancia que tiene para los hombres el trabajo y transferencias fruto de la seguridad social y, para las mujeres, representa más el capital social y familiar que hayan logrado en su vida (Huenchuan, 2005b).

Por consiguiente, las desigualdades del mercado laboral tienen una fuerte repercusión en los sistemas de pensiones, en especial en términos de cobertura y beneficios. Es necesario tomar una perspectiva de género, no solo para el referido sistema, sino para todo el sistema de protección social y es que las mujeres han tenido, y posiblemente tengan, menor participación laboral e ingresos y como resultado, menor ahorro provisional, prestaciones y garantías públicas² (AMAFORE, 2015).

Además de la variable sexo, es importante recalcar la influencia que el ahorro tiene sobre los ingresos en la vejez, sobre todo porque sigue siendo un reto para las

² La manera en cómo se han organizado las actividades de hombres y mujeres es con base en la separación y la jerarquía, sobra decir que la participación histórica de los hombres está dirigida a lo público, mientras que las mujeres en lo doméstico; se ha buscado deslindar esta última relación con la idea de que no existe ninguna predisposición biológica de que las mujeres tengan que ser responsables de tareas que no suelen ser reconocidas como trabajo (AMAFORE, 2015).

políticas públicas; aunado a que se sigue delegando a la jubilación como una responsabilidad del Estado o familia. El ahorro es indispensable porque se transforma en ingreso futuro adicional a través de la acumulación de capital y que resulta en un complemento a las entradas de dinero usuales (AMAFORE, 2015).

1.3.2.1 Ahorro

El ahorro juega un papel importante para el retiro, sobre todo en países con sistemas de capitalización individual. Es más, la toma de decisiones respecto a la acumulación de riqueza y cómo se acumula o elige el momento del retiro laboral y su efecto en el monto de la pensión son imprescindibles para prever la jubilación; el ahorro parece ser clave en los sistemas de contribución definida y esta anticipación financiera permite alcanzar cierta seguridad económica en la vejez (AMAFORE, 2015).

La teoría del Ciclo Vital³ propone a los individuos como planificadores de su consumo y ahorro durante largos periodos con el fin de distribuir el consumo en sus vidas. Asimismo, parte del supuesto de que las personas eligen un estilo de vida estable, entonces, contribuye a relacionar el consumo y ahorro con la distribución de la población por edades, siendo su hipótesis central que los individuos ahorran en mayor medida para financiar su vejez (Serrano-Espinosa & León-Andrade, 2015).

Incertidumbres como la reducción temporal del ingreso debido al desempleo, la presencia de alguna enfermedad o discapacidad, provocan que los individuos optimicen la distribución de su consumo a lo largo de sus vidas (Lera-López, 1997) y busquen ahorrar

³ Propuesta por Franco Modigliani. Premio Nobel de Economía 1985. El modelo del ciclo vital de los ahorros explica el comportamiento de los ahorros personales, como los ingresos de los agentes económicos suelen subir en la juventud, alcanza su pico en la madurez y empieza a descender hasta llegar a niveles muy bajos en el retiro y el deseo de mantener cierto tipo de consumo, obliga a los jóvenes a ahorrar para en la senectud comenzar el desahorro (Ruiz-Durán, 2003).

para protegerse de estas contingencias, cuyo monto e intensidad varían de acuerdo con la edad, situación particular, carácter y situación socioeconómica. Sólo se reduce cuando existen mecanismos de protección social (Arriaga, 2010).

Como la seguridad social influye en las dinámicas económicas, especialmente en el empleo, ahorro privado y redistribución del ingreso; es objeto de interés analizar la asociación entre el ahorro y la seguridad social. El primer aspecto es una fuente de financiamiento para la inversión, consolidándolo como un deseo de seguir consumiendo en la vejez, por lo que la existencia de seguridad social tendría que reducir el ahorro⁴ (Moreno *et al.*, 2000).

Cada vez se hace más necesario promover acciones que anticipen y contemplen aspectos previsorios que permitan construir mejores condiciones de vida. El bienestar financiero está ligado con el personal, social y familiar y como los sistemas de pensiones no garantizan estabilidad económica una vez que la población se retira laboralmente, se enmarca la necesidad de fomentar el ahorro voluntario como un complemento de los gastos en la vejez (Serrano-Espinosa & León-Andrade, 2015).

1.3.3 Biografías generacionales

Las biografías generacionales aluden a cómo ha ido envejeciendo la población mayor (Huenchuan & Guzmán, 2007). Entonces, es menester puntualizar el crecimiento que ha tenido el grupo de edad porque potencialmente puede provocar un desequilibrio en el sistema de pensiones, además obliga a la población próxima a entrar a la vejez a

⁴ Para ilustrar mejor lo que se plantea, Moreno *et al.*, (2000) proponen un ejemplo: un agente económico que sabe que va a tener cierta cantidad de beneficios que le permitan continuar con su mismo nivel de consumo, entonces no habrá un incentivo de ahorrar durante su edad productiva; por otro lado, un agente económico que sabe que tendrá pocos beneficios durante su retiro, se verá obligado a incrementar su ahorro.

permanecer en el mercado laboral por razones de subsistencia lo que puede representar un problema en materia económica y social.

La transición demográfica está tornando a la población más longeva, prueba de ello es que, en México, al inicio del milenio las personas de 60 años o más representaban 9.23% de la población total, diez años después tuvo un incremento de casi un punto porcentual (0.97%), para finalmente terminar en 13.44% en 2020. Resalta, además, el hecho que la población mayor presenta una tendencia femenina. Véase **GRÁFICA 1.1**.

En el mismo sentido, esta transición se refleja en las pirámides poblaciones de 2000 y 2020 (**GRÁFICA 1.2**); en primer instancia, la pirámide del año 2000 tiene una gran concentración de población joven en la base que se va desvaneciendo según se avanza hacia la cúspide de los grupos de edad, este tipo de pirámide es común en países con alta natalidad, esperanza de vida baja y alta mortalidad, como lo era México hasta ese entonces.

GRÁFICA 1.1 Población nacional por grupo quinquenal de edad según sexo, 2000 a 2020



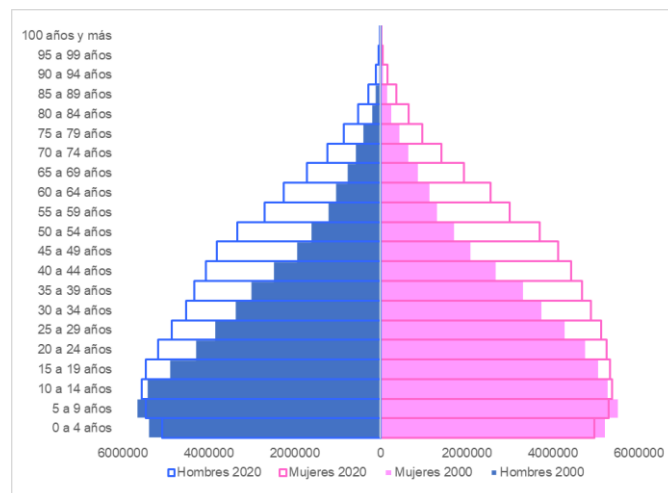
FUENTES: Elaboración propia con base en Instituto Nacional de Estadística y Geografía (INEGI). XII Censo General de Población y Vivienda 2000, Censo de Población y Vivienda 2010 y Encuesta Nacional de Empleo (nueva versión) 2020.

NOTAS: Para 2000, la información está referida al 14 de febrero.
Para 2010, la información está referida al 12 de junio.
Para 2020, la información corresponde al tercer trimestre.

Para 2020, se ve un cambio drástico en el aspecto demográfico, pues en la forma de la pirámide se muestra una base con ligeramente menos población respecto con los grupos quinquenales intermedios, mientras que en la cumbre existe una proporción importante de personas, visualmente notable, lo que parece indicar un descenso importante en la natalidad, tasas de mortalidad controladas y una esperanza de vida más alta.

La presencia de personas jóvenes (menores de 15 años) disminuye si se le compara con la población de hace 20 años, continua una buena proporción de personas en edades jóvenes, aunque también se incrementa en edades adultas avanzadas. Se hace más notorio un ensanchamiento de la cúspide, lo que hace más evidente el envejecimiento de la población; por tanto, es lógico pensar que en los próximos años existirá un contingente mayor de población de 60 años y más.

GRÁFICA 1.2 Población total por grupo quinquenal según sexo, 2000 y 2020



FUENTES: Elaboración propia con base en INEGI. XII Censo General de Población y Vivienda 2000 y Censo de Población y Vivienda 2020.

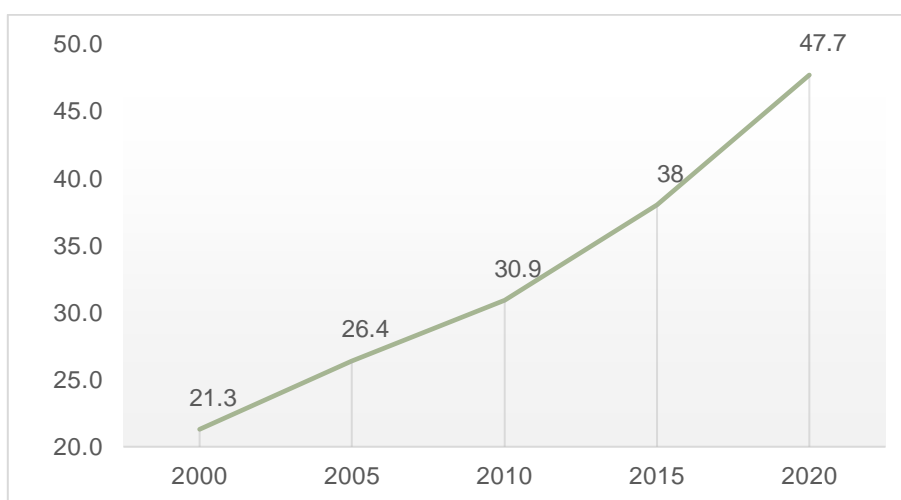
NOTAS: Para 2000, la información está referida al 14 de febrero. Para 2020, la información está referida al 15 de marzo.

Asimismo, el país está en un punto en el que cuenta con un enorme potencial productivo para la economía y, al mismo tiempo, se encuentra en un momento adecuado

para hacer frente a sus desafíos demográficos. Actualmente, el índice de envejecimiento es de 47.7%, es decir, hay aproximadamente 48 personas adultas mayores por cada cien niños y jóvenes de entre cero y 14 años; indicador que en el año 2000 era poco menos de la mitad. Véase la **GRÁFICA 1.3**.

Por otro lado, aun teniendo características específicas, la población mayor tiene rasgos culturales y generacionales que sesgan su cultura económica y los predispone a cierto tipo de consumo, mismos que los hace tendientes a la subsistencia, al ahorro tradicional y a la contracción y control de gastos, inclusive por arriba de su poder adquisitivo real (Sánchez-Vera, 2000).

GRÁFICA 3. índice de envejecimiento poblacional, en porcentaje, de 2000 a 2020



FUENTE: Elaboración propia con base en INEGI. XII Censo General de Población y Vivienda 2000, II Censo de Población y Vivienda 2005, Censo de Población y Vivienda 2010 y Censo de Población y Vivienda 2010.

NOTAS: Para 2000, la información está referida al 14 de febrero.

Para 2005, la información está referida al 17 de octubre.

Para 2010, la información está referida al 12 de junio.

Para 2015, la información está referida al 15 de marzo.

Para 2020, la información está referida al 15 de marzo.

Un caso particular es la situación económica de los adultos que participaron en el milagro mexicano porque su vida productiva sucedió cuando el país tuvo un crecimiento de tres décadas; lo que deja una incertidumbre para futuras generaciones sobre la carga de la vejez, aquellas que han crecido en contexto de crisis económicas y la tendencia es

la reducción del número de hijos, cuyas educación y empleabilidad digna son restringidas (Pedrero-Nieto, 1999).

En el contexto matrimonial, la posición de cada cónyuge está en función de los ingresos que aporta, y esto no hace más que poner en desventaja a la mujer mayor, pues tienen mayor dependencia económica, situación que se traslada en su momento de edades activas en bajas tasas de ocupación. Otra vertiente de mucho peso es el aspecto cultural, donde la mujer administra el dinero que recibe de su pareja, pero que, irónicamente, las decisiones económicas son muy masculinas (Sánchez-Vera, 2000).

Las mujeres de edad son las que tienen un espectro biológico muy fuerte sobre sus espaldas resultado del aumento en la esperanza de vida: la viudez, que se percibe como un descenso social y que va más allá de la pérdida de la pareja, significa la pérdida de una entidad independiente. Por otro lado, un aspecto cultural que afecta a la población femenina mayor es que puede verse relegada socialmente por estereotipos sexistas y por situaciones asociadas con la edad (OIT, 2002).

Específicamente, las mujeres adultas mayores siguen teniendo un rol doméstico **(TABLA 1.2)**, suelen ser señaladas como el principal recurso de mujeres jóvenes trabajadoras y se consolida esta ayuda intergeneracional como una solidaridad femenina, sujeta a características de consanguineidad y género. Por su parte, los abuelos contribuyen en situaciones concretas que no interfieren demasiado en sus hábitos y costumbres (Muñoz-Cobos & Espinosa-Almen, 2008).

En México (2020), con estas situaciones de género se comprueba que la actividad más frecuente de las mujeres de edad es el quehacer doméstico (90.02%), mientras que para sus pares masculinos solamente 9.98%. Solamente 32.65% de las mujeres adultas

mayores se encuentra pensionada y jubilada, situación en la que está 67.35% de los hombres del mismo grupo de edad. Ver **TABLA 1.2**.

Además, las adultas mayores que viven solas, en zonas rurales o sin ingresos garantizados las que son particularmente vulnerables en lo que compete en seguridad económica. En adición, son más sensibles aquellas que no tienen bienes, con ahorros escasos o nulos, sin pensión o algún sistema de protección social, las que no tienen hijos ni familiares, son las que se enfrentan a un futuro incierto y tienen probabilidades de caer en la indigencia (Zuluaga-Callejas & Robledo-Marín, 2016).

TABLA 1.2 Condición de actividad de la población adulta mayor por sexo, 2020

	PEA 60+	PNEA ⁵ 60+				
		Quehaceres domésticos	Pensionados y jubilados	Con alguna discapacidad permanente	Otros no económicamente activos	Estudiantes
TOTAL, 60+	28.608%	34.985%	20.659%	2.139%	13.592%	0.016%
HOMBRES	69.368%	9.980%	67.347%	43.573%	56.154%	66.619%
MUJERES	30.632%	90.020%	32.653%	56.427%	43.846%	33.381%

FUENTE: Elaboración propia con base en INEGI. Encuesta Nacional de Ocupación y Empleo (ENOE) 2020. Tercer trimestre.

Entonces, la relación entre el sexo y el envejecimiento reside en comprender como se conectan con el poder, privilegios y bienestar en la sociedad. La construcción social del género es distinta en cada etapa del ciclo vital y obviamente depende de la edad cronológica, fisiológica y social para los hombres y mujeres; es inherente al paso del tiempo y, por supuesto, de los acontecimientos que suceden en las etapas de la vida (Huenchuan, 2005b).

Otro problema que ha surgido en los últimos años es que la organización familiar ha cambiado, en el sentido que las parejas y vida conyugal están tendiendo a ser

⁵ Población No Económicamente Activa.

inestables, lo que pone en jaque a los sistemas de protección social, ya que casi todos fueron diseñados y construidos a inicios y transcurso del siglo XX, cuando la estructura familiar y laboral, tanto de hombres y mujeres, correspondían a patrones culturales distintos (Bertranou, 2005).

1.3.4 Protección social

La protección social fue un concepto revolucionario y comprensivo de la seguridad social cuando se presentó en 2005 por la OIT; entendiéndose como la totalidad de intervenciones de instituciones públicas y privadas cuyo propósito es apaciguar una serie de riesgos y necesidades para los hogares de los individuos. A su vez, esto se basa en la conjunción contemporánea de tres pilares (Bertranou, 2005):

- Acceso a bienes y servicios básicos indispensables.
- Protección y prevención.
- Promoción de oportunidades.

Dentro de los programas de protección social se entiende a la cobertura como la proporción de personas que reciben un beneficio (población objetivo). Consta de dos fases: la primera, el momento en que el trabajador contribuye al sistema y acumula sus beneficios; mientras que la segunda alude a la recepción de los derechos monetarios, que es cuando los individuos llegan a edades avanzadas (60 años o más), no es más que la cobertura de la población envejecida (Rofman, 2005).

En general, los sistemas de seguridad social son un espejo de las oportunidades y amenazas que tiene la población para un ingreso sustituto digno en la vejez y, a su vez, de acceder a la atención médica institucional (Huenchuan & Guzmán, 2007). Se hace hincapié en las características de los sistemas de protección social mientras se

avanza en el tiempo y en especial, para el momento en que la población llega a la vejez (Pérez-Ortiz, 1996).

Sin embargo, incluso con la relación positiva entre la seguridad social y económica en la senectud, no significa que se posean ambas, porque se puede acceder a la seguridad social y no disponer de seguridad económica en la vejez (Ivonet-Munder & Estrada-Hernández, 2017). Ejemplo de ello es que, de forma generalizada, las condiciones de seguridad económica en América Latina y El Caribe son raquílicas y poco equitativas (CEPAL/CELADE, 2003).

Como ya se ha planteado de forma intuitiva, los sistemas de pensiones, tanto en acceso, cobertura y montos, tienen una relación con las condiciones de vida y bienestar de las personas mayores, pero, aun así, es poco posible aseverar una generalidad sobre los sistemas, sobre todo si se toma en cuenta las tendencias de algunas regiones geográficas de encaminarse a sistemas de capitalización individual, así como sus implicaciones en el ámbito fiscal (Aranibar, 2001).

Específicamente, en América Latina la transición a este tipo de sistemas inicia en 1981 con Chile; poco a poco se fue implementando la capitalización individual en el resto de los países, siendo el último país Panamá en 2008. Cada país tiene su propio organismo regulador e institución de capitalización individual que se encargan de su regulación y gestión; para el caso mexicano son la Comisión Nacional de Ahorro para el Retiro (CONSAR) y Administradora de Fondos para el Retiro (AFORE), respectivamente. Véase **TABLA 1.3**.

TABLA 1.3 Países de América Latina con Sistema de Capitalización Individual

País	Año	Organismo	Institución de Capitalización Individual
Bolivia	1997	Fiscalización y Control de Pensiones y Seguros	Administradora de Fondos de Pensiones
Chile	1981	Superintendencia de Pensiones	Administradora de Fondos para el Retiro
Colombia	1994	Superintendencia Financiera	Administradora de Fondos para el Retiro y Cesantía
Costa Rica	2000	Superintendencia de Pensiones	Operadoras de Planes de Pensiones Complementarios
El Salvador	1998	Superintendencia del Sistema Financiero	Administradora de Fondo de Pensiones
México	1997	Comisión Nacional de Ahorro para el Retiro	Administradora de Fondos para el Retiro
Panamá	2008	SIACAP	Administradora de Fondos para el Retiro
Perú	1993	Superintendencia de Banca, Seguros y AFP	Administradora Privada de Fondos para el Retiro
República Dominicana	2003	Superintendencia de Pensiones	Administradora de Fondos de Pensiones
Uruguay	1996	Superintendencia de Servicios Financieros	Administradora de Fondos de Ahorro Previsional

FUENTE: Elaboración propia con base en Federación Internacional de Administradoras de Fondos de Pensiones (FIAP), 2016.

La totalidad de estos países con este sistema tiene un esquema multipilar, adaptado por cada marco legal de acuerdo con las necesidades de su población; en líneas generales, consta de tres lineamientos básicos: pilar no contributivo financiado con presupuesto público, pilar obligatorio que usualmente incluye un programa de reparto y uno de capitalización individual y, por último, el pilar voluntario que busca incentivar fiscalmente el ahorro complementario (FIAP, 2016).

En cierto modo y como lo hace notar Bertranou (2005), se produce una paradoja de la protección; si bien, existe cierto porcentaje menos vulnerable, quienes acceden a más y mejor protección derivado de ciertos factores, pero principalmente producto del mercado laboral, donde únicamente se segrega positivamente a los mejores posicionados y por ende, son los que son favorecidos en relación con la cobertura. Entonces, una problemática es la baja extensión y calidad en la cobertura.

Merece la pena subrayar que el financiamiento de la protección social para la población mayor debe gestionarse a través de esquemas solidarios, pero que,

lamentablemente, no son fáciles de diseñar ni mucho menos de implementar en economías frágiles. Aun así, la alternativa acarrea un envejecimiento con mayores enfermedades no mortales, las que generan discapacidad, desembocando en una mayor carga económica y social para la familia (CEPAL/CELADE, 2003).

Se apela a los criterios de universalidad, solidaridad y eficiencia en la gestión financiera, dejando de favorecer a los estratos con mejor nivel socioeconómico o que hagan mayores aportaciones; como se pretende una vida digna en la vejez, se deben evitar desigualdades en los servicios de salud y seguridad social. Lo importante es cimentar un mínimo solidario, que sea independiente de la capacidad contributiva de las personas (CEPAL/CELADE, 2003).

1.3.5 Pobreza en la vejez

El estudio de la pobreza en la vejez se hace como un factor complementario de la seguridad económica, puesto que no figura como uno de los aspectos que determinan la posición económica propuestos por Huenchuan & Guzmán (2007) (ver **FIGURA 1.1**), aun así, tiene tal importancia que se hace referencia a este tema en esta sección.

Desde una perspectiva comparativa con otros grupos de edad, la población en edades avanzadas incide menos en la pobreza, pero, aun así, los porcentajes en que afecta a los adultos mayores son bastante altos (Nava-Bolaños *et al.*, 2016). Este menor acaecimiento se puede explicar porque se encuentran en la última etapa del ciclo económico familiar, la que inicia con la emancipación de los hijos y culmina con la muerte de los progenitores (Huenchuan & Guzmán, 2007).

Aun así, el mayor obstáculo para un envejecimiento decente y seguro es la pobreza⁶. Es bastante normal que los ingresos de los adultos mayores se reduzcan o que incluso dependan de las transferencias de carácter público o privado, es entonces cuando cobra sentido que la cobertura del sistema de pensiones y la generación de oportunidades laborales respondan a las características del grupo de edades avanzadas y sean relevantes (AMAFORE, 2015).

Es sabido que una población envejecida acarrea problemas serios en los sistemas de pensiones en materia de solvencia y sostenibilidad; situaciones que impiden que las personas cuenten con recursos económicos para solventar sus gastos en la etapa final del ciclo vital (CEPAL/CELADE, 2003); entonces la pobreza en la vejez supone mayor vulnerabilidad y acrecienta las condiciones de inseguridad económica que los adultos de edad raramente pueden solucionar (Huenchuan & Guzmán, 2007).

No es más que un resultado de desigualdad extrema. La pobreza en el grupo etario de interés está relacionada con ciertas fases de fragilidad en su ciclo vital, entonces, lamentablemente, la edad se convierte en el determinante de la fragilidad y gracias a esta variable, descienden bruscamente de un nivel de subsistencia a uno de pobreza con una aparente facilidad si se le compara con otras etapas de la vida (Huenchuan, 2005a).

Aun así, hay que tener cuidado con lo que caracteriza a la pobreza en la vejez, la muerte del cónyuge o de los hijos y el estatus de salud son caldo de cultivo para que las

⁶ Una persona se encuentra en un estado de pobreza cuando tiene al menos una carencia de tipo social, ya sea en rezago educativo, acceso a salud, alimentación y seguridad social, así como de calidad y espacios en la vivienda y, además, que su ingreso no sea suficiente para adquirir los bienes y servicios que requiere para satisfacer sus necesidades alimentarias y no alimentarias (CONEVAL (Consejo Nacional de Evaluación de la Política de Desarrollo Social), 2016).

personas caigan en la pobreza con mayor facilidad que en otros momentos de la vida. En el mismo sentido, se tiene que considerar que la posición económica de los adultos mayores está marcada por factores generacionales que favorecen su situación económica, haciendo difícil una homogeneidad para su estudio (Huenchuan, 2005b).

Las transiciones que hacen las personas hacia el retiro y viudez vienen con la reducción de ingresos y aumenta la probabilidad de la pobreza en hogares con adultos mayores, asimismo también influyen características individuales, generacionales y de la acumulación de activos, que quedan especialmente arraigadas por las condiciones de mercado, cobertura de la seguridad social y permanencia en la participación económica (Garay & Montes de Oca, 2006).

Factores como las características del empleo y del sistema de seguridad social vigente son lo que suelen estar presente en el estudio de la pobreza respecto con otros grupos de edad, no son útiles para las desigualdades dentro de una misma generación; si bien, son evidentes entre hombres y mujeres mayores, muchos de éstos son resultado de las condiciones de la actividad laboral de las mujeres o de limitaciones legales y normas de la seguridad social (Huenchuan, 2005a).

Es evidente que la pobreza en la vejez también tiene una fuerte marca de género; es más probable que las mujeres sean pobres por más tiempo durante su senectud que sus pares masculinos y, en el mismo sentido, es más fácil que una adulta mayor sea viuda a que se vuelva a casar (OIT, 2002), por lo que un factor que puede mermar el poder adquisitivo de la población mayor femenina es la viudedad (Sánchez-Vera, 2000).

Previamente se establece a la seguridad económica como un derecho humano fundamental que el Estado debe complementar de forma integral a través de programas,

sistemas y acciones de carácter político, social y económico que permitan a los adultos de edad asegurar su propia subsistencia y que también, ofrezcan las condiciones necesarias para salir de la pobreza a la totalidad de la población mayor (Cuentas, 2002).

Las estrategias de reducción de pobreza tienen que incorporar acciones que permitan el bienestar desde una perspectiva multidimensional, incluyendo a las personas mayores como a sus familias. Las políticas públicas necesitan fomentar la vida activa que logren el envejecimiento digno; el éxito de éstas puede desembocar en la disminución de la informalidad laboral o un aumento en la cobertura de pensiones no contributivas (AMAFORE, 2015).

1.4 Fuentes de la seguridad económica en la vejez

La inserción laboral de los adultos mayores no resulta una cuestión voluntaria, tiene el propósito de subsistencia con condiciones precarias, ya que la generalidad es que encuentren trabajos no remunerados donde carecen de cobertura médica y social (Garay & Montes de Oca, 2006). Entonces, los mecanismos que garantizan seguridad económica son la participación económica (mercado), la seguridad social (Estado) y apoyos familiares (familiar) (Huenchuan & Guzmán, 2007).

De esta triada generadora de ingresos para la población de edad, las primeras dos fuentes son de carácter formal, mientras que la última es informal. Los orígenes del dinero corresponden a salarios, remuneraciones de trabajo, pensiones, jubilaciones, seguros de salud y subsidios por discapacidad; por otro lado, los apoyos familiares comprenden ayuda monetaria, en especie o regalos que usualmente provienen de su círculo más cercano (Madrigal-Martínez, 2010).

La tenencia de activos se añade como una fuente de seguridad económica, en especial la vivienda propia. Asimismo, se plantea la inserción al mercado laboral de los adultos mayores como una estrategia de compensar la desprotección del Estado en lo que compete en seguridad social, sobre todo para aquellos que mientras estaban en edad productiva no pudieron reunir los requisitos necesarios para acceder a una pensión (AMAFORE, 2015).

Todas estas dimensiones se van acoplando a lo largo de la vida del individuo; pero no es una generalidad que se fortalezcan en todos los casos, ya sea por otros eventos vitales y generacionales. Una persona mayor tendrá cierta situación económica como resultado de su historia laboral y familiar, accesibilidad a la protección social, ahorros, bienes acumulados y características sociodemográficas; entonces cualquier desventaja en algún ámbito se reflejará en el nivel de seguridad económica en la senectud (Madrigal-Martínez, 2010).

1.4.1 Participación económica

Como es de esperarse, las personas mayores tienen una preferencia al ocio y en general, se encuentra que las personas con pensiones más cuantiosas tienden a jubilarse antes; sin embargo, la seguridad social y los planes de pensiones son los que determinan el comportamiento jubilatorio de las personas, y es que, generalmente, los incentivan para que continúen trabajando hasta edades avanzadas (van Gameren, 2008).

Es así como los niveles de participación económica en los adultos mayores suelen ser altos por la escasa cobertura en materia de pensiones, lo que, a su vez, condiciona su acceso a la salud y, por supuesto, termina provocando vulnerabilidad, no sólo en el aspecto monetario, sino también en los servicios médicos (Garay & Montes de Oca,

2006). Las jubilaciones, pensiones y trabajo se consolidan como las fuentes formales de ingresos más importantes durante la vejez (Huenchuan & Guzmán, 2007).

Elementos como la formación, escolaridad y adiestramiento son caldo de cultivo de las oportunidades económicas y capacidad de inserción en el mercado de trabajo, pero es la escolaridad la variable que sobresale e incrementa los requerimientos necesarios para funcionar en el referido mercado, ofreciendo ventajas y mayores oportunidades en el manejo de tecnologías entre edades más avanzadas (Nava-Bolaños & Ham-Chande, 2014).

En el mismo sentido, algunos indicadores no hacen más que poner de relieve cuales son las oportunidades laborales de la población adulta mayor según ciertas características; por ejemplo, la escolaridad máxima que tiene en conjunto este grupo sería análoga al primer año de secundaria, mostrando pocas diferencias entre hombres y mujeres, los años de escolaridad para jefes de hogar de edad avanza también ronda en los siete años de instrucción.

La participación económica en la vejez es un fenómeno impulsado por varias causas más allá de la calidad y extensión de la cobertura de los sistemas, como el entorno macroeconómico y desempeño del mercado de trabajo. En el mismo sentido, el estado de salud de las poblaciones de edad ha estado mejorando, es una generalidad que la expectativa de vida está en incremento y seguramente se mantendrá así en los próximos años (Bertranou & Velasco, 2006).

Independientemente del estado de salud que las personas mayores pudieran enfrentar, el papel que tengan en sus familias o su propia disposición por el trabajo, la mayoría sigue laborando como respuesta a una necesidad económica que los coloca en

condiciones poco favorables; situación que no hace más que romper el estigma de que la población mayor deja de trabajar por la disminución de sus obligaciones o que gozan de los beneficios de la seguridad social (Madrigal-Martínez, 2010).

Para ilustrar mejor esto, se explican a continuación los motivos de mayor permanencia o reinserción laboral de las personas que son exclusivos para el grupo poblacional de edades avanzadas, usando información de encuestas de hogar de ocho países de América Latina y El Caribe (Argentina, Brasil, Chile, Colombia, Costa Rica, Guatemala, México y Perú) cada uno con distintos niveles de desarrollo, mercados laborales y sistemas de jubilación (Weller, 2018):

- Mejores niveles de salud y mayor esperanza de vida estimulan el interés de los adultos mayores en seguir laborando.
- Escasez de oferta laboral e insostenibilidad de los sistemas de pensiones, tanto en los de reparto como los de ahorro individual; mismos que pueden fomentar políticas que incrementen la edad legal de jubilación, imponiendo una vida laboral más extendida.
- Disminución de brechas generacionales de participación y ocupación entre hombres y mujeres, al producirse un incremento gradual de las mujeres en actividades laborales, además de mayor esperanza de vida para ellas y altos niveles de viudez.
- Menor presencia de hogares multigeneracionales.

Estar vigente en sociedad para los adultos mayores se puede lograr a través de ingresos decentes recompensa de trabajos o por jubilación y, por supuesto, contar con la posibilidad de ser partícipes en la comunidad a través del empleo remunerado o

voluntario u otras actividades según sus cualidades y preferencias; es aquí donde nuevamente la protección social y el trabajo tienen un papel indispensable, sobre todo si se hace en condiciones de libertad, equidad, seguridad y dignidad (OIT, 2002).

En México, la participación económica que tienen los adultos mayores es de 28.61%; es decir, aproximadamente, 29 de cada 100 adultos mayores está en alguna actividad económica o en busca de alguna; en el mismo sentido, haciendo una desagregación sexo resulta que 43.31% de la población masculina de edad está en la referida situación, mientras que sus pares femeninas solamente 16.18%, según datos de la ENOE, 2020. Véase **TABLA 1.4**.

Parece menester conceptualizar a la informalidad desde la perspectiva de las condiciones y características de los trabajadores; entonces de acuerdo con Martínez-Restrepo *et al.*, (2015) la informalidad es aquella que contempla a todos los trabajadores que no son contratados bajo las normas legales en aspectos relacionados con la calidad del empleo; es el conjunto de actividades voluntariamente disimuladas a las autoridades con el propósito de no pagar impuestos a la renta o al valor agregado.

TABLA 1.4 Población Mayor Económicamente Activa y su participación económica, 2020

	Población total 60+	Población Económicamente Activa						Participación económica ⁷
		Total, PEA	Sólo PEA	PEA y estudian	PEA y quehaceres domésticos	PEA y apoyos al hogar	PEA y otras actividades	
TOTAL, 60+	17 025 218	4 870 564	931 067	22 953	3 286 741	514 032	115 771	28.608%
HOMBRES	7 801 805	3 378 603	908 448	18 836	1 839 417	496 579	115 323	43.305%
MUJERES	9 223 413	1 491 961	22 619	4 117	1 447 324	17 453	448	16.176%

FUENTE: Elaboración propia con base en INEGI. ENOE, 2020. Tercer trimestre.

⁷ Corresponde al grado de participación de la población en el mercado de trabajo, se calculó a través del cociente de la población económicamente activa de 60 años y más entre la población total del mismo grupo, se presenta en porcentaje.

La informalidad laboral tiene asociado un alto grado de precarización, grupos poblacionales considerados vulnerables son los más propensos a este tipo de empleo. Se entenderá como grupo vulnerable a aquel que por alguna característica demográfica, económica, social, política o cultural tiene mayor riesgo de que sus derechos sean violentados; siendo límites o inaccesibilidad para un trabajo decente un claro ejemplo de violación de derechos (Cuevas-Rodríguez *et al.*, 2016).

La generalidad es que los países en desarrollo, como México, suelen tener una población mayor que tiende a trabajar en categorías ocupacionales no asalariadas, dejándolos claramente en mayor desventaja que los jóvenes. Por otro lado, algo positivo de esta situación es la reducción relativa de horas de la jornada laboral que tienen los trabajadores de edad, en comparación con otros grupos de trabajadores (Bertranou & Velasco, 2006).

Como señalan Bertranou & Velasco (2006), las crecientes tasas de participación laboral entre los adultos mayores, sobre todo en mujeres, provoca que la mayoría de la población trabaje en la informalidad, lo que podría estar intrínsecamente relacionado con las reformas en los sistemas de pensiones en décadas pasadas que limitaron las condiciones de acceso a beneficios a edades más tardías o con una baja tasa de reemplazo, incentivando a los mayores mantenerse ocupados durante más tiempo.

En el grupo de interés, existen estimaciones que ponen de relieve la ocupación informal según el grupo de edad. En 2020 y para la población mayor, poco más de dos millones de personas trabajan en el sector informal, de las cuales 65.73% son hombres; asimismo, prevalece ocupación informal masculina en ámbitos agropecuarios y para las

mujeres es caso contrario, situaciones que no sorprenden por los roles de género que caracterizan a los adultos mayores (**revisar 3.3**). Véase **TABLA 1.5**.

En específico, la situación social de la población mayor mexicana se puede sintetizar como: a menor seguridad social, mayor participación económica y mayor marginación, por lo que la asociación negativa entre la participación económica y cobertura en el sistema de pensiones del Estado tiene una fuerte alusión con los grados de marginación que señala el Consejo Nacional de Población (CONAPO) (Garay & Montes de Oca, 2006).

Siguiendo con el punto anterior, en México las personas mayores pueden seguir trabajando o estar en alguna actividad económica dentro o fuera del mercado laboral (formal o informal), siendo más evidente en zonas menos urbanizadas (las poblaciones se consideran rurales con menos de 2 500 habitantes, según INEGI), en estas localidades, la población mayor suele seguir en actividades relacionadas a la subsistencia (véase **TABLA 1.5**), como ganadería y agricultura (Montes de Oca & Hebrero, 2006).

TABLA 1.5 Indicadores estratégicos de la informalidad laboral para la población 60+, 2020

	Población ocupada	Ocupación formal	Ocupación informal				
			Subtotal	Sector informal	Fuera del sector informal		
					Subtotal	Ámbito no agropecuario	Ámbito agropecuario
TOTAL, 60+	3 067 562	722 104	2 345 458	1 192 434	1 153 024	417 521	735 503
HOMBRES	2 080 140	538 355	1 541 785	640 022	901 763	220 460	681 303
MUJERES	987 422	183 749	803 673	552 412	251 261	197 061	54 200

FUENTE: Elaboración propia con base en INEGI. Encuesta Nacional de Ocupación y Empleo. Informalidad laboral. Indicadores estratégicos. Primer trimestre de 2020.

Entre más alto sea el grado de marginación, la informalidad laboral aumenta, confirmándose la hipótesis de que las personas en condición de pobreza e incapacidad

de incorporarse al mercado formal buscan opciones de sobrevivencia en los mercados laborales secundarios o informales. Esta dinámica conduce a un círculo vicioso que profundizan y acrecientan las condiciones de pobreza o precarizan las condiciones laborales (Cuevas-Rodríguez *et al.*, 2016).

La propensión a la actividad laboral de la población de la tercera edad y su relación con la protección social están acompañados con el nuevo paradigma asociado a la necesidad del envejecimiento activo, cuya polémica radica en saber si puede existir una reversión en esta tendencia y cuáles serían las políticas públicas necesarias para atender el proceso vía de extensión de las vidas laborales y restringir el temprano acceso a la seguridad social por problemas de financiamiento (Bertranou, 2005).

La tendencia del envejecimiento activo debe estar acompañado con el diseño de políticas públicas, incentivando la demanda de empresas por mano de obra de mayor edad, así como en los sistemas de pensiones. La disminución de horas trabajadas conforme se avanza en el ciclo vital parece ser un asomo de flexibilidad para las necesidades de los mayores, pero aun así falta mucho por hacer en términos de políticas para proteger a la población de edad (Bertranou & Velasco, 2006).

1.4.1.1 Envejecimiento activo

Tradicionalmente se ha pensado que la salida laboral de los adultos mayores es para darle espacio a los más jóvenes, lo que ha dejado de ser completamente cierto por la actual fragmentación del mercado laboral; es evidente la necesidad de estudio y diseño de políticas públicas que ayuden a mejorar la situación laboral y la protección social; es menester un enfoque comprensivo que fortalezca la protección de los ingresos y la potencialidad productiva de la fuerza laboral en la vejez (Bertranou, 2005).

La experiencia de la jubilación se ha consolidado como una solución reciente⁸ para el desafío del envejecimiento, al considerar beneficios de una vivacidad física, mental, sexual, social y laboral, convirtiendo a la vejez en una etapa de actividad e independencia (Lassen, 2015). Es el proceso óptimo de las oportunidades de salud, seguridad y participación social, económica, cultural, espiritual y cívica en la que los individuos mejoran su calidad de vida en la medida que envejecen (Vega & Ayala, 2015).

La población mayor debe tener la oportunidad de trabajar hasta cuando ellos lo quieran y sean capaces de hacerlo, de forma satisfactoria y en posibilidad de acceder a la educación y programas de capacitación⁹ (Nava-Bolaños & Ham-Chande, 2014). En el mismo sentido, supone una visión de envejecimiento funcional, cuyas características principales se muestran en la **FIGURA 1.3**.

Se puede considerar al envejecimiento activo como una herramienta que induce el cambio social, haciendo uso de formas clásicas como la educación adecuada y sana (productividad continuada y alto nivel de actividad), cambios infraestructurales (centros de actividades y residencias para adultos mayores en donde éstos participen en actividades diarias) y, por supuesto, reformas en el mercado laboral (retraso en la edad de jubilación y transiciones más flexibles hacia la jubilación) (Lassen, 2015).

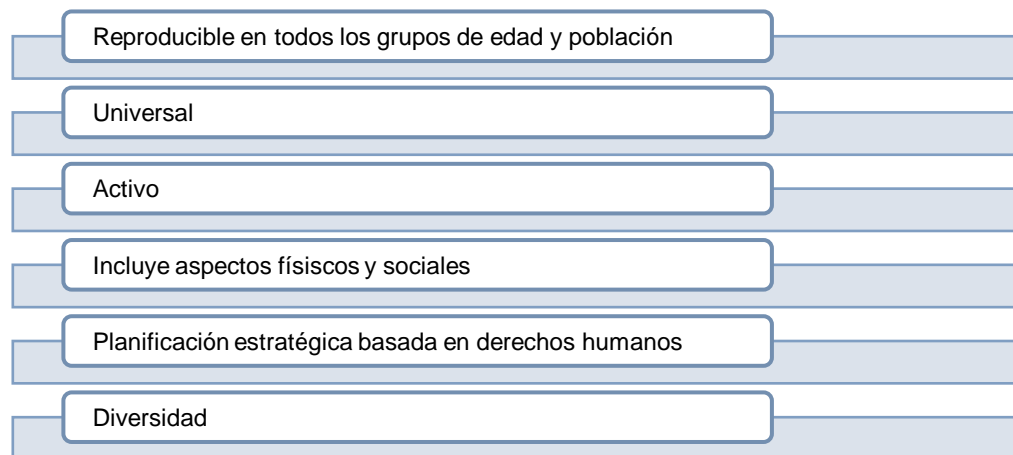
Un problema que tiene el envejecimiento activo es que parece excluir a los adultos mayores más vulnerables, dependientes, menos capaces o gravemente enfermos

⁸ Surgió a finales de la década de los 90 con eventos de las Naciones Unidas “Año Internacional de las Personas de Edad: hacia una sociedad para todas las edades”, “II Asamblea Mundial sobre el envejecimiento: construir una sociedad para todas las edades” y “Año Europeo del envejecimiento activo y de la solidaridad intergeneracional”.

⁹ En Madrid, abril de 2002, se reconocieron los derechos económicos y participación en la sociedad de los adultos mayores.

(Boudiny , 2013). Al ser un proceso transversal, crear políticas de vejez que se centren en la última etapa del ciclo vital, sería un error. Los países podrán implementar estrategias que mejoren la salud, participación voluntaria y seguridad de los mayores (Zuluaga-Callejas & Robledo-Marín, 2016).

FIGURA 1.3 Características del envejecimiento activo



FUENTE: Elaboración propia con base en (Muñoz-Cobos & Espinosa-Almen, 2008).

Las actividades que deben hacerse en materia de política pública van desde la preparación de la jubilación hasta eliminar la discriminación por edad en ambientes laborales. Asimismo, se apela a la creación de organizaciones de grupos productivos de personas mayores, a la participación económica, protección de los derechos laborales y atención preferente a la población en pobreza e indigencia y mejoramiento del sistema de seguridad social (CEPAL/CELADE, 2003).

Incrementar el ingreso de los adultos mayores debe ser resultado de acciones como empleos con mayor valor social para garantizar su seguridad económica (Rubalcava-Montero, 2019). A su vez, las políticas públicas se deben centrar en la capacidad productiva y eliminación de obstáculos por edad, fomentando su contratación,

promoviendo una imagen realista de sus capacidades y elevando las tasas de participación laboral para las mujeres (Pessanha & Alves, 2019).

1.4.2 Seguridad social

Teniendo en cuenta a Bertranou (2005), tradicionalmente la seguridad social se considera como el régimen del Estado que garantiza protección en caso de accidentes o enfermedades de trabajo o comunes, desempleo, maternidad, invalidez, vejez, jubilación, sobrevivencia o muerte con prestaciones por hijos y familia, ya sea a través de dinero o especie. Incluye prestaciones de salud, prevención, rehabilitación y cuidados de largo plazo, así como fondos de previsión, asistencia y seguro social.

Se constituye como una herramienta que garantiza protección y autonomía a los individuos, bajo regímenes y programas establecidos por el Estado; sin embargo, en países en desarrollo, como es el caso de México, un porcentaje muy pequeño está en esquemas de protección social, por lo que la cobertura es bastante baja y los beneficios son raquíticos para solventar los gastos personales propios y de los dependientes económicos (Madrigal-Martínez, 2010).

Simplemente, en materia de pensiones¹⁰, México tiene baja cobertura en el rubro contributivo, 53% de las mujeres y 43% hombres, cuyos beneficios se traducen en montos promedio de \$5 128 y \$6 602, respectivamente. Por otro lado, los beneficiarios de ambos sexos de una pensión no contributiva reciben \$600, representando 13% del ingreso monetario individual para las mujeres y 6% para hombres; en general, 26% de los adultos mayores no recibe ingresos por pensión de ningún tipo (CONSAR, 2018).

¹⁰ Los datos y cifras son para poner en relieve cual fue la situación pensionaria en México, 2018, siendo éste el año de estudio del presente trabajo.

En lo que respecta en seguridad social y en concordancia con la Encuesta Nacional de Empleo y Seguridad Social (2017), solamente 11.93% de la población de 60 años y más no tiene afiliación a alguna institución de salud pública y de la población cubierta, 47.87% es derechohabiente del IMSS, seguida de la afiliada al Seguro Popular con 36.84% y al ISSSTE con 10.37%, en el sector privado sólo 0.42%. No existen muchas diferencias entre ambos sexos. Véase **TABLA 1.6**.

Además del seguro médico y la seguridad social, las pensiones están ligadas con el trabajo formal. En especial, el acceso a pensiones suele estar más acotados porque se necesita de un mínimo de contribuciones y las transferencias entre fondos son limitadas, desembocando en una tasa de remplazo (la pensión como porcentaje de los ingresos previos a la jubilación) variable respecto al número de años cotizados e ingresos laborales previos (van Gameren, 2008).

TABLA 1.6 Condición de afiliación a los sistemas de seguridad social para la población adulta mayor, en porcentaje

	Sin afiliación	Tipo de institución					
		Afiliados	IMSS	ISSSTE	Seguro Popular	Otra institución pública	Institución privada
TOTAL 60+	11.926%	88.047%	47.865%	10.370%	36.840%	4.506%	0.419%
HOMBRES	12.619%	87.359%	47.665%	9.604%	37.488%	4.859%	0.385%
MUJERES	11.351%	88.619%	48.029%	10.997%	36.309%	4.218%	0.447%

FUENTE: Elaboración propia con base en INEGI. Encuesta Nacional de Empleo y Seguridad Social. 2017

Sobre todo, los nuevos sistemas de pensiones que están basados en el ahorro individual fungen como un complemento para la seguridad en los ingresos en la vejez, en especial para aquellos que tienen una participación en los mercados laborales formales. Aunque estos mismos sistemas no cuentan con herramientas para los que están fuera de estos mercados, para los que se ven obligados a salir y los que no tienen recursos suficientes, condenándolos a la pobreza (CEPAL/CELADE, 2003).

Esto no hace más que generar desigualdades e inequidades, es así como tanto el desempeño laboral y las diferencias de acceso a la protección social generan y perpetúan estas situaciones; se debe focalizar su reducción a través de la ampliación de la cobertura y estandarización de los programas, además de la eliminación de mecanismos regresivos en el financiamiento, de la promoción y cooperación para el diálogo para las reformas de los sistemas (Bertranou, 2005).

Entonces, se pone de relieve la situación actual de la jubilación, siendo el dispositivo burocrático que convierte legalmente en ancianos a la población. Aun cuando se le puede comprender como una herramienta que legaliza la expulsión de la fuerza laboral que se considera obsoleta y la cambia en virtud de las necesidades de un patrón específico de país, el fenómeno tiene también otros alcances individuales y colectivos (Aranibar, 2001).

Es así como las pensiones no contributivas se convierten en un mecanismo para abolir la pobreza en hogares con personas de la tercera edad; otra situación a considerar es el papel que tienen en el bienestar económico en los mismos hogares (Huenchuan, 2005a). Empero, suelen otorgar prestaciones en efectivo modestas y uniformes para los riesgos de vejez, invalidez y muerte, situación desventajosa para los excluidos de la seguridad social y en inseguridad económica (Bertranou, 2005).

Estudios demuestran que pensiones de tipo no contributiva son un aliciente para la pobreza y de costo razonable para todos los países (CEPAL/CELADE, 2003). La población mayor que no recibe ingresos por concepto de jubilación, pensión o trabajo remunerado tiene a la familia y redes sociales como su soporte económico y si se pierde

alguno de estos determinantes, un porcentaje considerable caería en situación de pobreza (Huenchuan & Guzmán, 2007).

1.4.3 Transferencias familiares

Se entiende a las redes de apoyo como la práctica que involucra relaciones interpersonales que vinculan a una persona con su entorno, permitiendo mejorar o mantener su bienestar material, físico y emocional para evitar un deterioro que podría generarse ante dificultades. Estas redes son especialmente importantes para los sectores desfavorecidos, creando una alternativa social para la supervivencia y que complementa los mecanismos formales (Estado y mercado) (Guzmán *et al.*, 2003).

Los apoyos familiares dentro de la seguridad económica han cobrado especial importancia, sobre todo en lo que respecta a la composición de ingresos de las personas y hogares. Conceptos como “transferencias familiares”, “rentas provenientes de asistencia privada” y “ayudas familiares” aluden a un mismo contenido como una suerte de sinónimos: apoyos, ya sea en dinero o en especie, hacia personas que los necesitan (Huenchuan, 2005b).

La situación socioeconómica de la población vieja no se puede analizar solo desde la perspectiva de la protección social; el factor que también tiene un peso significativo es el apoyo que les brinda su familia, como una especie de compensación de las instituciones de seguridad social, cuyas coberturas no son universales. Pero es en la vejez donde se viven situaciones transitorias que modifican el tamaño de la estructura familiar (Garay & Montes de Oca, 2006).

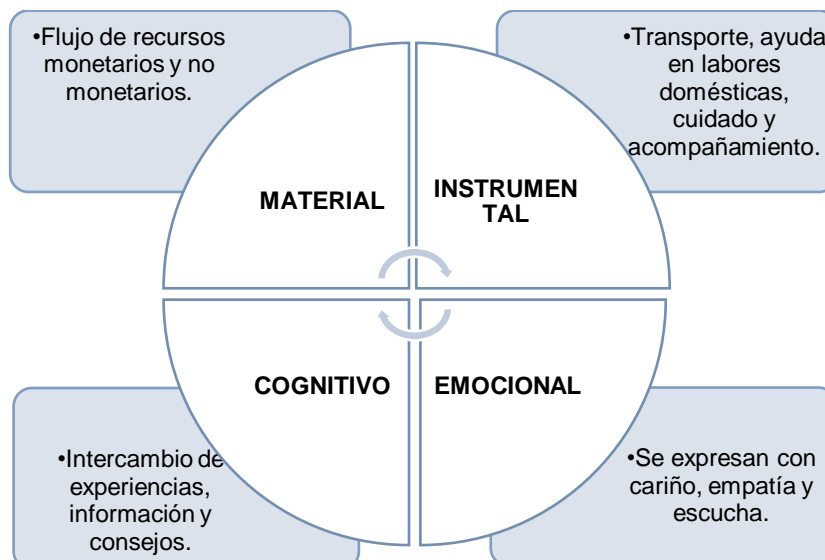
El apoyo de tipo familiar está sujeto a las condiciones socioeconómicas, del tamaño y composición de los hogares, formas de organización y la tendencia en la reducción de

hijos (Nava-Bolaños & Ham-Chande, 2014; Guzmán *et al.*, 2003). Cuando un individuo no logra acceder a un nivel de seguridad económica en la tercera edad, entran en acción diferentes mecanismos de transferencia de tipo familiar que usualmente son reciprocas, formando un intercambio (Huenchuan, 2005b).

Según Guzmán *et al.*, (2003) existen cuatro categorías de apoyos o transferencias, véase **FIGURA 1.4**.

La gran mayoría de las mujeres mayores vive alejada de la pobreza debido a su alta probabilidad de recibir transferencias familiares, lo que funge como un aliciente a sus escasas contribuciones de la seguridad social; mientras que sus pares masculinos siguen siendo proveedores incluso en la vejez, permitiéndoles cierto grado de cuidado y calidad de vida, empero, sus redes de apoyo familiar son más reducidos, que a la larga puede mermar su bienestar (Huenchuan, 2005b).

FIGURA 1.4 Categorías de transferencias familiares



FUENTE: Elaboración propia con base en (Guzmán *et al.*, 2003).

Además, cuando Garay & Montes de Oca (2006), analizan la relación entre el nivel socioeconómico con el arreglo familiar que tiene el adulto mayor, las autoras encuentran

que, conforme a mayor posición económica, mayor probabilidad de vivir en hogares nucleares; en sentido opuesto, gente de edad con menor capacidad tienden a residir en hogares extensos, pudiendo ser una estrategia de apoyo mutuo, misma que está limitada y no siempre se dan en condiciones de solidaridad genuina.

Es así como la dimensión familiar sigue prevaleciendo como un instrumento que define a la seguridad económica de las personas de edad, y si se considera que prevalece la idea tradicionalista de que los hijos deben mantener económicamente a sus padres, en especial en países como México, como una especie de obligación moral y de recompensa por la crianza, se arraiga aún más esta fuente de ingresos en la población de edad avanzada.

En la siguiente sección se aborda el método estadístico que se emplea para dar respuesta a lo que se ha planteado en este capítulo y cómo se busca cuantificar un fenómeno predictivo de carácter categórico, es ideal el uso de la regresión logística. Se establece su eficacia, los supuestos en los que basa, cómo se hace la evaluación global del modelo, de los coeficientes y de los errores.

CAPÍTULO II: ANÁLISIS DE DATOS DE RESPUESTA BINARIA. REGRESIÓN LOGÍSTICA

Este capítulo se hace una descripción técnica sobre la regresión logística, se empieza con su definición, el tipo de variables que se puede llegar a requerir, así como los métodos de selección de las variables para encontrar la parsimonia del modelo. Al mismo tiempo, se demuestra que el modelo logit es ideal en la estimación de la probabilidad de que un adulto mayor se encuentre trabajando, así como su poder clasificatorio.

Los métodos de regresión son integrales para analizar datos si se trata de describir asociaciones entre variables dependientes e independientes; el propósito de estos métodos es el mismo que cualquier otra técnica de modelación: encontrar el mejor ajuste parsimonioso para los datos. Dentro de estos métodos de regresión podemos encontrar a la regresión logística, la cual resuelve el caso en que la variable dependiente es discreta y busca dar respuesta a problemas de diferentes campos en los cuales otras técnicas estadísticas no son útiles (Hosmer & Lemenshow, 2000).

Permite hacer mediciones del riesgo de padecer ciertas situaciones a través de modelos predictivos complejos, tener control sobre posibles variables confusas y averiguar interacciones, siempre que se trate de una perspectiva dicotómica (Sagaró del Campo & Zamora-Matamoros, 2019). Se consolida como una técnica estadística sólida, cuya dificultad radicaba en la averiguación de la estimación de los coeficientes, y es el algoritmo de Walter-Duncan el que lo soluciona, pero que, por su naturaleza, el uso de una computadora se hace imprescindible (De la Fuente-Fernández, 2011).

2.1 Definición de regresión logística y conceptos preliminares

El análisis de regresión logística es una técnica estadística multivariante de dependencia que se encarga de relacionar una variable con un conjunto de variables independientes.

Su objetivo se centra en las predicciones del comportamiento de algún fenómeno a través de la estimación de probabilidades de un suceso definido por la variable dependiente; es decir, se pretende pronosticar la pertenencia a un grupo (López-Roldán & Fachelli, 2015).

De acuerdo con Fiuza-Pérez & Rodríguez-Pérez (2000) la regresión logística tiene una doble función: explicativa y predictiva. Su uso predictivo es el más frecuente y extendido para el estudio de riesgos, en cualquiera de las vertientes de pronóstico, estudios prospectivos metódicos de caso-control y ensayos clínicos; respecto al fin explicativo, hace una descripción útil basada en información reducida, facilitando el tratamiento cuantitativo de los datos.

Adicionalmente, tiene el propósito de explicar una característica cualitativa, misma que puede ser (Fiuza-Pérez & Rodríguez-Pérez, 2000; López-Roldán & Fachelli, 2015):

- Modelo Binomial. Una cualidad binomial, es el modelo más utilizado y explica un suceso o característica dicotómica.
- Modelo Multinomial. Una cualidad que puede tomar más de una categoría diferente, exhaustiva y mutuamente excluyente, que suele ser politómica.
- Modelo anidado. Una cualidad que representa una decisión que supone un encadenamiento.
- Modelo ordenado. Una cualidad con varias modalidades que tienen un orden natural.

La primera modalidad, como su mismo nombre sugiere y que es objeto de estudio para el presente trabajo de investigación, consta de una variable dependiente cualitativa con dos etiquetas que corresponden a la ausencia y presencia de cierta característica o

suceso; si sólo se contempla una variable independiente, se trata de una regresión logística simple y si se consideran dos o más se denomina regresión logística múltiple (López-Roldán & Fachelli, 2015).

El interés de la regresión logística binaria gira en torno a la evaluación de la ocurrencia o no ocurrencia de la característica de estudio, así como el efecto de uno o más antecedentes sobre el hecho de que el suceso se produzca; si se denota como Y a la variable dependiente dicotómica, sólo se admiten dos valores: toma el valor de uno si el hecho sucede y, para el caso contrario, cuando el suceso no ocurre, toma el valor de cero (Silva-Aycaguer, 1995).

Consiste en obtener una función clasificatoria de los datos en grupos preestablecidos por los dos valores de la variable dependiente (Fiuza-Pérez & Rodríguez-Pérez, 2000). Entonces, la regresión logística procura expresar la probabilidad de que ocurra el suceso en función de las variables independientes, también llamadas covariables¹¹, que se presumen relevantes; representando la flexibilidad en las variables de entrada uno de sus mayores atractivos (Silva-Aycaguer, 1995).

Adicionalmente, la regresión logística mezcla el análisis de contingencia con los modelos log-lineales y el análisis de regresión por mínimos cuadrados; en ambas situaciones se enfrentan a limitaciones que esta técnica multivariante resuelve eficientemente, para el primer caso los dos modelos de dependencia no pueden usar variables continuas y el segundo caso las variables categóricas no funcionan siempre como buenas predictoras (López-Roldán & Fachelli, 2015).

¹¹ Esta denominación es bajo el contexto de la regresión logística.

Se procede a establecer el origen que pueden tener las variables explicativas: categóricas, cualitativas, dicotómicas como politómicas, continuas o nominales (Silva-Aycaguer, 1995); además, se busca comprender su efecto individual como su posible interacción, asimismo denotar la importancia entre cada una de estas variables con la dependiente y clasificar los datos en la correspondiente categoría de acuerdo con la probabilidad de pertenencia (López-Roldán & Fachelli, 2015).

En el mismo sentido, todas las técnicas estadísticas que involucran la asociación entre variables miden objetos o sujetos. Para efectos de comodidad en el presente trabajo se entiende como variable respuesta, dependiente o de resultado a aquella que fluctúa en torno a otras denominadas exploratorias, predictivas o independientes, todas comprendiéndose como sinónimos; la primera variable es de tipo aleatorio, mientras que las covariables son observaciones o medidas no aleatorias (Dobson, 2002).

Aun así, ambos tipos de variables pueden caer en alguna de las siguientes escalas (Dobson, 2002):

- Nominal. Los valores que representan no siguen un orden específico; aquí entran las variables binarias, dicotómicas o binomiales para el caso de dos grupos y si tienen más categorías se denominan policotómicas, politómicas o multinomiales.
- Ordinal. Indica la existencia de un orden natural o rango entre las categorías. Tanto esta como la anterior clasificación en conjunto se denominan categóricas o discretas.
- Continua. Las observaciones están en el conjunto de los números reales, incluye intervalos y razones. El término cuantitativo se usa para este tipo de variable.

Finalmente, la regresión logística se diferencia de la regresión múltiple o el análisis discriminante porque cuando se estima el modelo se busca la predicción y explicación de las relaciones que influyen en la categoría de los datos, así como controlar su efecto, lo que se logra a través de la probabilidad de un suceso; entonces, los valores que resulten deben estar acotados entre cero y uno, haciendo a esta técnica multivariante la herramienta perfecta (Hair *et al.*, 1999).

2.2 Diferencias con el análisis de regresión lineal y análisis discriminante

El concepto de regresión tiene implícita la formulación matemática que relaciona variables correlacionadas, y si se quiere lograr esto en el campo estadístico, se acude a la regresión lineal simple o múltiple. Esta técnica está sustentada con el método aritmético de mínimos cuadrados, pero cuando la variable respuesta sólo puede tomar dos valores, al evaluar la función los valores resultantes carecerán de sentido, por lo que la regresión lineal debe ser descartada (Fiuza-Pérez & Rodríguez-Pérez, 2000).

En otras palabras, lo que distingue a la regresión logística con la lineal es la naturaleza dicotómica o binaria que tiene la variable dependiente, diferencia que se refleja en la elección de un modelo paramétrico y en los supuestos que sigue. Un hecho importante es que los principios que rigen un análisis mediante regresión lineal también sirven de guía para un análisis basado en la regresión logística (Hosmer & Lemenshow, 2000).

El modelo de regresión lineal tiene como característica principal que sus variables, ya sea la dependiente o las independientes, son todas cuantitativas (López-Roldán & Fachelli, 2015); aun así, permite que las exploratorias sean categóricas con su debida

transformación en *dummy*¹² (Alderete, 2006). Además, tiene que establecer una serie de condiciones rigurosas para su aplicación que lo terminan limitando en su uso, particularmente en estudios por encuesta (López-Roldán & Fachelli, 2015).

De forma general, el resultado clave de la asociación entre la variable dependiente y las predictivas para cualquier problema de regresión es la media condicional de Y dado x, es decir, $E(Y|x)$; para el caso lineal, se asume que este valor esperado se representa matemáticamente a través de la siguiente ecuación que se mueve en la recta real (Hosmer & Lemenshow, 2000):

$$E(Y|x) = \beta_0 + \sum \beta_i * X_i \quad (2.2.1)$$

A su vez, para el caso dicotómico, y como se ha establecido anteriormente, $E(Y|x)$ oscila gradualmente en el intervalo cerrado acotado entre cero y uno, lo que desemboca en una representación gráfica curva con forma de S. En la literatura se usan diferentes distribuciones acumulativas conocidas para el modelo binario del valor esperado de Y dado x, pero siguiendo a Hosmer & Lemenshow (2000) se usa la distribución logística para efectos del presente trabajo.

Asimismo, en la regresión lineal una observación específica de Y se expresa como se muestra en **(2.2.2)**, donde ε es el error que expresa la desviación de la media condicional. Es usual que $\varepsilon \sim N(0, \sigma^2)$, de lo que se deriva que la distribución condicional de Y dado x sigue una distribución normal con media $E(Y|x)$ y varianza constante;

¹² Las variables independientes categóricas deberán contener parámetros para los distintos niveles de un factor, los elementos elegidos tienen que incluir todos los parámetros apropiados para cada observación (Dobson, 2002). Requiere que las variables se deconstruyan en diversas variables dicotómicas ficticias como categorías tenga la variable menor una, la que fungirá como referencia (López-Roldán & Fachelli, 2015).

situación que no ocurre cuando se presenta en la regresión logística binaria (Hosmer & Lemenshow, 2000).

$$y = E(Y|x) + \varepsilon \quad (2.2.2)$$

Entonces, el valor para una observación de Y dicotómica se muestra en la ecuación (2.2.3); de ésta surgen dos conclusiones: la primera es que la distribución condicional de Y es binomial con probabilidad $\pi(x)$ ¹³ y la segunda es que la misma distribución referida describe los errores con media cero y varianza $[\pi(x) * (1 - \pi(x))]$, resultando ser la base del análisis bajo un modelo logístico (Hosmer & Lemenshow, 2000).

$$y = \pi(x) + \varepsilon$$

$$\varepsilon = \begin{cases} 1 - \pi(x), & \text{si } y = 1 \\ -\pi(x), & \text{si } y = 0 \end{cases} \quad (2.2.3)$$

La obtención de los coeficientes de la regresión logística, así como sus errores, se hace mediante la técnica de máxima verosimilitud para que se maximice la probabilidad de obtener los valores de Y (López-Roldán & Fachelli, 2015); mientras que, en la regresión lineal se estiman los valores de los parámetros de tal forma que la suma de los errores cuadrados de predicción se minimicen, procedimiento que recibe el nombre de mínimos cuadrados ordinarios (Hair *et al.*, 1999).

Respecto a la distribución condicional de Y, es necesario hacer hincapié que al ser binomial sólo tiene dos resultados posibles que son opuestos, siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones; al no poder distribuirse normal, se ajusta con eficacia a la distribución binomial y, consecuentemente, la varianza

¹³ Para simplificar la notación se usa $E(Y|x) = \pi(x)$. La forma específica que tiene el modelo y la transformación que implica se estudia en la sección 4 de este capítulo.

no puede ser constante, presentándose situaciones de heterocedasticidad en los datos (Alderete, 2006).

Por otra parte, la regresión logística binaria es equivalente al análisis discriminante de dos grupos, pero la primera suele ser más adecuada en muchos casos. Mientras que la segunda técnica multivariante al tener la capacidad de tratar a más de dos grupos pasa a llamarse análisis discriminante múltiple; pero que, en líneas generales, se puede aplicar en situaciones donde se trata de identificar el grupo de pertenencia a través de las variables independientes seleccionadas (Hair *et al.*, 1999).

La regresión logística binaria tiene mayor flexibilidad en cuanto al supuesto de normalidad y la homocedasticidad (es la igualdad en las varianzas: la variable dependiente dicotómica no debe tener varianza constante); en el mismo sentido, es estadísticamente más potente y robusta que el análisis discriminante múltiple cuando no se cumplen estos supuestos¹⁴ (Alderete, 2006).

Es así como, desde la perspectiva matemática, la regresión logística es una herramienta fácil de usar, bastante flexible en los supuestos que requiere comparándola con la regresión lineal o análisis discriminante, además de que tiene una interpretación bastante significativa tanto a nivel individual como global (Hosmer & Lemenshow, 2000); más adelante, en las secciones 2.6 y 2.8 respectivamente, se tratan a detalle estos aspectos.

¹⁴ El análisis discriminante implica obtener un valor teórico a partir de una combinación lineal de dos o más variables independientes que discriminen mejor a los grupos definidos *a priori*. El carácter no métrico de la variable resultado se adecúa con predicciones de pertenencia al grupo con base en una puntuación discriminante, misma que requiere de un cálculo de puntuaciones de corte y la asignación de observaciones de los grupos (Hair *et al.*, 1999).

2.3 Supuestos

Como cualquier otra técnica multivariante, la regresión logística binaria requiere el cumplimiento de supuestos; y es que, antes de construir el modelo se necesitan considerar ciertos aspectos como precaución para que tenga suficiente validación estadística, entre ellos están: el tamaño de la muestra, tratamiento de covariables, así como sus criterios de inclusión y exclusión, multicolinealidad y monotonía (Sagaró del Campo & Zamora-Matamoros, 2019).

El **tamaño de la muestra** que requiere la regresión logística debe ser grande. La literatura establece criterios específicos a tomar en cuenta, debido a que si ésta resulta pequeña se puede caer en errores estándar grandes y en sobreajustes en los coeficientes; por convención se adopta tener al menos diez observaciones por cada covariable para obtener resultados concisos y significativos (Sagaró del Campo & Zamora-Matamoros, 2019).

El uso de relativamente múltiples variables con pocos casos de estudios representa uno de sus mayores atractivos. Se sugiere que si una de las variables dicotómicas, en especial la variable dependiente, no tiene al menos diez casos en cada uno de sus posibles resultados, las estimaciones obtenidas no son confiables; para las covariables, la inclusión de un gran número de ellas puede indicar falta de estudio sobre el problema (Domínguez-Alonso & Aldana-Padilla, 2001).

El número de variables depende estrechamente del tamaño de la muestra. Pero más allá de esto, el centro de atención son los requisitos que deben de cumplir las variables para incluirlas o excluirlas del modelo, sobre todo si se adecuan con las preguntas de investigación; por otro lado, las covariables pueden ser medidas en

cualquier escala de forma inicial, pero es a través del análisis que pueden necesitar de alguna transformación (Sagaró del Campo & Zamora-Matamoros, 2019).

Al tratar variables categóricas nominales y ordinales y como se presenta en la sección 2, se les tiene que codificar como *dummy* considerando que para las últimas las categorías tienen un orden preestablecido. En el caso de las continuas, el mismo modelo logístico asume cambios unitarios entre éstas y la dependiente en cualquier punto de la escala covariable, por lo que los cambios son de la misma magnitud en Y, asemejándose a la linealidad¹⁵ (Sagaró del Campo & Zamora-Matamoros, 2019).

Se hace hincapié que en las variables ordinales también se puede asumir que su escala funciona aproximadamente a un nivel cuantitativo y que la distancia entre las categorías contiguas es la misma. Algunas veces conviene categorizar a las variables continuas porque a veces producen cambios unitarios que pueden ser intrascendentes o no ser constantes a lo largo del rango de las variables (Domínguez-Alonso & Aldana-Padilla, 2001).

Respecto **al tratamiento de las variables**, para incluir alguna variable exploratoria conviene hacer un análisis bivariado previo para averiguar asociaciones entre covariables y la variable dependiente. Sagaró del Campo & Zamora-Matamoros (2019) sugieren incluir a aquellas que tuvieron una relación suficiente con Y ($p -$

¹⁵ Algunas veces se considera un supuesto por sí mismo para el caso de las variables continuas, se trata de una relación lineal entre estas variables y sus respectivos resultados con transformación logit (Stoltzfus, 2011).

value¹⁶ < 0.25). Se descartan a las redundantes o muy relacionadas, o las que sean intervinientes a causa del mismo pronóstico en estudio.

En el mismo sentido, existen dos conceptos que no deben pasar desapercibidos: interacción y factor de confusión. El primero alude a que la influencia de una de las variables tiene sobre la probabilidad se modifica en función del valor de otra, por lo que es necesario introducir una nueva que sea producto de las anteriores; su uso debe estar justificado con que la influencia de una variable sobre la función logística varía en relación de otra variable del modelo (Domínguez-Alonso & Aldana-Padilla, 2001).

Al ingresar interacciones en un modelo de regresión, lleva implícito una jerarquía, es decir, se tienen que incluir interacciones de orden superior como inferior, así como los términos independientes de las variables. Por otro lado, un factor de confusión se produce cuando una variable excluida del modelo no muestra una relación significativa con la variable dependiente, pero sí de forma indirecta al relacionarse con otras covariables ligadas con la variable de estudio (De la Fuente-Fernández, 2011).

Conviene distinguir que se debe impedir la **multicolinealidad**¹⁷ porque provoca varianzas y covarianzas grandes, lo que desemboca en intervalos de confianza amplios para los coeficientes y, además, puede que algunas covariables se concluyan no significativas cuando se esperaba lo contrario *a priori*. Para su diagnóstico, se puede utilizar el factor de inflación de la varianza, los autovalores o el índice de condición (Sagaró del Campo & Zamora-Matamoros, 2019).

¹⁶ Se entiende como la probabilidad para declarar la significación de una prueba, en otras palabras, se busca que la probabilidad sea lo suficientemente pequeña como para rechazar la hipótesis planteada (Romero-Suárez, 2012).

¹⁷ Es una situación en donde los coeficientes del modelo de regresión tienen una fuerte correlación entre sí.

Cuando las variables individuales están muy correlacionadas con otras, desemboca en situaciones indeseables, específicamente, trae consecuencias indeseables al obtener un resultado inestable en el sentido que pequeños cambios en los datos puede causar grandes cambios en el estimador de máxima verosimilitud. Para elegir el mejor subconjunto de las covariables se puede calcular a través del factor de inflación de varianza para cada variable (VIF) (Hosmer & Lemenshow, 2000):

$$VIF = \frac{1}{1-R_i^2} \quad (2.3.1)$$

R_i^2 = coeficiente de determinación de la covariable X_i respecto a las otras.

Concerniente a los autovalores e índice de condición ambos suministran la cantidad de dimensiones entre variables independientes. Siguiendo a Vilar del Hoyo *et al.* (2008) cuando algunos autovalores están cercanos a cero, significa que las covariables están altamente correlacionadas; mientras que el índice de condición es la raíz cuadrada de la proporción entre el máximo autovalor y cada sucesivo autovalor, si tiende a ser mayor indica mayor dependencia.

En caso de presentarse y recordando que el objetivo es evitarla, la multicolinealidad se puede resolver a través de tres opciones: aumentar el tamaño de la muestra, omitir las covariables teóricamente menos relevantes, satisfactorias o que presenten más valores faltantes, o transformarlas centrándolas respecto a la media, estandarizando, usando escalas elaboradas o con ayuda de variables sintéticas previo análisis de componentes principales (Sagaró del Campo & Zamora-Matamoros, 2019).

Otro supuesto importante de recordar es la **independencia de los errores**, se busca que los resultados de la muestra estén separados entre sí, evitando duplicidad; y si llegan a existir resultados repetidos u correlacionados, los errores también lo estarán,

infringiendo el supuesto (Stoltzfus, 2011). Existen métodos para analizar datos correlacionados usando la técnica de regresión logística pero no son de interés para la presente investigación.

Por último, para que el modelo de regresión logística binario sea coherente por sí mismo y en sus resultados, es menester que cuente con **monotonía** en la relación entre las covariables y la probabilidad del evento estudiado; es decir, se debe evitar que esta probabilidad buscada aumente o disminuya drásticamente para ciertos recorridos de valores de las covariables en forma individual (Sagaró del Campo & Zamora-Matamoros, 2019).

2.4 Transformación *logit* y estimación del modelo logístico

La regresión logística es un modelo que permite analizar si una variable dicotómica depende de ciertas variables relevantes denominadas covariables (Alderete, 2006) a través de una probabilidad de ocurrencia del suceso, hecho o fenómeno de interés (Silva-Aycaguer, 1995). La variable dependiente, tal como sugiere el nombre de la técnica estadística que está siendo estudiada, debe ser binaria aleatoria y se puede definir como sigue (Dobson, 2002):

$$Y = \begin{cases} 1, & \text{si ocurre el suceso} \\ 0, & \text{si no ocurre el suceso} \end{cases} \quad (2.4.1)$$

Se sustenta a través de la principal característica de la distribución binomial, la probabilidad de éxito o la ocurrencia del suceso, la probabilidad de fracaso o la no ocurrencia del suceso y por el cociente de ambas (Alderete, 2006), nótese que deben ser excluyentes y exhaustivos de un fenómeno representando por Y . Si la probabilidad de que suceda uno de ellos es $\pi(x)$ y la probabilidad de que suceda el otro es igual a $1 - \pi(x)$, tal como se presenta a continuación (López-Roldán & Fachelli, 2015):

$$Y = \begin{cases} 1, & \Pr(Y = 1) = \pi(x) \\ 0, & \Pr(Y = 0) = 1 - \pi(x) \end{cases} \quad (2.4.2)$$

Una aclaración importante es que no es importante a qué se le denomina $Y = 1$ y a qué $Y = 0$, es decir, para lo que se quiere resolver no es trascendente como se delimita la variable dependiente, pero, al interpretar los coeficientes se debe recordar como está definida. Por otra parte, la forma analítica en que esta probabilidad se relaciona con las covariables $X_i \forall i \in \{1, 2, \dots, p\}$ y suponiendo que son p , se expresa en **(2.4.3)** en su forma más general (Silva-Aycaguer, 1995):

$$\Pr(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} \quad (2.4.3)$$

La expresión anterior se conoce como función logística, responde a la formulación matemática de la curva logística¹⁸ $\left(f(x) = \frac{1}{1+e^{-x}}\right)$ para el caso de regresión logística binaria (López-Roldán & Fachelli, 2015), en donde $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son los parámetros del modelo y e funge como la función exponencial¹⁹ (Silva-Aycaguer, 1995), misma que describe las relaciones entre las variables, mientras que los parámetros de la ecuación se interpretan de forma multiplicativa (López-Roldán & Fachelli, 2015).

La ecuación **(2.4.3)** también mide la probabilidad de que la variable dependiente tome el valor de uno o la presencia de la característica estudiada en función de las covariables. El modelo se formaliza con ayuda de la curva logística, en donde se plantea

¹⁸ Nótese que la curva logística sigue una distribución logística con media cero y varianza uno. La definición opera con la función exponencial $f(x) = e^x$ así como con su función inversa $f(x) = \ln x$ (Silva-Aycaguer, 1995).

¹⁹ Al producir la función exponencial valores mayores que cero para cualquier argumento en un intervalo cerrado entre cero y uno (Domínguez-Alonso & Aldana-Padilla, 2001).

una función matemática en forma sigmoidea como su representación gráfica y que verifica las siguientes cuatro propiedades (López-Roldán & Fachelli, 2015):

— Sus valores oscilan entre cero y uno, permitiendo una interpretación en términos probabilísticos.

— Límite inferior es cero.

$$\lim_{x \rightarrow -\infty} \frac{1}{1 + e^{-x}} = \frac{1}{1 + e^{-(-\infty)}} = \frac{1}{1 + e^{\infty}} = \frac{1}{\infty} = 0$$

— Límite superior es uno.

$$\lim_{x \rightarrow +\infty} \frac{1}{1 + e^{-x}} = \frac{1}{1 + e^{-(\infty)}} = \frac{1}{1 + e^{-\infty}} = \frac{1}{1 + 0} = 1$$

— Cuando la variable independiente toma el valor de cero, la función vale $\frac{1}{2}$.

$$f(0) = \frac{1}{1 + e^{-0}} = \frac{1}{1 + 1} = \frac{1}{2}$$

Entonces, la función logística **(2.4.3)** halla para cada observación o dato, sujeta al conjunto de las covariables, la probabilidad de que se presente el efecto estudiado. Para esta ecuación se plantea una transformación logarítmica, denominada *logit*, esta ecuación, que consiste en convertir la probabilidad $\pi(x)$ en *odds*, de aquí nace la ecuación de la regresión logística, bastante similar a la de la ecuación de la regresión lineal (Fiuza-Pérez & Rodríguez-Pérez, 2000).

A través de la transformación logarítmica la relación de dependencia se puede interpretar en términos lineales y es que **(2.4.3)** permite asignar valores a las covariables para generar valores de Y de la misma forma en cómo se haría en la regresión lineal, siendo la diferencia principal que las predicciones se sitúan siempre en el intervalo (0,1),

lo que facilita la interpretación de resultados y parámetros en términos de probabilidad y que pronostiquen un comportamiento (López-Roldán & Fachelli, 2015).

Los coeficientes en la regresión logística cuantifican el cambio producido por variaciones unitarias de las covariables en el logaritmo neperiano del cociente de probabilidades de los dos sucesos, aludiendo nuevamente a la transformación *logit* y al concepto de *odd* (López-Roldán & Fachelli, 2015), teniendo presente que una de las cualidades de la distribución binomial es el uso esta división para averiguar cuanto más probable es el éxito que el fracaso (Alderete, 2006).

Formalmente, el *odd* es la división de la probabilidad de que ocurra un suceso y la probabilidad de que no ocurra (López-Roldán & Fachelli, 2015), es un número que indica cuánto es más probable se produce un hecho frente a que no ocurra el mismo hecho en cuestión (Silva-Aycaguer, 1995); siguiendo López-Roldán & Fachelli (2015) y Silva-Aycaguer (1995) esta razón se puede expresar como sigue:

$$odd = \frac{\text{Probabilidad de que ocurra el suceso}}{\text{Probabilidad de que no ocurra el suceso}} = \frac{\Pr(Y = 1)}{\Pr(Y = 0)} = \frac{\pi(x)}{1 - \pi(x)} \quad (2.4.4)$$

En el contexto latino es más común tratar con probabilidades (Silva-Aycaguer, 1995), del mismo modo, se puede obtener una equivalencia para pasar de los *odds* a las probabilidades (2.4.5) (López-Roldán & Fachelli, 2015). Entre la probabilidad y el *odd* existe una asociación directa, si la primera aumenta, el *odd* también; si $\pi(x) = 0$, el *odd* también valdrá cero, pero en medida que la probabilidad tiende a uno, el *odd* tiende al infinito (Silva-Aycaguer, 1995).

$$\pi(x) = \frac{odd}{1 + odd} \quad (2.4.5)$$

Específicamente, un valor de uno en el *odd* significa que existe equiprobabilidad en las categorías de la variable dependiente, en cambio, un valor mayor de uno indica que la categoría estudiada tiene mayor probabilidad de ocurrencia; aun así, su principal inconveniente es que su valor oscila entre cero e $+\infty$. Con ayuda de la transformación logarítmica en los *odds* proporciona una medida de cuán importante es para el análisis de datos categóricos, variando entre $(-\infty, +\infty)$ (Alderete, 2006).

En las ecuaciones **(2.4.4)** y **(2.4.5)** se mide que tan probable es un suceso, o bien, su riesgo. El concepto de riesgo relativo indica el cociente de probabilidades de un suceso en condiciones distintas, en otras palabras, es la razón de dos *odds*, la cual se le conoce como *Odd Ratio* (OR) (López-Roldán & Fachelli, 2015), correspondiente a un hecho bajo cierta condición entre los que le corresponden bajo otra. Esta definición cobra especial interés en el contexto de la regresión logística (Silva-Aycaguer, 1995).

El OR funge como el indicador básico de la regresión logística, se encuentra determinado por el exponencial del coeficiente de regresión (β_i) de cada covariable. Al mismo tiempo, esta técnica multivariante exige una relación lineal entre el logaritmo neperiano del *odd* y las ya referidas covariables, lo que conlleva la necesidad de una transformación de por medio (Sagaró del Campo & Zamora-Matamoros, 2019).

Los posibles cuatro valores de la regresión logística se pueden presentar en una tabla de contingencia de 2×2 . Los resultados entre los datos con $x = 1$ se entienden como $\pi(1)/[1 - \pi(1)]$; asimismo, el *odd* resultante entre los datos con $x = 0$ se conceptualizan como $\pi(0)/[1 - \pi(0)]$. Entonces, el (OR) es la proporción de *odds* de $x = 1$ comparado con los *odds* de $x = 0$ se formula de la siguiente manera (Hosmer & Lemenshow, 2000):

$$OR = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} = e^{\beta_i} \quad (2.4.6)$$

La relación presentada en **(2.4.6)** concentra la potencia estadística que posee la regresión logística como una herramienta de investigación. El OR es una asociación que aproxima que tan probable es que el resultado deseado se presente respecto a si no se presenta. Esta interpretación se basa en que aproxima el riesgo relativo, este parámetro es $\pi(1)/\pi(0)$, de aquí se deriva que el OR se aproxima con el riesgo relativo si $[1 - \pi(0)]/[1 - \pi(1)] \approx 1$ (Hosmer & Lemenshow, 2000).

A partir de las ecuaciones **(2.4.3)**, **(2.4.4)** y **(2.4.6)** se obtienen las siguientes relaciones (López-Roldán & Fachelli, 2015):

$$\begin{aligned} \frac{\Pr(Y = 1)}{\Pr(Y = 0)} &= \frac{\pi(x)}{1 - \pi(x)} = \frac{\frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}}{1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}} \\ &= \frac{\frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}}{\frac{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} - 1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}} \\ &= \frac{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}{e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} (1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)})} \\ &= \frac{1}{e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \end{aligned}$$

Efectuando las operaciones algebraicas el resultado queda expresado en la siguiente ecuación (López-Roldán & Fachelli, 2015):

$$\frac{\Pr(Y = 1)}{\Pr(Y = 0)} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \quad (2.4.7)$$

Este resultado es bastante importante, pues muestra directamente una medida de riesgo relativo a través de un OR de un perfil específico respecto a otro en términos de

los parámetros de la regresión logística ($\beta_0, \beta_1, \beta_2, \dots, \beta_p$) (Silva-Aycaguer, 1995). Por último, si a la ecuación anterior se le aplica logaritmo base e , resulta la transformación *logit*, permitiendo caracterizar al modelo en su forma lineal y aditiva (López-Roldán & Fachelli, 2015):

$$\log\left(\frac{\Pr(Y = 1)}{\Pr(Y = 0)}\right) = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.4.8)$$

Es un modelo logístico lineal porque lo es la escala del logaritmo de la razón de los productos cruzados (*odd ratio*), variando en toda la escala de los números reales (Alderete, 2006). A su vez, la función del modelo tiene una interpretación como el logaritmo neperiano de los *odds*, ampliamente usado para datos binomiales. El modelo logístico puede generalizarse en su forma matricial como se muestra en la siguiente ecuación (Dobson, 2002):

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = X_i^t \beta \quad (2.4.9)$$

Cabe destacar que esta transformación permite un análisis similar al de la regresión lineal múltiple (Ibarra & Michalus, 2010). En la ecuación de arriba (2.4.9), X_i es un vector de medidas continuas correspondiente a las covariables y variables *dummy* correspondiente a los niveles de los factores, así como β es el vector de parámetros; este modelo es una poderosa técnica análoga a la regresión lineal y al ANOVA (Análisis de Varianza en inglés) para variables continuas (Dobson, 2002).

Otra forma más convencional de representar este valor teórico es como una combinación lineal de variables con ponderaciones determinadas empíricamente, esta forma representa una relación exclusiva multivariante con coeficientes que muestran el peso relativo que tiene cada covariable; en donde π_i representa la proporción o

probabilidad de observar a la categoría a predecir, β_0 es una constante y X_i son las covariables (Alderete, 2006):

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.4.10)$$

El modelo de regresión logística es uno de carácter no lineal que se basa en la técnica de máxima verosimilitud de los datos para calcular estimaciones de los coeficientes y sus errores, y para ello requiere seguir algoritmos iterativos como el Newton-Raphson, que tras algunas fases sucesivas ajusta el modelo (López-Roldán & Fachelli, 2015). Al mismo tiempo, el modelo debe cumplir con dos aspectos básicos: parsimonia y congruencia (Sagaró del Campo & Zamora-Matamoros, 2019).

2.5 Máxima verosimilitud

La naturaleza no lineal de la transformación logística (2.4.8) requiere que, para el tratamiento de los coeficientes, se utilice iterativamente la estimación de máxima verosimilitud con el propósito de encontrar el valor más probable de los coeficientes; es por ello por lo que se plantea el uso del valor de la verosimilitud, reemplazando al tradicional método de mínimos cuadrados para calcular la medida de ajuste global del modelo (Hair *et al.*, 1999).

El Método de Máxima Verosimilitud (MMV), es una técnica de estimación que es ampliamente referida en diferentes situaciones, debido a que resulta ser mejor si se le compara con la de mínimos cuadrados ordinarios, la de momentos o la de cuantiles, sobre todo cuando la muestra es grande. El MMV utiliza toda la información de los parámetros y arroja un estimador totalmente flexible con mejores propiedades asintóticas (Omari *et al.*, 2018).

Para definir formalmente el MMV, sean X_1, X_2, \dots, X_n una muestra aleatoria de observaciones independientes e idénticamente distribuidas de una población desconocida y $X = x$ una realización específica de la variable aleatoria o del vector X con función masa o función de densidad²⁰ $f(x, \theta)$ en donde θ es un vector de parámetros a estimar, el objetivo estadístico es inferir θ a partir de los datos y obtener una función de verosimilitud de la variable aleatoria (Omari *et al.*, 2018).

La función de verosimilitud es la función masa o de densidad de los datos x expresada en relación con los parámetros θ ; entonces, dado que X_1, X_2, \dots, X_n posee una función conjunta $f(X_1, X_2, \dots, X_n | \theta)$ para cada muestra observada con distintos datos $\{x_i\}, \forall i = 1, 2, \dots, n$. La función de verosimilitud se define como sigue (Omari *et al.*, 2018):

$$L(\theta) = L(\theta | x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (2.5.1)$$

Dos resultados surgen de **(5.1)**: es no lineal en $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ y que requiere de un método iterativo para su solución, busca proveer una idea de partida para elegir un estimador asintóticamente eficiente para un parámetro o conjunto de parámetros $\hat{\theta}$ como el valor estimado que es más probable para la totalidad de los datos; el estimador de máxima verosimilitud, $\hat{\theta}$, se obtiene mediante la maximización de la función de verosimilitud $L(\theta)$, tal como se muestra a continuación (Omari *et al.*, 2018):

$$\hat{\theta}(x) = \arg \max L(\theta)^{21} \quad (2.5.2)$$

²⁰ Variará dependiendo si el vector de variables aleatorias es discreto o continuo.

²¹ Es el argumento del máximo, son los puntos del dominio en la que los valores son maximizados, es decir, son los argumentos en que las salidas de la función son los más grande posible.

Específicamente, la estimación de modelos no lineales (como es el caso de la regresión logística) presentan cierto nivel de complejidad debido a que la maximización no se puede resolver de forma analítica, teniendo que recurrir a métodos numéricos y la solución que obtengan debe ser lo suficientemente aproximada al valor óptimo. La inferencia estadística que sustenta a los modelos de este tipo se basan en distribuciones asintóticas de máxima verosimilitud (Aubone & Wöhler, 2000).

Este estimador de máxima verosimilitud (EMV) no tiene una distribución normal, es sesgado, no es de mínima varianza y tiende a tener propiedades desconocidas para muestras pequeñas, pero a medida que la muestra va adquiriendo robustez un mejor estimador se va adquiriendo; estas propiedades asintóticas afectan a las pruebas que se realicen sobre los parámetros, por lo que se debe tener en especial consideración el tamaño de la muestra (Aubone & Wöhler, 2000).

Tomando como punto de partida una expresión como $l_t = \mu(\theta, t) + \varepsilon_t^{22}$ con ε_t como el error aleatorio y $\mu(\theta, t)$ como el valor esperado de la variable l_t , la función de verosimilitud de una muestra aleatoria con n observaciones independientes también puede tomar la siguiente expresión (Aubone & Wöhler, 2000):

$$L(\theta, l_1, l_2, \dots, l_n) = f(l_1|\theta) * f(l_2|\theta) * \dots * f(l_n|\theta) \quad (2.5.3)$$

En la ecuación **(2.5.2)** se establece que la estimación de máxima verosimilitud es el valor que maximiza la función **(2.5.1)** o equivalentemente a **(2.5.3)**, también suele usarse el logaritmo neperiano de la función de verosimilitud, $\ln L(\theta)$ debido a que es más fácil trabajar con sumas que con productos; entonces, el EMV es aquel que sea la

²² Ver ecuación **(2.3)**.

solución de la siguiente ecuación, misma que es condición necesaria de extremo y en donde $\ln L(\theta)$ depende de la muestra y del vector θ (Aubone & Wöhler, 2000):

$$\frac{\partial \ln(L(\theta))}{\partial \theta_i} = 0; \quad \forall i = 1, \dots, m \quad (2.5.4)$$

Es así como **(2.5.4)** se transforma en un sistema de ecuaciones no lineales llamadas normales, cuya solución arroja a los EMV y que por su naturaleza no se puede resolver fácilmente de forma analítica, siendo esta la razón por la que se debe recurrir a métodos numéricos como Newton-Raphson. Asimismo, estos EMV tienen propiedades deseables, como que se distribuyen asintóticamente con distribución normal y es asintóticamente insesgado (Aubone & Wöhler, 2000).

Entonces, dentro del contexto de la regresión logística y tomando como punto de partida la ecuación **(2.4.1)**, se tiene que para n variables aleatorias independientes Y_1, Y_2, \dots, Y_n con probabilidad $\Pr(Y_j = 1) = \pi_j$, se deriva en la siguiente función de probabilidad conjunta, la cual pertenece a la familia exponencial (Dobson, 2002):

$$\prod_{j=1}^n \pi_j^{Y_j} (1 - \pi_j)^{1-Y_j} = \exp \left[\sum_{j=1}^n Y_j \log \left(\frac{\pi_j}{1 - \pi_j} \right) + \sum_{j=1}^n \log(1 - \pi_j) \right] \quad (2.5.5)$$

Ahora, si se considera el caso en donde todas las probabilidades sean la misma, se puede definir una nueva variable Z como la suma de Y_1, Y_2, \dots, Y_n , fungiendo como el número de sucesos en n intentos y se distribuye binomial (n, π) . Por último, tomando en cuenta N variables aleatorias independientes Z_1, Z_2, \dots, Z_N que alude al número de sucesos en N distintos subgrupos; si $Z_i \sim \text{Bin}(n_i, \pi_i)$ la función de log-verosimilitud es (Dobson, 2002):

$$l(\pi_1, \pi_2, \dots, \pi_N | z_1, z_2, \dots, z_N) = \left[\sum_{i=1}^N z_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + n_i \log(1 - \pi_i) + \log\binom{n_i}{z_i} \right] \quad (2.5.6)$$

Los EMV se obtienen resolviendo iterativamente la ecuación **(2.5.7)**, en donde el superíndice (m) indica la m – ésima aproximación, b es el vector de los parámetros estimados, en donde \mathfrak{J} representa la matriz de información y U es el vector de derivadas parciales (Dobson, 2002). Lo que se quiere calcular es la estimación de los parámetros a partir de la ya referida matriz **(2.5.8)**, en donde la primera columna se compone sólo de unos o ceros (Silva-Aycaguer, 1995).

$$\mathfrak{J}^{(m-1)}b^m = \mathfrak{J}^{(m-1)}b^{(m-1)} + U^{(m-1)} \quad (2.5.7)$$

$$\begin{bmatrix} Y_1 & X_{11} & X_{1k} \\ \vdots & \ddots & \vdots \\ Y_n & X_{n1} & X_{nk} \end{bmatrix} \quad (2.5.8)$$

En líneas generales, si el número de individuos clasificados de forma correcta es mayor, entonces se proporcionan buenos resultados cuando se prediga el valor de Y para cualquier objeto de estudio (Ferrán, 2001). Con todo esto se busca estimar, para cualquier adulto mayor mexicano de 60 años, la probabilidad de que se encuentre en informalidad laboral durante su retiro laboral, no sólo se trata de determinar el modelo multivariado, sino que se clasifique correctamente en términos de probabilidad.

2.6 Interpretación de coeficientes

El propósito de usar la regresión logística para analizar datos de respuesta binaria se condensa en la estimación de parámetros, **(ver 2.4.10)**, donde en la parte izquierda de la ecuación es el logaritmo neperiano del *odd* del resultado esperado, β_0 es la ordenada al origen y $(\beta_1, \beta_2, \dots, \beta_p)$ son los coeficientes de la pendiente de la recta, mismos que

representan el logaritmo neperiano del OR para cada covariable o factor de riesgo X (Fiuza-Pérez & Rodríguez-Pérez, 2000).

El coeficiente de regresión β_i es el cambio que se produce en la transformación *logit* por cada cambio unitario que tiene su respectiva variable independiente²³. A su vez, una buena covariable genera una curva logística con elevada pendiente, es decir, en términos de valor absoluto el valor del parámetro es alto; por el contrario, si el valor se acerca a cero, su capacidad predictora se reduce. Se busca encontrar variables con el mayor coeficiente asociado (López-Roldán & Fachelli, 2015).

Al efectuarse la transformación *logit*, se facilita la interpretación de los coeficientes, así como su efecto en la probabilidad. Si resulta ser mayor a cero, el antilogaritmo asociado es mayor a uno y el OR aumenta y por ende la probabilidad de ocurrencia también lo hace; por otro lado, ocurre lo contrario si la probabilidad de no ocurrencia es la que crece, $\beta_i < 0$. Un coeficiente igual a cero equivale a un valor de uno, lo que no provoca cambios en el *odd* (Hair *et al.*, 1999).

Siguiendo a Silva-Aycaguer (1995), existe un modo equivalente al ya mencionado: β_i indica la variación en el $\ln(odds)$ cuando X_i incrementa una unidad y las otras variables permanecen inalteradas. En el mismo sentido, si β_i es positiva es sinónimo de que la función es creciente, un coeficiente mayor a cero significa que $P(Y = 1)$ (**ver 2.4.3**) crece cuando la covariable lo hace; la función es decreciente en viceversa (Domínguez-Alonso & Aldana-Padilla, 2001).

²³ También se entienden como la medida de cambio en el OR, al estar expresados en logaritmos deben ser transformados para su interpretación (Alderete, 2006).

Ya se estableció que un coeficiente con signo positivo alude que $P(Y = 1)$ crece cuando lo hace la variable; sin embargo, el sentido cualitativo de este hecho depende de lo que representen tanto la covariable en cuestión como el suceso $Y = 1$. La función logística siempre tiene forma de S estilizada, siendo los coeficientes los que determinan cuan abrupta puede llegar a ser; por otro lado, β_0 representa la altura a la cual la curva corta al eje de las ordenadas (Silva-Aycaguer, 1995).

Por otro lado, surge un aspecto intrínseco a esta sección: una vez realizado algún ajuste (**revisar 2.8**), es posible que se quiera averiguar cuál es la influencia de las variables en el resultado; a modo ilustrativo, se puede querer saber cuál de todas las variables explicativas es la que más influye en el valor que asuma $P(Y = 1)$ y la idea de asumir que la variable de mayor peso en β_i es la que otorga mayor peso a la variable de respuesta se hace incorrecta (Silva-Aycaguer, 1995).

Finalmente, se debe tener cuidado con la interpretación de los OR o $\ln(odds)$ ajustados estadísticamente ya que ésta depende de la adecuación de los supuestos de linealidad y pendiente constante; cualquier irregularidad puede hacer que el ajuste sea inútil (Hosmer & Lemenshow, 2000). Toda esta sección se sustenta en el hecho que los coeficientes son resultado de un análisis que arroja la importancia de cada variable y obtener criterios de clasificación (López-Roldán & Fachelli, 2015).

2.7 Selección de variables

El primer paso en el proceso de análisis de cualquier modelo lineal generalizado²⁴ es que las variables se prueben en el marco de los objetivos de investigación y de criterios

²⁴ Son aquellos modelos que constan de tres partes: uno aleatorio en la variable dependiente, uno sistemático en las covariables y uno de enlace, para el presente caso, la función logística (López-Roldán & Fachelli, 2015).

teóricos que justifiquen la relación de dependencia. Se insta a proceder desde una perspectiva lógica deductiva con ayuda de un modelo definido, con la posibilidad de usar criterios estadísticos adicionales para seleccionar adecuadamente las variables finales (López-Roldán & Fachelli, 2015).

En segundo lugar, se sugiere determinar el nivel de asociación entre cada covariable de forma individual con la variable dependiente, para desechar aquellas con poca relación y que permitan agrupaciones de valores de las variables. Asimismo, se debe efectuar un análisis para contrastar multicolinealidad e indagar el nivel de asociación entre las colineales y la variable dependiente, solamente así se obtiene el mejor modelo tentativo (López-Roldán & Fachelli, 2015).

Formalmente, Hosmer & Lemenshow, (2000) proponen una serie de pasos que ayudan al completo entendimiento de lo referido. En primera instancia se debe realizar un análisis univariado; para las variables categóricas se sugiere como herramienta a la tabla de contingencia de la variable dependiente *versus* los niveles de las predictoras, mientras que para las continuas el uso de *two-sample t-test* que contraste cuales se incluyen en el modelo.

Un problema surge en el primer paso al relegar a las predictoras que tienen una asociación individual pobre pero que en conjunto no; entonces, el segundo paso es la selección de variables que entran al análisis multivariado; en la sección 3 se establece un $p - value < 0.25$ ya que, si se usa un nivel de significancia más tradicional se pueden excluir a aquellas que pueden llegar a ser importantes, aunque al mismo tiempo puede agregar a las que aporten poco (Hosmer & Lemenshow, 2000).

El tercer paso es verificar la importancia de cada covariable (**revisar 2.8.2**), así como la comparación de cada coeficiente con su modelo univariado respectivo; si no pasan estas pruebas se deben eliminar y crear un nuevo modelo, después se estima $-2LL^{25}$, prestando atención a las variables con grandes cambios, se hace de forma iterativa hasta que todas las importantes se incluyan. Se sugiere que cualquier variable no seleccionada en el modelo original se integre (Hosmer & Lemenshow, 2000).

Este penúltimo paso concluye tras haber borrado, reajustado y verificado las veces necesarias las variables propuestas hasta llegar al denominado modelo preliminar. A continuación, se revisan las potenciales interacciones, justificando con pruebas estadísticas su uso en el modelo, para que se culmine con el llamado modelo preliminar final y verificando en dónde es lineal en las variables continuas (Hosmer & Lemenshow, 2000).

En adición de la selección de variables independientes, la correcta estrategia de la construcción del modelo garantiza que la regresión logística obtenga buenos resultados. Los enfoques que se aplican a las técnicas de regresión son tres, cada uno con un propósito diferente: directo, secuencial y escalonado; las cuales no son necesariamente intercambiables porque pueden desembocar en diferentes estadísticas para los mismo datos (Stoltzfus, 2011).

Usar la estrategia más adecuada para los objetivos de la investigación es crucial (Stoltzfus, 2011); aun así, la más usual es la regresión paso a paso/escalonada (*Stepwise Regression*), misma que consiste en construir modelos sucesivos diferentes,

²⁵ -2 veces el logaritmo de la verosimilitud.

cuyo criterio de la variable es la significación estadística del coeficiente en turno; dos variantes surgen de este método: *Forward Stepwise* y *Backward Stepwise*, donde es lógico suponer que las variables irán entrando o saliendo una a una del modelo (Fiuza-Pérez & Rodríguez-Pérez, 2000).

En la *Forward Stepwise* se trata a las variables de tal forma que se comienza con un modelo sin ninguna variable, se va añadiendo a aquellas más significativas una tras otra y se detiene hasta que, con ayuda de una regla de detención preestablecida, se incluyan a todas las variables y la selección directa habrá terminado y se tendrá un modelo con variables con valores menores a un umbral (es un valor fijo, normalmente *Akaike Information Criterion* (AIC) o *Bayesian Information Criterion*) (Choueiry, 2020).

Respecto a la *Backward Stepwise* empieza con un modelo que contiene todas las variables iniciales, llamado modelo completo, continua con la eliminación de variables menos significativas una tras otra y se culmina con el cumplimiento de la regla de detención, lo que se logra cuando las variables restantes tienen un valor menor a un umbral preestablecido o, por otro lado, hasta que no quede ninguna variable en el modelo (Choueiry, 2020).

En el mismo sentido, emplear *Stepwise Selection* provee un entendimiento rápido y efectivo de un gran número de variables y ajusta simultáneamente ecuaciones logísticas, identifica a las predictoras como candidatas únicamente con motivos estadísticos, por lo que las variables deben ser examinadas detenidamente; por otro lado, el uso de interacciones en la regresión logística puede contribuir valiosamente en la construcción del modelo (Hosmer & Lemenshow, 2000).

Recomendaciones surgen en este punto, la primera de ellas es sobre las variables independientes, y es que, cuando una de ellas entra (sale) del modelo, todas sus variables *dummy* se deben de incluir (excluir) con el propósito de no agregar (eliminar) solamente sus categorías (Hosmer & Lemenshow, 2000). La segunda es que se debe evitar la inclusión del modelo una variable cuando ya hay otras que aporten la misma información (Fiuza-Pérez & Rodríguez-Pérez, 2000).

En especial, iniciar con el modelo completo tiene la ventaja de que todos los efectos de las covariables se consideran simultáneamente, esta situación cobra importancia en presencia de multicolinealidad ya que el *Backward Stepwise* las obliga a permanecer en el modelo, mientras que la *Forward Stepwise* no se podría ingresar ninguna de ellas. A su vez, su cualidad reproducible y objetiva en la reducción de variables independientes comparada con la elección manual, elimina cierta tendencia de los investigadores/analistas hacia la prueba de hipótesis (Choueiry, 2020).

Asimismo, surgen dos limitaciones en el uso de *Stepwise Regression*, y es que ésta no considera todas las combinaciones posibles de las covariables (Choueiry, 2020). La última precaución es que, a pesar del amplio uso de esta estrategia, su uso se encuentra mermado porque se basa en la selección automática de variables que tienden a usar factores aleatorios o arrojar modelos que no son lógicos bajo ciertas perspectivas (Stoltzfus, 2011).

Lamentablemente, no hay forma metódica ni lógica para prevenir estas situaciones y, considerando que el mayor peso se lo llevan las interpretaciones escuetas del analista que no tiene en cuenta los pros y contras de la estrategia al momento de elegir un método de aplicación. Naturalmente, todos los modelos necesitan de una

validación estadística para poder sacar buenas conclusiones y es aquí en donde cobra especial importancia la pericia y experiencia del investigador.

Se está tratando con procesos en los que se estiman varias ecuaciones de regresión logística a partir de diferentes modelos alternativos entre los que se elige el que mejor se ajuste a los datos para explicar la variable dependiente. En general, estos procedimientos se apoyan de *softwares* estadísticos, mismos que incorporan procesos automatizados de selección del modelo, siendo de gran ayuda en esta tarea (López-Roldán & Fachelli, 2015).

A modo de recapitulación, en la regresión logística se admite la idea de ir eliminando variables previamente elegidas y para ello se emplean pruebas estadísticas de significancia: el estadístico de Wald, la puntuación eficiente de Rao (Ferrán, 2001) y el valor del logaritmo de la verosimilitud (Hair *et al.*, 1999); todo esto con la idea de obtener el mejor ajuste mientras se minimiza el número de parámetros del modelo (Hosmer & Lemenshow, 2000).

2.8 Evaluación del modelo ajustado

Evaluar la pertinencia de las covariables que expliquen mejor Y es necesario al tratarse de cualquier modelo. Para lograrlo, se dispone de una amplia gama de pruebas estadísticas de significancia, tanto para establecer la bondad de ajuste del modelo, su capacidad explicativa o su eficacia predictiva (López-Roldán & Fachelli, 2015) para determinar si las covariables se asocian con el desenlace de interés más allá de lo que podría ocasionar el azar²⁶ (Sagaró del Campo & Zamora-Matamoros, 2019).

²⁶ Por ejemplo, puede considerarse un valor de $p < 0.05$ si se ha fijado un nivel de significación de 5%.

2.8.1 Evaluación del ajuste global del modelo

Existen medidas comparables con el coeficiente de determinación de la regresión múltiple para evaluar la capacidad predictiva del modelo de regresión logística, mismas que se han desarrollado con la intención análoga a la regresión de mínimos cuadrados ordinarios (Smith & McKenna, 2013). La primera de ellas, la prueba R^2 Cox – Snell, tiene un funcionamiento exacto como el R^2 lineal, donde mayores valores indican un mejor ajuste y cuya fórmula es (Hair *et al.*, 1999):

$$R_{C-S}^2 = 1 - \left(\frac{L(Null)}{L(Full)} \right)^{2/n} \quad (2.8.1.1)$$

En donde el cociente se conforma con las funciones de verosimilitud del modelo sólo con el coeficiente de intercepto como numerador y el modelo completo considerado como denominador, n alude al número de casos/sujetos; este estadístico se interpreta como la mejora en la media cuadrada geométrica. Sin embargo, tiene una enorme limitante y es que no tiene un rango predefinido que ayude a su comprensión y potencialmente puede exceder la unidad (Smith & McKenna, 2013).

Es la R^2 Nagelkerke soluciona la principal limitante del estadístico anterior modificando el rango entre cero y uno, tal como se muestra en la ecuación (2.8.1.2) (Hair *et al.*, 1999), cuyo divisor representa el mayor valor posible (Smith & McKenna, 2013). Es un estadístico con valores superiores en relación con el anterior; estas pruebas de significancia que muestran la variabilidad explicada arrojan valores muy bajos en relación con su análogo en la regresión lineal clásica (López-Roldán & Fachelli, 2015).

$$R_N^2 = \frac{1 - \left(\frac{L(Null)}{L(Full)}\right)^{2/n}}{1 - [L(Null)]^{2/n}} \quad (2.8.1.2)$$

La Pseudo R^2 es el estadístico más sencillo que se ha presentado porque refleja la minimización de la estimación de la regresión logística como la variación que éste mismo contabiliza (Smith & McKenna, 2013). El valor máximo que llega a alcanzar se obtiene cuando se ajusta el modelo saturado (Hosmer & Lemenshow, 2000); esta prueba es uno menos la proporción de la log-verosimilitud del modelo completo y la del modelo de sólo el intercepto (Smith & McKenna, 2013):

$$R_P^2 = 1 - \frac{LL(Full)}{LL(Null)} \quad (2.8.1.3)$$

El valor que otorga $LL(Null)$ es análogo a la suma de cuadrados totales, así como $LL(Full)$ es con la suma de cuadrados residual en el contexto de la regresión lineal; tal como funciona el R^2 lineal, si $R_P^2 = 0$ la explicación de la varianza de la variable dependiente por la covariable es nula y es perfecta cuando $R_P^2 = 1$. Hay que prestar atención en la interpretación porque la variación Pseudo R^2 es distinta en la regresión logística (Alderete, 2006).

Surge una alternativa que también mide el ajuste del modelo a través del cambio del estadístico $-2LL$, en donde la razón de verosimilitud oscila entre cero y uno; trata de calcular la verosimilitud del modelo completo $L(Full)$ y se efectúa su diferencia con el modelo nulo $L(Null)$ **(2.8.1.4)**, resultando en un estadístico que se distribuye χ_p^2 . Se obtiene un modelo perfecto de dependencia que se ajusta a los datos cuando su valor es cero, es decir, implica que $L = 1$ (López-Roldán & Fachelli, 2015).

$$-2LL = L(Null) - L(Full) \quad (2.8.1.4)$$

En el procedimiento de máxima verosimilitud se eligen las estimaciones que los resultados observados hagan más posibles, siendo que la verosimilitud es la probabilidad de los resultados observados dadas las estimaciones de los parámetros, se espera que sea una cantidad pequeña; entonces, un buen modelo es aquel con una verosimilitud grande o, por el contrario, un valor $-2LL$ pequeño (Alderete, 2006). Es equivalente al estadístico F en la regresión lineal (Ibarra & Michalus, 2010).

Las primeras tres pruebas sólo son útiles en la etapa de construcción del modelo como estadísticos para evaluar modelos competidores. Ahora bien, recordando que la bondad de ajuste se centra en la evaluación de los valores ajustados por las covariables del modelo, surgen en la regresión logística varias formas de medir los valores observados y ajustados, recordando que estos últimos dependen de su probabilidad estimada (Hosmer & Lemenshow, 2000).

Dos medidas ayudan en esta faena; el primero de ellos es el estadístico χ^2 – Pearson, mismo que se muestra en la siguiente ecuación (2.8.1.5). Surgen algunos conceptos que merecen ser esclarecidos para la comprensión de las venideras fórmulas: se consideran p variables independientes $x = (x_1, x_2, \dots, x_p)$, y_j es el número de respuestas positivas entre m_j sujetos con $x = x_j$ y J es el número de valores distintos de $x = (x_1, x_2, \dots, x_p)$ (Hosmer & Lemenshow, 2000).

$$\chi^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2 = \sum_{j=1}^J \left[\frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \right]^2 \quad (2.8.1.5)$$

La segunda medida es la Devianza, se entiende como la prueba de la razón de verosimilitud de un modelo saturado con j parámetros *versus* el modelo ajustado con $p +$

1 parámetros; la expresión se muestra en **(2.8.1.6)**, en donde el signo positivo o negativo es igual al de $(y_j - m_j \hat{\pi}_j)$ en el estadístico χ^2 – Pearson (Hosmer & Lemenshow, 2000). Ambos estadísticos indican un mal ajuste del modelo cuando sus valores de prueba son más altos, traduciéndose en diferencias altas (Stoltzfus, 2011).

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2 \quad (2.8.1.6)$$

$$\text{con } d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}$$

Por otro lado, Hosmer y Lemenshow propusieron un estadístico que mide la correspondencia de los valores reales y predichos de la variable dependiente; un mejor ajuste tiene una menor diferencia (Hair *et al.*, 1999). Se basa en el agrupamiento de percentiles con $g = 10$ grupos, los que se refieren como deciles de riesgo. Se obtiene calculando χ^2 – Pearson de tablas de $g \times 2$ de las frecuencias observadas y estimadas (Hosmer & Lemenshow, 2000):

$$\hat{C}^{27} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (2.8.1.7)$$

En donde n'_k es el número total de sujetos en el k – ésimo grupo, C_k es el número de covariables en el k – ésimo decil, $O_k = \sum_{j=1}^{C_k} y_j$ es el número de respuestas entre C_k covariables y $\bar{\pi}_k = \sum_{j=1}^{C_k} \frac{m_j \hat{\pi}_j}{n'_k}$ es la probabilidad media estimada; \hat{C} se aproxima a $\chi^2_{(g-2)}$.

Su principal ventaja es que proporciona un valor único fácilmente interpretable, pero cuya

²⁷ Si el ajuste es bueno se espera un valor alto de probabilidad, superior o igual al nivel de significación propuesto (López-Roldán & Fachelli, 2015).

desventaja radica en una potencial desviación importante en el proceso de agrupamiento de datos individuales (Hosmer & Lemenshow, 2000).

El más bajo de los deciles del estadístico está compuesto por aquellas observaciones que son menos probables de suceder, lo contrario sucede en las observaciones más plausibles; si el modelo tiene un buen ajuste los sujetos estarán en los deciles de mayor riesgo, en cambio, un modelo pobremente ajustado tiene como resultado sujetos distribuidos entre deciles de riesgo para ambos casos de la variable dependiente binaria (Stoltzfus, 2011).

La prueba de hipótesis sobre la que descansa este estadístico es que $H_0 =$ El modelo de regresión logística se ajusta a los datos, mientras que la alternativa es $H_a =$ El modelo de regresión logística no se ajusta a los datos. Como además se ajusta a una distribución Ji-Cuadrada el criterio de decisión es $\chi^2 \geq \chi^2_{(g-2)}$ para que no se rechace H_0 , si lo anterior no llega a suceder significa que el modelo ajustado no es el adecuado (Huamaní-Miranda, 2014).

Finalmente, existen otras medidas distintas que son útiles para medir el poder predictivo del modelo de regresión logística²⁸, las tablas de clasificación y las curvas ROC (*Receiver Operating Characteristic curve*). Las primeras son el resultado de la clasificación cruzada de Y con una variable dicotómica, cuyos valores se derivan de las probabilidades logísticas estimadas que fungen como herramienta para predecir la pertenencia a un grupo (Hosmer & Lemenshow, 2000).

²⁸ Estas medidas de diagnóstico ayudan a determinar si el ajuste general del modelo permanece intacto en todas las configuraciones posibles de las variables independientes.

Es importante definir un punto de corte c que sirve para compararlo con cada probabilidad estimada, si excede c se cataloga a la variable como uno, caso contrario cero, usualmente este punto es 0.5, pero no es una regla estricta. Presenta una enorme limitante cuando el modelo logístico predice la pertenencia con precisión de acuerdo con algún criterio, se puede inferir que puede proporcionar evidencia de que el modelo ajusta, situación que no siempre pasa (Hosmer & Lemenshow, 2000).

La tabla de clasificación muestra la distribución de los valores observados en contraste con los estimados obtenidos a partir del modelo. La capacidad que tenga el modelo ajustado para estimar el suceso de interés con valor de uno se entiende como sensibilidad; por otro lado, la capacidad que tenga el modelo no estime el evento de interés con valor de cero, se llama especificidad (Huamaní-Miranda, 2014), conceptos que también son útiles para las curvas ROC.

La eficacia predictiva de la curva ROC depende de un punto de corte en la especificidad y sensibilidad para clasificar el resultado de prueba como positivo, discriminando entre sujetos/objetos que experimentan el resultado de interés y los que no. El foco de atención es el área bajo la curva porque proporciona una medida de discriminación cuya regla general es la siguiente, que varía entre cero y uno (Hosmer & Lemenshow, 2000):

- ROC = 0.5; no hay discriminación
- $0.7 \leq \text{ROC} < 0.8$; discriminación aceptable **(2.8.1.8)**
- $0.8 \leq \text{ROC} < 0.9$; discriminación excelente
- ROC ≥ 0.9 ; discriminación sobresaliente

Además, el área surge de los pares ordenados conformados por los individuos en los que si ocurrió el evento con los que no ocurrió, representando la probabilidad de ocurrencia del evento de interés. La curva que resulta de la ROC tiene el indicio de que cuan más esté alejada de la diagonal principal (recta de 45° con área bajo la curva de 0.5) significa un mejor diagnóstico, siendo la curva ideal la que tiene una especificidad y sensibilidad de uno (Huamaní-Miranda, 2014).

2.8.2 Interpretación de resultados de las variables independientes

Para contrastar la bondad de ajuste de cada coeficiente, Sagaró del Campo & Zamora-Matamoros, (2019) proponen el empleo de las pruebas G de razón de verosimilitud, Omnibus, Wald y Score; a su vez, Silva-Aycaguer, (1995) y Stoltzfus, (2011) complementan este análisis con los intervalos de confianza de los OR. Para propósitos de esta investigación sólo se tratan algunos de los estadísticos mencionados por ser los de uso tradicional en la regresión logística.

El estadístico de Wald es un análogo a la prueba *t-Student* para la regresión múltiple, sirve para validar los parámetros y proporciona la significación estadística para cada coeficiente estimado, a excepción en casos donde el resultado sea excesivamente alto, se procederá a usar el estadístico de puntuación de Rao (Hair *et al.*, 1999). Parte de la hipótesis nula de que el coeficiente a nivel individual es cero, sigue una distribución normal estándar y tiene la siguiente forma (Hosmer & Lemenshow, 2000):

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} = \frac{\hat{\beta}_i}{\sqrt{\widehat{Var}(\hat{\beta}_i)}} \quad (2.8.2.1)$$

En otras palabras, se busca contrastar si los coeficientes son significativamente distintos de cero (López-Roldán & Fachelli, 2015). La ecuación anterior **(2.8.2.1)** se entiende como la razón entre el coeficiente de regresión y su error típico, sigue una distribución cuadrada Ji-Cuadrado con un grado de libertad, haciéndola idónea para datos categóricos (Alderete, 2006). Por ejemplo, los coeficientes serán significativos si tienen una probabilidad menor a 0.05 (López-Roldán & Fachelli, 2015).

La prueba de la razón de la verosimilitud está basada en el estadístico que se presenta en la ecuación **(2.8.2.2)**. Bajo la hipótesis nula de que los coeficientes $(\beta_1, \beta_2, \dots, \beta_p)$ con $\beta_i \neq$ intercepto de las covariables son iguales a cero, la distribución del estadístico es χ_p^2 (Hosmer & Lemenshow, 2000). Menores valores indica un mejor ajuste del modelo ya que se comparan las probabilidades predichas con las observadas (Hair *et al.*, 1999).

$$G = -2 \ln \left(\frac{\text{verosimilitud CON la variable}}{\text{verosimilitud SIN la variable}} \right) \quad \mathbf{(2.8.2.2)}$$

En secciones precedentes se ha remarcado la importancia del resultado que otorga e^{β_i} **(ver 2.4.6)**, a tal grado que tiene su propia interpretación y por consecuencia, se hace indispensable la construcción de un intervalo de confianza. Se muestra en **(2.8.2.3)**, en donde el error estándar $SE(\hat{\beta}_i)$ está asociado con la estimación de β_i , que es un resultado que cualquier software estadístico arroja de forma regular y toma en cuenta $\alpha = 5\%$, aunque no es norma (Silva-Aycaguer, 1995).

$$\exp[\hat{\beta}_i - 1.96SE(\hat{\beta}_i)], \quad \exp[\hat{\beta}_i + 1.96SE(\hat{\beta}_i)] \quad \mathbf{(2.8.2.3)}$$

En general, e^{β_i} se puede interpretar como la cantidad de veces en que el fenómeno de estudio es más probable que ocurra a que no suceda (Hosmer &

Lemenshow, 2000). Asimismo, el Intervalo de Confianza (IC) a 95% funge como medida de precisión para los OR en el sentido que tan probables son los resultados en la población no medida. Si IC supera la unidad no hay diferencias significativas en los datos de estudio (Stoltzfus, 2011).

Es menester hacer una recomendación cuando se quiera cuantificar la aportación de un coeficiente: efectuar de forma individual la multiplicación de la estimación del coeficiente por la raíz cuadrada de la estimación de la varianza muestral de la variable independiente asociada, el resultado que arroja se denomina coeficiente estandarizado; es especialmente útil para comparar la magnitud de los efectos debido a las variables sobre la probabilidad que se estima (Silva-Aycaguer, 1995).

Una consecuencia inmediata es que los OR ya están ajustados porque representan contribuciones individuales de las variables independientes quitando los efectos de otras variables en el modelo. A su vez, también influye si la covariable asociada es de carácter continuo o categórico; en el primer caso se debe identificar una unidad de medida (el coeficiente estandarizado) que muestre el grado de cambio asociado con la covariable (Stoltzfus, 2011).

2.8.3 Análisis de residuos

Una vez que se tenga un modelo bien ajustado, es necesario hacer un análisis diagnóstico posterior para comprobar el cumplimiento de los supuestos (**revisar 2.3**), indagar si existe alguna otra función, logística o no, que describa mejor a los datos o bien, si pudiera haber datos raros o con excesiva influencia en la estimación de los parámetros. Este análisis se realiza mediante el análisis de residuos y análisis de medidas de influencias (Sagaró del Campo & Zamora-Matamoros, 2019).

Respecto al diagnóstico del modelo, éste se hace mediante el análisis de residuos del modelo y su dominio en la estimación de los parámetros; mientras que las medidas de influencia, tal como su nombre sugiere, cuantifican la influencia que cada observación tiene sobre los parámetros o predicciones de tal forma que cuan más grande son, mayor es el poder que ejerce la observación sobre la estimación del modelo (De la Fuente-Fernández, 2011).

Lo que pretende explicar esta sección es la importancia de estudiar la falta de ajuste a nivel individual y como repercuten a nivel general, específicamente en los modelos lineales generalizados este análisis puede develar una posible mala elección de la función de enlace o si el efecto de las predictoras no es lineal; los tres tipos de residuos que se abordan tienen en común que la distribución de los residuos tiene una distribución normal con media cero y varianza menor a uno (Cañadas-Reche, 2013).

Los residuos de Pearson se expresan en **(2.8.3.1)**, en donde y_j es el número de respuestas $y = 1$ entre m_j individuos con $X_j = x \forall j = 1, 2, \dots, p$ variables con algunos valores x repetidos; este estadístico es similar con el residuo estudentizado dentro de la regresión lineal, por lo que un residuo mayor a dos en valor absoluto se interpreta como un dato atípico. El valor calculado de este estadístico es la suma de los cuadrados de sus residuos (Solano-Dávila *et al.*, 2007).

$$r_j = \frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad (2.8.3.1)$$

Los residuos de Pearson estandarizado se define en la ecuación **(2.8.3.2)**, en donde h_j es el *leverage*, es decir, el elemento asociado con la diagonal principal de la

matriz H^{29} . Por último, los residuos de deviance se encuentran definidos en la ecuación **(2.8.1.6)**, cuyo valor calculado también es la suma de los cuadrados de sus residuos; si su magnitud es mayor a cuatro en valor absoluto, entonces la observación estudiada se considera atípica (Solano-Dávila *et al.*, 2007).

$$r_{sj} = \frac{r_j}{\sqrt{1 - h_j}} \quad \text{(2.8.3.2)}$$

Al tener un gran número de observaciones, es útil mostrar gráficas de los residuos, siendo la más común la comparación de los residuos frente a los valores predichos; si ésta no llega a tener valores repetidos, debería ser una banda horizontal concéntrica en cero y si representa un buen ajuste da lugar a puntos alrededor de la recta de pendiente uno y constante cero para los datos agrupados, en caso contrario los valores predichos son cercanos a sus observaciones (Cañadas-Reche, 2013).

Finalmente, las medidas de influencias más usuales son las distancias de Cook, estadística Delta Ji-Cuadrada de Pearson, Delta Devianza y DFBETAS, los últimos develan en cambio en los coeficientes estimados cuando se elimina cada una de las observaciones y es especialmente útil para saber sobre cual variable es influyente cada observación. También cuenta con representaciones gráficas de diagnóstico³⁰ para una descripción visual de la información proporcionada (Solano-Dávila *et al.*, 2007).

²⁹ Es la matriz de predicción, el *leverage* de la observación i es el i –ésimo de la diagonal principal, $0 < h_j < 1$.

$$H = X_*(X_*'X_*)^{-1}X_*' \text{ con } X_* = V^{1/2}X; V = \text{diagonal}[\hat{\pi}_j(1 - \hat{\pi}_j)]$$

³⁰ Delta Ji-Cuadrado *versus* probabilidad estimada, Delta Devianza *versus* probabilidad estimada, Distancia de Cook *versus* probabilidad estimada, Delta Ji-Cuadrado *versus* leverage, Delta Devianza *versus* leverage y Distancia de Cook *versus* leverage.

La distancia de Cook ΔB_j **(2.8.3.3)** mide la influencia de la estimación de los parámetros; si es mayor a la unidad la observación es influyente en los valores de los parámetros estimados. El estadístico Delta Ji-Cuadrada de Pearson $\Delta \chi_{P(j)}$ **(2.8.3.4)** resulta de eliminar del modelo las observaciones de la variable x_j , si tiene valores grandes se les considera influyentes sobre los valores estimados de los parámetros (Solano-Dávila *et al.*, 2007).

$$\Delta B_j = \frac{r_{sj}h_j}{1 - h_j} \quad \mathbf{(2.8.3.3)}$$

Tanto la Delta Ji-Cuadrada de Pearson como Delta Devianza miden como el modelo estimado es afectado por sus parámetros; cuyos valores grandes indican que el modelo se mejora al excluir esa observación de los datos; la convención es que valores mayores a cuatro indican influencia sobre las estimaciones; la última prueba es el resultado de relegar las observaciones de la variable x_j , se muestra en la ecuación **(2.8.3.5)** (Solano-Dávila *et al.*, 2007).

$$\Delta \chi_{P(j)} = \frac{r_j^2}{1 - h_j} \quad \mathbf{(2.8.3.4)}$$

Todas las pruebas de significancia, tanto a nivel global como individual, tienen la capacidad de evaluar las covariables; sin embargo, es el investigador quien a través de su pericia se auxilia de éstas para lograr sus objetivos y comprender su resultado.

En el capítulo que sigue, por efectos prácticos, se emplean algunas de las pruebas estadísticas que se abordan aquí, además se hace un recuento de la aplicación, justificación e interpretación de los resultados a los que llega el modelo final de regresión logística sobre los datos de la ENIGH 2018.

CAPÍTULO III: FACTORES ECONÓMICOS Y SOCIODEMOGRÁFICOS QUE INFLUYEN EN LA PARTICIPACIÓN DE ADULTOS MAYORES EN LA INFORMALIDAD LABORAL

En este capítulo se presenta la fuente de datos empleada, a la vez que hace una descripción estadística de las características sociodemográficas y económicas que inciden en la participación del trabajo informal a través del modelo de regresión logit. Se cierra el capítulo con la aplicación del modelo e interpretación de los resultados a los que se llegaron.

El análisis de datos que se presenta como sustento de este tema de investigación se encuentra basado en la ENIGH, la nueva ola de 2018, cuyo propósito general es proporcionar información y un panorama estadístico sobre los ingresos y gastos de los hogares mexicanos en relación con su monto, procedencia y distribución, así como sus características sociodemográficas y de ocupación (INEGI, 2018).

La desagregación que ocupa esta encuesta es a nivel nacional y por entidad federativa, el esquema de muestreo es probabilístico y, al mismo tiempo, el diseño es estratificado, bietápico³¹ y por conglomerados, donde la unidad de muestreo es la vivienda, la de observación es el hogar y la de análisis son el hogar, la vivienda y sus habitantes, por lo que los resultados que obtiene la ENIGH son generalizables a toda la población con una muestra efectiva total de 87 826 viviendas (INEGI, 2018).

La ENIGH 2018 permite analizar la coincidencia entre la heterogeneidad de los adultos mayores y su relación con su trayectoria laboral y fuentes de ingresos, considerando aquellos que la literatura reconoce como factores primordiales, así como

³¹ También llamado análisis de conglomerados en dos fases. Está diseñada para descubrir agrupaciones naturales de un conjunto de datos para la generación de criterios de información, frecuencias de los conglomerados y estadísticos descriptivos por conglomerados, así como gráficas de barras, sectores y de importancia de las variables (Rubio-Hurtado & Vilá-Baños, 2016).

la descripción de características sociodemográficas de este grupo de edad; posteriormente se aborda el análisis estadístico de estos factores y su incidencia en la informalidad laboral bajo un modelo de regresión logística binaria.

3.1 Descripción de los factores seleccionados³²

Al ser un tema que ha cobrado relevancia por los movimientos demográficos de las últimas décadas, la participación económica de sectores poblacionales de edades avanzadas es prueba de que en esta etapa de la vida no hay un ingreso asegurado proveniente de otras fuentes ajenas al empleo, lo que atenta a su derecho al retiro laboral en la vejez; entonces, conocer los factores que obligan a este sector continuar en actividades remuneradas en aras de tener estabilidad económica se hace menester.

En el rubro de las características sociodemográficas, las investigaciones de Madrigal-Martínez (2010), Ramos (2016) y Millán-León (2010) coinciden en enlistar a la edad, sexo, estado conyugal, tamaño de la localidad, parentesco, alfabetismo, escolaridad y tipo de hogar como variables que definen la trayectoria personal de la población de estudio ya sea que por cuestiones de género o sociales se les adjudica responsabilidades consigo mismos o en sus círculos más cercanos.

En líneas generales, la **TABLA 3.1** presenta una descripción estadística de las variables anteriormente mencionadas y que son usadas con el mismo propósito en este trabajo académico: entender el perfil individual y del hogar de la población mayor. Se observa que 54.22% de los adultos mayores se encuentran entre los 60-69 años, así

³² Los resultados que se describen en esta sección son gracias al análisis descriptivo de la ENIGH 2018.

como que la mayoría de los hombres de edad se encuentran casados o unidos, mientras que la mujeres se enfrentan a una mayor esperanza de vida y a la viudez.

La edad es el indicador más importante porque muestra el punto del ciclo vital en el que se encuentran las personas, y tal como éste alude, marca una referencia de cómo deberían ser sus situaciones social y económica. En México, existe una prevalencia de mujeres adultas mayores, quienes al tener mayor esperanza de vida respecto a sus pares masculinos, se exponen a situaciones que pueden poner en riesgo su calidad de vida. Véase **TABLA 3.1**.

En el mismo sentido, es de esperar que exista un decremento gradual en las proporciones de adultos mayores conforme se avanza en los grupos quinquenales, producto del curso de la vida. Por otro lado y con respecto a su estado civil, 53.23% de la población de estudio se encuentra casada, siguiendole la viudez con 25.61% (de los cuales 18.91% son mujeres y el restante 6.7% son hombres) y viviendo, predominantemente, en hogares nucleares (46.46%).

Esto último se refuerza con el hecho que 63.13% de los adultos mayores son jefes de familia, siendo la mayoría hombres, a quienes se les otorga el título de proveedores económicos, situación en la que no se encuentran la mayoría de las mujeres de edad; por lo que cobra importancia la posición en la jerarquía familiar y como sean las relaciones afectivas y económicas para continuar laborando en la tercera edad.

También es importante revisar su residencia en zonas rurales y urbanas, principalmente porque existen ciertos beneficios y desventajas al residir en ciertas ubicaciones geográficas como acceso a servicios, ayudas gubernamentales, así como en diversidad en las oportunidades de empleo. Según datos de la ENIGH 2018, casi 41%

de los adultos mayores viven en localidades con menos de 2 500 habitantes, seguidos por los que viven en localidades con más de 100 mil (34.43%) (ver **TABLA 3.1**).

Además, los datos en este rubro muestran que la mayoría de las adultas mayores suelen vivir en zonas urbanas (39.58%) (con más de 100 000 habitantes), mientras que los adultos de edad, en su mayoría habita en localidades rurales (44.11%) (con menos de 2 500 habitantes), lo que, de acuerdo con Madrigal-Martínez (2010) es consecuencia inherente de la migración, principalmente femenina, del campo a la ciudad en la tercera edad.

Otro indicador que refleja la calidad de vida de las personas es el estrato socioeconómico al que pertenecen, en concordancia con las características físicas y equipamiento de las viviendas, una gran parte de la población mayor se encuentra catalogada en el estrato “medio bajo” (50.73%), seguido del “bajo” (24.98%) y un mínimo porcentaje en el grupo de clase alta (5.32%), lo que remarca las pecarias y desiguales situaciones en las que vive este grupo etario.

Asimismo, una variable que cobra peso es el alfabetismo, puesto que delimita la calidad de vida y empleabilidad, en general, la brecha de analfabetismo es pequeña para la población de estudio, permaneciendo en desventaja las mujeres con 22.95% respecto a sus pares. Todos estos factores son caldo de cultivo para definir el perfil ocupacional de una persona de edad avanzada, es decir, la ausencia o presencia de estas características los predispone a la participación en el mercado laboral.

TABLA 3.1 Características sociodemográficas de la población mayor de 60 años, 2018

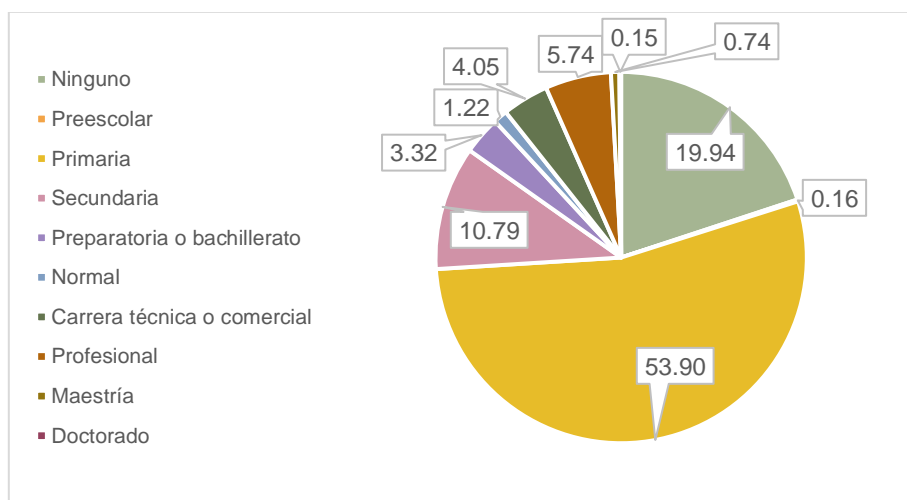
	TOTAL		HOMBRES		MUJERES	
	32485	100%	15376	100%	17109	100%
Edad						
60-64	9900	30.48%	4693	30.52%	5207	30.43%
65-69	7711	23.74%	3629	23.60%	4082	23.86%
70-74	5707	17.57%	2789	18.14%	2918	17.06%
75-79	4160	12.81%	1978	12.86%	2182	12.75%
80-84	2600	8.00%	1206	7.84%	1394	8.15%
85-89	1560	4.80%	711	4.62%	849	4.96%
90 y más	847	2.61%	370	2.41%	477	2.79%
Estado conyugal						
Unión Libre	2416	7.44%	1533	9.97%	883	5.16%
Casado(a)	17292	53.23%	9878	64.24%	7414	43.33%
Separado(a)	1773	5.46%	735	4.78%	1038	6.07%
Divorciado(a)	823	2.53%	289	1.88%	534	3.12%
Viudo(a)	8319	25.61%	2177	14.16%	6142	35.90%
Soltero(a)	1862	5.73%	764	4.97%	1098	6.42%
Alfabetismo						
Si	26073	80.26%	12890	83.83%	13183	77.05%
No	6412	19.74%	2486	16.17%	3926	22.95%
Parentesco						
Jefe de hogar	20507	63.13%	13347	86.80%	7160	41.85%
Localidad						
100 000 y más	7061	34.43%	4227	31.67%	2834	39.58%
15 000 a 99 999	2398	11.69%	1501	11.25%	897	5.24%
2 500 a 14 999	2662	12.98%	1732	12.98%	930	5.44%
Menos de 2 500	8386	40.89%	5887	44.11%	2499	14.61%
Estrato socioeconómico						
Bajo	5123	24.98%	3661	27.43%	1462	20.42%
Medio bajo	10403	50.73%	6676	50.02%	3727	52.05%
Medio alto	3891	18.97%	2304	17.26%	1587	22.16%
Alto	1090	5.32%	706	5.29%	384	5.36%
Clase del hogar						
Unipersonal	4101	20.00%	1739	13.03%	2362	32.99%
Nuclear	9528	46.46%	7506	56.24%	2022	28.24%
Ampliado	6684	32.59%	3982	29.83%	2702	37.74%
Compuesto	171	0.83%	107	0.80%	64	0.89%
Corresidente	23	0.11%	13	0.10%	10	0.14%

FUENTE: Elaboración propia con base en ENIGH 2018.

NOTA: Las variables localidad, estrato socioeconómico y clase del hogar están en función de la unidad de observación, por lo que la suma es $n = 20\ 507$, que es el número de hogares en los que habitan adultos mayores.

Igualmente sobresale el hecho que las féminas de edad tienen menores niveles de estudio, puesto que 22.24% de ellas no cuentan con alguna instrucción educativa, frente a 17.38% de los hombres en la misma situación. Predomina la educación primaria como el nivel máximo de estudios (54%) y en menor proporción se encuentran los estudios universitarios y de posgrado para la población mayor (aproximadamente 8%), veáse **FIGURA 3.1**.

FIGURA 3.1 Máximo nivel de estudios de los adultos mayores, 2018, en porcentajes.



FUENTE: Elaboración propia con base en ENIGH 2018.

En lo que respecta a los ingresos que tiene la población mayor, éstos pueden provenir de diferentes vías, aunque la literatura señala tres: transferencias sociales o por parte del Estado como pensiones y subsidios, las del mercado como salarios o retribución del trabajo y, por último, transferencias familiares por ayuda monetaria; el monto y formas de conseguirlas diferencian a los adultos de edad entre aquellos que cuentan con una o varias de estas fuentes de ingreso (Millán-León, 2010).

Para analizar la fuente de ingresos de la población mexicana mayor de 60 años y más, se tiene que comprender que el origen de éste determina su situación económica. Cabe resaltar que la mayoría de las encuestas que realiza el INEGI, incluyendo la ENIGH, usan como referencia a los hogares para medir ingresos, bienes y características que no necesariamente son exclusivas del adulto mayor, sino de la vivienda, situación que permite que hayan múltiples ingresos de un mismo integrante del hogar.

Lo anterior se refuerza con el hecho que por compartir un espacio físico implica distribuir los recursos con los cohabitantes, entonces, se parte del supuesto que los bienes e ingresos de los individuos son compartidos, sin importar quien sea el proveedor de dichos recursos (Madrigal-Martínez, 2010). Es así que al revisar los datos de los adultos mayores se devela que la principal fuente de ingresos de los hogares de los adultos de edad es el programa de 65 y más, una pensión no contributiva.

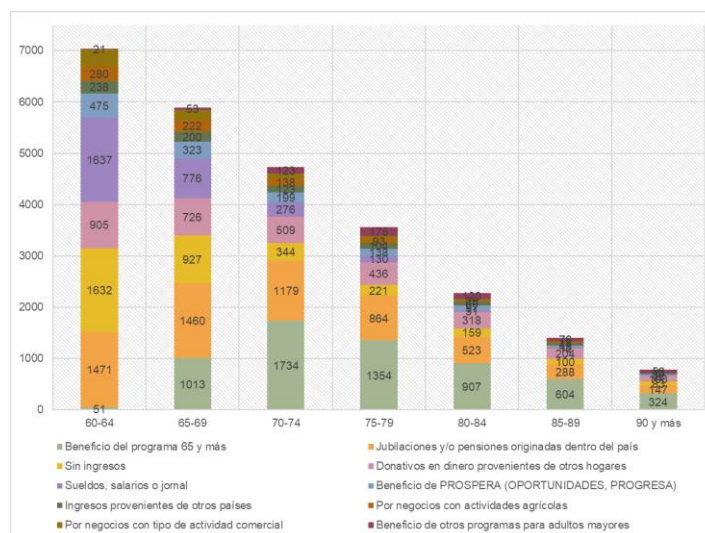
En general, las principales fuentes de ingreso de los adultos mayores son por salario, sueldo o jornal, ingresos provenientes de otro país y actividades de tipo comercial; mientras que no es común que un adulto mayor obtenga remuneración económica por rendimientos producto de algún instrumento financiero³³ o agrícola. Además, alrededor de la quinta parte de la población de estudio no cuenta con ninguna clase de ingresos, poniendolos en situaciones de vulnerabilidad. Véase **FIGURA 3.2.**

Esto no hace más que resaltar que los adultos mayores siguen participando en el mercado laboral; si bien, tener distintas fuentes de ingreso puede ser una manera de asegurar cierta cuantía para el estilo de vida de las personas, se sabe que la aparición de unas puede significar la ausencia de otras, por ejemplo, el trabajo puede aparecer

³³ Se encuentra dentro del rubro “Otros”, no se incluyó como un apartado por sí mismo.

cuando hay ausencia de pensiones y/o jubilaciones y, ante desventajas individuales o del hogar en general (Madrigal-Martínez, 2010).

FIGURA 3.2 Fuentes de ingreso de los adultos mayores, 2018 según grupo de edad (en habitantes)



FUENTE: Elaboración propia con base en ENIGH 2018.

Haciendo una desagregación por sexo, los hombres suelen tener ingresos provenientes de transferencias del Estado con jubilaciones y/o pensiones (23.37%) y por el mercado a través de sus salarios (13.75%). Por otro lado, las mujeres se benefician principalmente con programas sociales para adultas mayores (30.59%), transferencias familiares (13.55%) y remesas (3.28%); además, 16.24% carece de todo ingreso, situación que sólo afecta a 4.47% de los hombres. Ver **TABLA 3.2**.

Se distingue que el primer grupo quinquenal tiene mayores proporciones en los rubros de jubilaciones, salarios y, al mismo tiempo, son los que más carecen de ingresos, debido probablemente a su dependencia a las fuentes formales de ingreso o por alguna desventaja a la que se enfrentaron en su trayectoria laboral. A su vez, el ingreso por trabajo (salario) conforme se avanza en el ciclo vital pierde relevancia para los adultos mayores, mientras que para los beneficios sociales sucede a la inversa.

Para el grupo de 65 a 69 años el trabajo remunerado se complementa con beneficios del programa 65 y más, así como prevalece una proporción considerable de población sin ningún ingreso; mientras que para el resto de los grupos cobran mayor relevancia las transferencias provenientes del Estado y de tipo familiar en tanto disminuyen personas sin ingresos; en general, las entradas monetarias provenientes por negocios de tipo agrícola y comercial tienen menor importancia. Véase **FIGURA 3.2**.

TABLA 3.2 Fuentes de ingreso de los adultos mayores, 2018, según sexo

	HOMBRES		MUJERES	
Beneficio del programa 65 y más	2 321	15.095%	3 666	21.427%
Jubilaciones y/o pensiones originadas dentro del país	3 593	23.368%	2 339	13.671%
Sin ingresos	688	4.475%	2 778	16.237%
Donativos en dinero provenientes de otros hogares	882	5.736%	2 319	13.554%
Sueldos, salarios o jornal	2 114	13.749%	753	4.401%
Beneficio de PROSPERA (OPORTUNIDADES, PROGRESA)	116	0.754%	1 196	6.990%
Ingresos provenientes de otros países	239	1.554%	562	3.285%
Por negocios con actividades agrícolas	700	4.553%	85	0.497%
Por negocios con tipo de actividad comercial	272	1.769%	452	2.642%
Beneficio de otros programas para adultos mayores	255	1.658%	371	2.168%
Otros	4 196	27.289%	2 588	15.127%

FUENTE: Elaboración propia con base en ENIGH 2018.

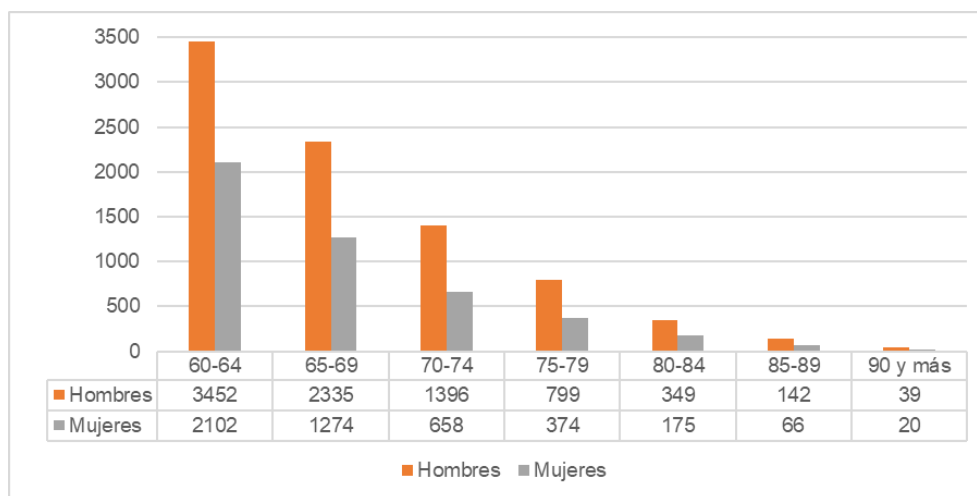
Como uno de los objetivos de este trabajo de investigación es demostrar que la mayoría de la población mexicana de 60 años o más continua laborando (independientemente del motivo y del sector de la economía en la que se emplean), en la **FIGURA 3.3** se muestra a los trabajadores de edad, situación en la que se encuentra 41% del grupo de interés³⁴; además, sigue siendo evidente la masculinización entre los que trabajan, y que, conforme aumenta la edad, la PEA disminuye drásticamente.

En el mismo sentido, una de las formas en las que se puede medir el nivel y calidad de vida de un país es a través de la cobertura en materia de seguridad social que

³⁴ En promedio, laboran 36 horas por semana.

tiene su población. De acuerdo con los datos de la encuesta, 61.47% de los adultos mayores no contribuye o ha contribuido a algún esquema de seguridad social durante su trayectoria laboral, mientras que el restante 38.5% lo hizo, en promedio, por 22 años³⁵.

FIGURA 3.3 Participación laboral de los adultos mayores, 2018, según grupo de edad.



FUENTE: Elaboración propia con base en ENIGH 2018.

Las transferencias se entienden como los ingresos en efectivo o en especie que reciben las personas sin tener la obligación de retribuirlos. La población mayor obtiene en promedio, \$14 478³⁶ en este rubro; además, haciendo una desagregación por quien otorga la transferencia, los adultos mayores reciben, en promedio, \$8 521 por jubilación/pensión, \$1 405 por bienes y servicios regalados por otros hogares, mientras que transferencias por otras instituciones públicas o privadas, \$987.

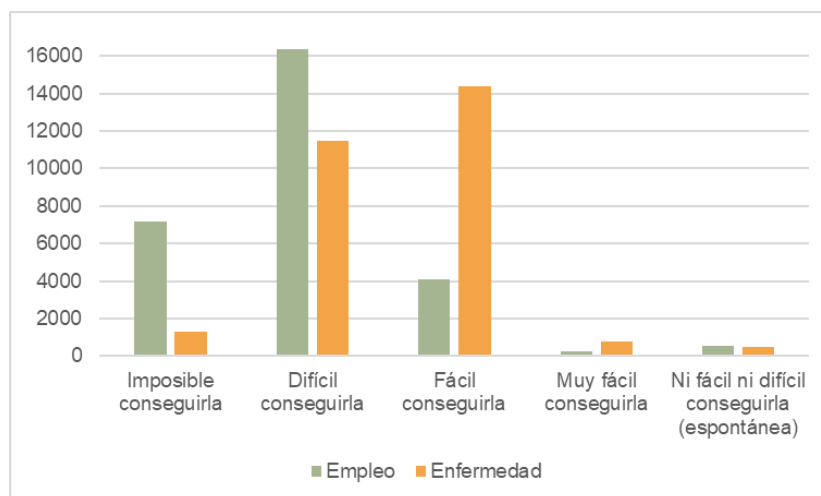
Respecto a las dificultades a las que se pueden enfrentar los adultos de edad, las dos más frecuentes son aquellas relacionadas con el empleo o la enfermedad (**FIGURA 3.4**), siendo que 50.43% percibe difícil conseguir ayuda para encontrar un trabajo,

³⁵ Bajo el Régimen 97, el trabajador necesita contar con un mínimo de 1250 semanas de cotización, aproximadamente 24 años.

³⁶ Suma de jubilación, becas, donativos, remesas, beneficios del gobierno, transferencias de otros hogares e instituciones. Todas las cifras presentadas en el rubro de transferencias son trimestrales.

mientras que solamente 0.8% lo considera sencillo; pero, contrariamente, 44.21% no se siente desprotegido ante una posible enfermedad, seguido de 35.39% que si lo ve como una adversidad.

FIGURA 3.4 Redes de apoyo³⁷ de los adultos mayores, 2018, según su percepción



FUENTE: Elaboración propia con base en ENIGH 2018.

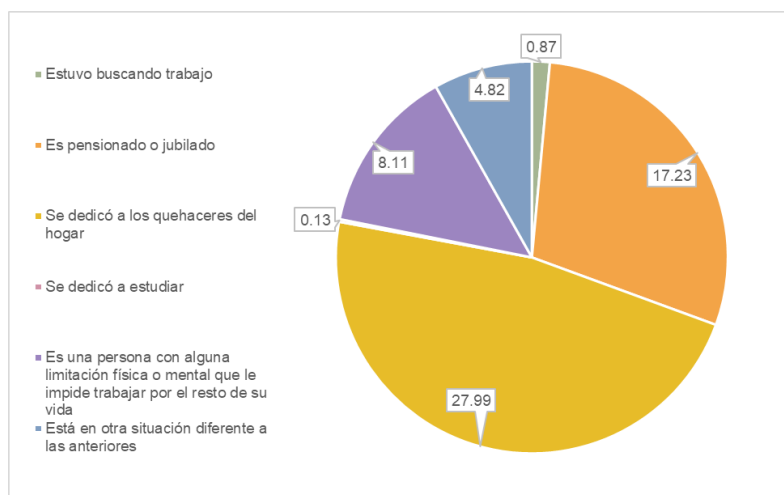
En relación con la actividad de las personas mayores que no son económicamente activas, aproximadamente 28% se dedica a labores domésticas, consagrando 20 horas por semana a realizarlas y de las cuales 93.33% son mujeres; después se encuentran las personas que son jubiladas o pensionadas (17.23%), 8.11% se tiene alguna limitación física o mental (47% son mujeres), mientras que sólo 0.8% estuvo buscando algún empleo, de los cuales 95% son hombres. Véase **FIGURA 3.5**.

En esta sección se puede visualizar la importancia que tienen cada uno de los factores que han sido seleccionados en orden de explicar la participación en la informalidad laboral en los adultos mayores, unos con mayor importancia que otros, pero que no se pueden descartar sin un sustento matemático, situación que se resuelve en el

³⁷ Refieren a la ayuda que pueden recibir las personas, refiriéndose no sólo a aquellas creados por o para los individuos sino aquellos generados de manera grupal, en donde se apela a la reciprocidad.

siguiente apartado; mientras tanto, en la **TABLA 3.3**, están las referidas variables con su codificación.

FIGURA 3.5 Actividad o situación de la PNEA, 2018 (en porcentajes)



FUENTE: Elaboración propia con base en ENIGH 2018.

3.2 Conceptualización y delimitación de la informalidad laboral.

Tradicionalmente, se percibe al trabajador informal como aquel que no hace contribución tributaria alguna y que no cuenta con prestaciones propias de la seguridad social; empero, como resultado de la evolución en donde impera la globalización y el capitalismo, surgen ciertos tipos de empleos que desafían esta conceptualización y en el que, además, existen diferentes consensos en la literatura para estudiar y comprender este fenómeno económico.

Es así como, para vislumbrar todas las dimensiones del fenómeno, se debe entender al sector informal ³⁸de la economía como al que abarca a las unidades económicas que buscan producir bienes y servicios con el fin primordial de generar empleo entre los involucrados, cuyas relaciones laborales son dadas por el parentesco

³⁸ Se debe excluir actividades relacionadas con la agricultura de subsistencia, si bien pueden ser catalogadas dentro de este sector, suelen ser objeto de estudio de sus propias políticas públicas.

o alguna relación social o personal, siendo sus características principales la baja operación, poca organización y el no registro (INEGI, 2014).

Lo anterior con el propósito de identificar eficazmente a la población que trabaja en este sector y así analizar sus condiciones de ocupación que vayan más allá de la recaudación fiscal, tomando en cuenta tanto estas disposiciones como las unidades económicas para decidir sobre todas las posibles modalidades laborales y lograr un esquema más incluyente para clasificar a la PEA empleada en dos categorías finales: formal e informal.

También cobra importancia comprender que es el empleo formal, de acuerdo con Enríquez y Galindo (2015), es aquel que incluye a trabajadores con un vínculo laboral establecido, cumpliéndose sus derechos a la seguridad social y beneficios no salariales; sin embargo, esto deja por fuera a trabajadores con contrato escrito pero que no tienen ningún esquema de seguridad social y, asimismo, a trabajadores que son catalogados como informales pero que pagan impuestos.

Por último y con el objetivo de indicar que parámetros se usan en este trabajo de investigación en lo que respecta a trabajadores informales, se tienen que asimilar los conceptos de trabajador subordinado y trabajador independiente como una manera de separarlos según la posición en la ocupación. El primer tipo alude a una persona con o sin pago que labora en una unidad económica que depende de un jefe, patrón o representante (INEGI, 2020).

TABLA 3.3 Variables que conforman el modelo

VARIABLE DEPENDIENTE				
Nombre	Etiqueta	Descripción	Clasificación	
Estatus laboral ³⁹	Y	El integrante del hogar con ciertas características que lo catalogan en la informalidad laboral.	1	Trabajador informal
			0	No es trabajador informal
VARIABLES INDEPENDIENTES				
<i>Características del hogar</i>				
Nombre	Etiqueta	Descripción	Clasificación	
Jefatura del hogar	Parentesco	Integrante del hogar al que sus miembros reconocen como el que contribuye al sostenimiento de este.	1	Es jefe del hogar
			0	No es jefe del hogar
Número de hijos	hijos_sob	Hijos sobrevivientes que tiene el/la integrante del hogar, independientemente donde residan.		
Tipo de hogar	clase_hog	Diferenciación de los hogares a partir del tipo de relación consanguínea, legal, de afinidad o de costumbre entre el jefe y los otros integrantes del hogar.	1	Unipersonal
			2	Nuclear
			3	Ampliado
			4	Compuesto
			5	Corresidente
<i>Características sociodemográficas</i>				
Sexo	Sexo	Distinción biológica que clasifica a las personas en hombres o mujeres.	1	Hombre
			0	Mujer
Edad	Edad	Años transcurridos entre la fecha de nacimiento de la persona y la fecha de la entrevista.		
Alfabetismo	Alfabetism	Situación que distingue a las personas según sepan leer y escribir un recado.	1	Sí
			0	No
Educación (escolaridad)	educa_jefe	Educación formal que recibió el integrante del hogar.	1	Sin instrucción
			2	Preescolar
			3	Primaria incompleta
			4	Primaria completa
			5	Secundaria incompleta
			6	Secundaria completa
			7	Preparatoria incompleta
			8	Preparatoria completa
			9	Profesional incompleta
			10	Profesional completa
			11	Posgrado
Estado conyugal	edo_conyug	Estado conyugal del integrante del hogar.	1	Vive con su pareja o en unión libre
			2	Está casado(a)
			3	Está separado(a)

³⁹ La construcción de esta variable se explica en la sección 2 de este capítulo.

			4	Está divorciado(a)
			5	Es viudo(a)
			6	Está soltero(a)
Tamaño de la localidad	tam_loc	Tamaño de la localidad donde reside el integrante del hogar.	1	Localidades con 100 000 y más habitantes
			2	Localidades con 15 000 a 99 999 habitantes
			3	Localidades con 2 500 a 14 999 habitantes
			4	Localidades con menos de 2 500 habitantes
Estrato socioeconómico	est_socio	Estrato socioeconómico del integrante del hogar.	1	Bajo
			2	Medio bajo
			3	Medio alto
			4	Alto
Tenencia de vivienda	Tenencia	Dominio o posesión legal de la vivienda por sus ocupantes.	1	Es rentada
			2	Es prestada
			3	Es propia, pero la están pagando
			4	Es propia
			5	Está intestada o en litigio
			6	Otra situación
Residentes	tot_resid	Número de personas que habitan la vivienda.		
Facilidad para encontrar un trabajo	redsoc_1	El nivel de dificultad o facilidad con que las personas podrían conseguir ayuda para conseguir un trabajo.	1	Imposible conseguirla
			2	Difícil conseguirla
			3	Fácil conseguirla
			4	Muy fácil conseguirla
			5	Ni fácil ni difícil conseguirla (espontánea)
Ayuda en caso de enfermedad	redsoc_2	El nivel de dificultad o facilidad con que las personas podrían conseguir ayuda para que lo(a) cuiden en una enfermedad.	1	Imposible conseguirla
			2	Difícil conseguirla
			3	Fácil conseguirla
			4	Muy fácil conseguirla
			5	Ni fácil ni difícil conseguirla (espontánea)
Tiempo de trabajo	hor_1	El tiempo, en horas, que las personas dedicaron a trabajar, por semana.		
Tiempo de cuidado	hor_4	El tiempo, en horas, que las personas dedicaron a cuidar, atender sin pago y de manera exclusiva a niños, ancianos, enfermos, discapacitados, por semana.		
Tiempo de quehacer	hor_6	El tiempo, en horas, que las personas dedicaron a los quehaceres del hogar, por semana.		
<i>Características de los ingresos</i>				
Transferencias	Transfer	Transferencias que recibe en total el integrante del hogar, es la acumulación de todas las que recibe, independientemente de la fuente que la otorga.	1	Percibe ingresos por transferencias
			0	No percibe ingresos por transferencias

Jubilación	Jubilacion	Jubilaciones, pensiones e indemnizaciones por accidente de trabajo, despido y retiro voluntario.	1	Percibe ingresos por jubilación
			0	No percibe ingresos por jubilación
Transferencias familiares	transf_hog	Bienes y servicios regalados por otros hogares.	1	Percibe ingresos por jubilación
			0	No percibe ingresos por jubilación
Transferencias institucionales	trans_inst	Bienes y servicios que las instituciones públicas o privadas otorgan a los integrantes del hogar.	1	Percibe ingresos por jubilación
			0	No percibe ingresos por jubilación
Inversiones	retiro_inv	Retiro de inversiones, ahorros, tandas, cajas de ahorro, etcétera.	1	Percibe ingresos por jubilación
			0	No percibe ingresos por jubilación
<i>Características laborales</i>				
Contribución a la seguridad social	ss_aa	Años de contribución a la seguridad social.		
Seguro popular	Segpop	Afilación al Seguro Popular.	1	Sí
			0	No
Atención médica	Atemed	Personas están o no afiliadas o inscritas a alguna institución que proporciona atención médica.	1	Sí
			0	No
Seguro voluntario	segvol_1	Tipo de seguro que el integrante del hogar ha contratado voluntariamente (SAR, AFORE o fondo de pensiones).	1	Sí
			0	No
PEA ⁴⁰	trabajo_mp	Si trabajó el mes pasado.	1	Sí
			0	No

FUENTE: Elaboración propia con base en ENIGH 2018.

En cambio, los trabajadores independientes son aquellos que dirigen su propia empresa o negocio, de tal forma que no tiene un jefe a quien rendir cuentas (INEGI, 2020), aquí se incluyen a los que trabajan por cuenta propia como a empleadores, el enfoque de interés es si la unidad económica pertenece al sector informal; si esto sucede los trabajadores (incluso subordinados) se consideran, en consecuencia, como informales o, en sentido contrario, están dentro de la ocupación formal (INEGI, 2014).

⁴⁰ No se encuentra en la base de datos final, sólo fue un referenciador para seleccionar a aquellos que tienen un trabajo.

Tomando en cuenta todas estas definiciones se consideran como trabajadores informales a aquellas personas que están en una de las siguientes circunstancias: (1) trabajadores sin un contrato que legalice su vínculo laboral, (2) trabajadores independientes o que laboren por cuenta propia sin seguridad social o, (3) trabajadores que no reciben pago en un negocio que puede ser o no del hogar. Se procede a explicar cada categoría.

El contrato laboral es un instrumento escrito en el que el empleado y la unidad económica establecen los derechos y obligaciones que rigen el vínculo; entonces, si éste es verbal o no existe, pone al trabajador en una situación de desventaja porque no conoce ni puede exigir sus beneficios, prestaciones y derechos y, a su vez, resulta ambigua su duración. Además, no es posible la retención de impuestos, por lo que se termina laborando en la informalidad.

Para la población objeto de estudio que trabaja, según datos que ofrece la ENIGH 2018, 30.8% no cuenta con un contrato que legalice su vínculo laboral; en el mismo sentido, del total que se encuentra en esta situación, 52.22% tampoco tiene prestaciones de seguridad social, en tanto que la totalidad de las personas sin contrato recibe un pago por sus servicios prestados; por otro lado, sólo 8% son aquellos con contrato, pero sin seguridad social.

Respecto a la categoría (3), INEGI (2014), cataloga como informal a todo trabajo no remunerado, independientemente si se desarrolla en empresas, instituciones formalmente constituidas o en negocios del hogar. Los resultados de la encuesta de referencia señalan que únicamente 5% de la población mayor a 60 años se encuentran

en esta situación; por otro lado, 40.61% de los adultos de edad sí reciben pago, mientras que el resto se reportan como trabajadores independientes.

Es común caer en la convención que un trabajador por cuenta propia o independiente es aquel que se relaciona con una unidad económica pertenece al sector informal; además, debido a la naturaleza de las variables de la ENIGH no es posible determinar verazmente si éstos trabajan en empresas del hogar legalmente constituidas, por lo que se dice que la persona tiene un empleo formal o informal en términos de su acceso a un esquema de seguridad social⁴¹.

Estas tres categorías buscan explicar en un sentido más amplio a la informalidad en la población mayor, sujeta a las limitaciones de las preguntas de los cuestionarios de la ENIGH, entendiéndose como informal a las actividades económicas dependientes o independientes realizadas por las personas que, de acuerdo con sus circunstancias, no pueden hacer efectivos sus derechos laborales, se enfrentan a desprotección legal y no hacen retribución tributaria.

3.3 Aplicación del modelo a los datos

La estrategia que se sigue consiste en estimar el modelo a partir de un enfoque que considera a un conjunto de factores sociodemográficos y económicos que representan a la población mexicana mayor de 60 años. Durante este proceso y al tomar como referencia la ENIGH, se deben omitir variables relacionadas con las características de las viviendas y hogares entrevistados por no estar dentro del objetivo principal de la investigación. Las variables elegidas *a priori* se encuentran en la **TABLA 3.3**.

⁴¹ De acuerdo con la ENIGH 2018, 71.51% de los trabajadores independientes no cuentan con seguridad social.

Su elección se sustenta vislumbrando la relevancia dada su naturaleza, así como su relación con el mercado de trabajo y seguridad social. Una parte del método estadístico, *stepwise regression*, también permite solucionar la cuestión de desestimar a las variables que no deben ser tomadas en cuenta en el modelo. En lo que respecta a la base de datos, se filtra, limitándola al rango de edad que va a los mayores de 60 años y a la población que reporta estar trabajando.

Del tamaño total de la muestra de 2018, 269 207 observaciones representan a 125 091 790 habitantes del país, de los cuales 32 485 (12.06%) son el grupo etario de interés y, finalmente, desagregando a la PEA, 13 181 (40.57%) cuenta con un empleo, ya sea formal o informal; esta variable, representada como “trabajo_mp”, no se incluye explícitamente en la función logística, pues se considera criterio para realizar las estimaciones únicamente a la población económicamente activa.

Finalmente, antes de aplicar formalmente el modelo estadístico y con el fin de solucionar datos vacíos, se realizó una imputación de datos. Al tratarse de los resultados de una encuesta censal, puede existir la probabilidad de encontrarse con espacios en blanco, cuyo origen pueden variar desde un “no lo sé” o frases incongruentes como respuesta, este sinfín de situaciones se conocen como datos faltantes y su existencia plantea un problema para el análisis de la información.

Específicamente, las variables “ss_aa”, “redsoc_1”, “redosoc_2”, “hor_1”, “hor_4”, “hor_6” presentan datos faltantes en una proporción importante; para solucionar esto, el paquete estadístico *RStudio* (versión 4.0.2, más adelante sólo *R*) en conjunto con la función “missForest⁴²” que se encuentra en la librería del mismo nombre, permite rellenar

⁴² Para más información consulte: <https://cran.r-project.org/web/packages/missForest/missForest.pdf>

datos de variables de cualquier tipo (continua y/o categórica) a través de una imputación no paramétrica de la estimación del error (*out-of-bag*).

Por otro lado, es importante remarcar la importancia que tiene el análisis bivariado de las covariables respecto a la variable dependiente, esto para dilucidar la calidad de la relación que tiene con cada una de estas; así, para ciertas variables, se busca una asociación inversa consistente, tal es el caso de la variable “edad”, ya que se espera que conforme se hace mayor la persona, menor probabilidad hay que se encuentre activa en el mercado laboral. También se espera asociaciones directas, por ejemplo, en “hor_1”.

De este primer contacto con la base de datos, “hor_1” y “hor_4”, aquellas relacionadas con la cantidad de horas que se usan para trabajar y a cuidar de otros, muestran consistencia con la variable de interés, probablemente influya más la actividad por si misma que el tiempo que se le dedica; otro caso es “hijos_sob”, ya que la tendencia entre la población mayor es tener una cantidad grande de hijos, como medida para asegurar un ingreso en la vejez.

Aunque también se debe considerar que esta variable está diseñada para el sexo femenino y los valores “-999” que refieren a los hombres pueden causar esta disonancia; esto puede ser una evidencia que ayude a eliminar estas covariables del modelo final. Para el caso de las variables continuas, se realiza el *t-test* de forma complementaria, a un nivel de significancia de 5%, no hay evidencia significativa para eliminar la mayoría de las variables, a excepción de “hor_4”.

Para terminar con el análisis bivariado, en las tablas de clasificación de la informalidad respecto a las variables categóricas, se ve que hay una menor proporción de adultos divorciados en “edo_conyug” ($4 = 0.6430678$), por lo que se toma como

categoría de referencia; lo mismo para "tam_loc" (1 - Localidades con 100 000 y más habitantes = 0.5822013), "est_socio" (4 - Alto = 0.4115523), "clase_hog" (5 - Corresidente = 0.6666667) y "tenencia" (3 - Es propia, pero la están pagando = 0.5).

Respecto a la *stepwise regression*, se emplea el método *Backward*; como parte de los requerimientos de la función de *R* que hará la selección final de variables, se necesita de un modelo inicial. Se empieza con un modelo que incluye a la totalidad de las covariables con algunas interacciones, se continua con un modelo que excluye a las variables que resultaron no significativas, así como otro que excluye interacciones y otro que excluye a ambas cuestiones; para elegir el que tenga menor AIC y Devianza.

En la **TABLA 3.4**, se nota que eliminar las variables que el análisis bivariado indica como no significativas aumenta el valor del AIC, así como la supresión de las interacciones, por lo que la elección del modelo preliminar es el primero, aquel que contempla todas las variables e interacciones, mismo que ayudará a la función "stepAIC" para realizar la selección de variables. Esta función comprueba como varía el AIC con la supresión de covariables y termina cuando ésta no disminuye el valor del primero.

TABLA 3.4 Resumen del modelo de *stepwise regression*

Y ~ parentesco + sexo + edad + alfabetism + edo_conyug + redsoc_2 + hor_6 + segpop + atemed + segvol_1 + hijos_sob + tenencia + tam_loc + est_socio + tot_resid + educa_jefe + transfer + jubilacion + trans_inst + hor_1 + hor_4 + parentesco: sexo + educa_jefe: jubilacion + sexo: alfabetism	
AIC	11 527
Devianza nula	14 904
Devianza residual	11 417

FUENTE: Elaboración propia.

NOTA: Las variables "edo_conyug", "tenencia" y "parentesco: sexo" no son relevantes a ningún nivel de significancia.

Para la conformación del modelo final, se toma como punto de referencia los resultados de la *Backward regression*: se usan 20 variables de las 27 originales, además

de 3 interacciones, en el segundo modelo se quitan las variables no significativas "edo_conyug", "tenencia" y "parentesco: sexo" y para el tercero se quitan las variables anteriores más "est_socio" y "educa_jefe: jubilacion". A pesar de que el tercer modelo es más parsimonioso, tiene un AIC mayor, el modelo final es el segundo. Ver **TABLA 3.5**.

TABLA 3.5 Comparación de modelos de regresión logística

	Modelo 1	Modelo 2	Modelo 3
AIC	11 527	11 539	11 594
Devianza nula	14 904	14 904	14 904
Devianza residual	11 417	11 451	11 532
<i>Modelo final</i>			
Y ~ parentesco + sexo + edad + alfabetism + redsoc_2 + hor_6 + segpop + atemed + segvol_1 + hijos_sob + tam_loc + est_socio + educa_jefe + transfer + jubilacion + trans_inst + hor_1 + hor_4 + tot_resid + jubilacion: educa_jefe + sexo: alfabetism			

FUENTE: Elaboración propia.

Una vez seleccionado el modelo final, el siguiente paso hacer una evaluación completa. Para los coeficientes, las pruebas estadísticas que se realizan son la de Wald, $-2LL$ y los OR; ninguna de estas arroja ningún resultado que represente alguna ineficacia en las covariables. Por otro lado, las pruebas para valorar el ajuste general son las de Hosmer-Lemeshow y Devianza, así como R^2 Nagelkerke, R^2 Cox – Shell y R^2 Mc Fadden. Se detallan los resultados en la siguiente sección. Véase **TABLA 3.6**.

TABLA 3.6 Resultados de la evaluación global del modelo final

AIC	11 539
Devianza nula	14 904
Devianza residual	11 417
Hosmer-Lemeshow	$p\text{-value} = 0.0154$ $X\text{-squared} = 18.899$
R^2 Nagelkerke	0.677
R^2 Cox – Shell	0.230
R^2 Mc Fadden	0.232

FUENTE: Elaboración propia.

Al tratarse la regresión logística de un método de clasificación también es interesante valorar el poder de categorización del modelo final. Esto se resuelve con una

tabla de clasificación, en cuyas filas se pone el estado real de la muestra y en las columnas la predicción (**TABLA 3.7**). Para evaluar cada observación se debe elegir un punto de corte, y si la probabilidad predicha por el modelo es mayor a éste se considera éxito y viceversa; los resultados varían dependiendo el punto de corte elegido (Cañadas-Reche, 2013).

TABLA 3.7 Tabla de clasificación para la regresión logística

	Éxito	Fracaso
Éxito	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Fracaso	Falsos Positivos (FP)	Verdaderos Negativos (VN)

FUENTE: Elaboración propia con base en (Cañadas-Reche, 2013).

Por defecto el punto de corte es 0.5, sin embargo, se busca obtener el punto de corte que maximice la clasificación correcta, la función en *R* “prediction” calcula los valores de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos para diferentes puntos de corte; mientras que con la función “performance”, se calculan varias medidas asociadas a la tabla de clasificación. El punto de corte que maximiza la tasa de clasificación correcta para la muestra es 0.515.

Otro resultado de interés es el área bajo la curva, medida que discrimina entre los sujetos que experimentan el resultado de interés y los que no, cuyos valores oscilan entre cero y uno. El área bajo la curva para la muestra es 0.818. Para efectos ilustrativos, se muestran dos gráficas de interés: la tasa de clasificación correcta bajo distintos puntos de corte y el área bajo la curva, mismos que se explican detalladamente en la siguiente sección.

Como penúltimo paso se analizan los residuos y los valores influyentes, lo que es fundamental para completar la evaluación del modelo y para detectar valores anómalos e influyentes en cada observación. El cálculo de residuos se efectúa en *R* y los resultados

se muestran en la **TABLA 3.8**, asimismo, al tener un alto número de observaciones, es conveniente hacer representaciones gráficas de los resultados, entre ellas destacan la de los residuos frente a probabilidades predichas y gráficas de diagnóstico.

TABLA 3.8 Resultados de los residuos significativos

Residuos de Pearson	693 (5.257%)
Residuos de Pearson estandarizados	696 (5.280%)
Residuos de Devianza	411 (3.118%)
Residuos de Devianza estandarizados	414 (3.140%)
Residuos estudentizados	415 (3.148%)

FUENTE: Elaboración propia.

NOTA: Los porcentajes aluden a los residuos significativos respecto al total de los datos.

Para finalizar con esta sección se realizan algunas pruebas de influencia: distancias de Cook y *leverage* y, como en las otras partes del análisis, algunas gráficas facilitan la interpretación y comprensión de los resultados que arrojan; *grosso modo*, ninguna distancia de Cook resulta significativa para los datos. También se hace la prueba para la detección de la colinealidad, los resultados se explican a detalle en la sección venidera.

3.4 Análisis de resultados

En esta sección se describen los resultados obtenidos de la aplicación del modelo de regresión logística en adultos mayores mexicanos con base en la ENIGH 2018. En la primera parte se describen los resultados desde una perspectiva estadística, así como su interpretación, mientras que en la segunda se muestra un panorama general de su situación sociodemográfica y económica de acuerdo con lo que se obtiene. Las conclusiones generales y recomendaciones se hacen en el siguiente capítulo.

TABLA 3.9 Modelo de regresión logística que identifica los determinantes del trabajo en el mercado informal en adultos mayores, 2018

		Coefficiente	Error estándar	z value	Pr(> z)	Límite inferior (2.5%)	Límite superior (97.5%)	OR	Límite inferior (2.5%)	Límite superior (97.5%)
(Intercepto)	***	2.628	0.365	7.208	0.000	1.913	3.343	13.847	6.800	28.414
parentescoTRUE	***	-0.415	0.068	-6.134	0.000	-0.548	-0.283	0.660	0.578	0.753
sexoTRUE	***	-1.022	0.252	-4.053	0.000	-1.516	-0.528	0.360	0.217	0.584
Edad	***	0.019	0.004	4.505	0.000	0.011	0.027	1.019	1.011	1.027
alfabetismTRUE	***	-1.110	0.189	-5.873	0.000	-1.480	-0.739	0.330	0.224	0.471
redsoc_2	*	-0.087	0.034	-2.541	0.011	-0.155	-0.020	0.916	0.857	0.980
hor_6	**	0.010	0.003	3.251	0.001	0.004	0.016	1.010	1.004	1.016
segpopTRUE	*	0.153	0.067	2.274	0.023	0.021	0.285	1.165	1.021	1.330
atedmedTRUE	***	-1.319	0.066	-19.877	0.000	-1.450	-1.189	0.267	0.235	0.304
segvol_1TRUE	***	-1.004	0.162	-6.193	0.000	-1.322	-0.686	0.366	0.265	0.501
hijos_sob	***	0.001	0.000	3.345	0.001	0.000	0.001	1.001	1.000	1.001
tam_loc2	*	0.202	0.080	2.527	0.011	0.045	0.358	1.223	1.047	1.431
tam_loc3	.	0.148	0.086	1.735	0.083	-0.019	0.316	1.160	0.981	1.372
tam_loc4	*	0.157	0.076	2.060	0.039	0.008	0.307	1.170	1.008	1.359
est_socio4		-0.125	0.114	-1.092	0.275	-0.349	0.099	0.883	0.705	1.104
est_socio2		0.051	0.074	0.690	0.490	-0.094	0.196	1.052	0.910	1.217
est_socio1	**	0.322	0.103	3.137	0.002	0.121	0.524	1.380	1.129	1.688
educa_jefe6	***	-0.362	0.095	-3.816	0.000	-0.548	-0.176	0.696	0.578	0.839
educa_jefe3	***	0.332	0.081	4.096	0.000	0.173	0.491	1.394	1.189	1.633
educa_jefe7	***	-1.094	0.250	-4.383	0.000	-1.584	-0.605	0.335	0.205	0.547
educa_jefe11	***	-1.631	0.276	-5.898	0.000	-2.172	-1.089	0.196	0.112	0.332
educa_jefe8	***	-0.859	0.139	-6.194	0.000	-1.131	-0.587	0.423	0.323	0.556
educa_jefe10	***	-1.170	0.130	-9.000	0.000	-1.425	-0.915	0.310	0.240	0.400
educa_jefe5	**	-0.458	0.172	-2.659	0.008	-0.795	-0.120	0.633	0.453	0.891
educa_jefe1	**	0.398	0.124	3.207	0.001	0.155	0.641	1.488	1.169	1.902

educa_jefe9	***	-0.908	0.231	-3.927	0.000	-1.361	-0.455	0.403	0.256	0.634
educa_jefe2		-0.627	0.682	-0.920	0.357	-1.964	0.709	0.534	0.158	2.492
transferTRUE	***	0.292	0.071	4.111	0.000	0.153	0.431	1.339	1.165	1.538
jubilacionTRUE	**	-0.277	0.108	-2.577	0.010	-0.488	-0.066	0.758	0.614	0.936
trans_instTRUE	**	-0.148	0.054	-2.758	0.006	-0.254	-0.043	0.862	0.776	0.958
hor_1	***	-0.007	0.001	-5.610	0.000	-0.010	-0.005	0.993	0.990	0.995
hor_4	***	-0.016	0.003	-6.007	0.000	-0.021	-0.011	0.984	0.979	0.989
tot_resid	**	0.038	0.012	3.151	0.002	0.015	0.062	1.039	1.015	1.064
educa_jefe6: jubilacionTRUE	.	0.258	0.156	1.650	0.099	-0.048	0.564	1.294	0.953	1.757
educa_jefe3: jubilacionTRUE	**	-0.424	0.136	-3.109	0.002	-0.691	-0.157	0.654	0.501	0.855
educa_jefe7: jubilacionTRUE	***	1.522	0.418	3.641	0.000	0.703	2.342	4.583	2.033	10.524
educa_jefe11: jubilacionTRUE		0.112	0.401	0.280	0.780	-0.674	0.899	1.119	0.505	2.448
educa_jefe8: jubilacionTRUE	.	0.396	0.227	1.744	0.081	-0.049	0.841	1.486	0.951	2.318
educa_jefe10: jubilacionTRUE	**	0.636	0.202	3.148	0.002	0.240	1.032	1.889	1.271	2.807
educa_jefe5: jubilacionTRUE		0.434	0.314	1.383	0.167	-0.181	1.049	1.544	0.834	2.861
educa_jefe1: jubilacionTRUE		-0.261	0.196	-1.333	0.182	-0.644	0.123	0.770	0.526	1.132
educa_jefe9: jubilacionTRUE		0.324	0.368	0.880	0.379	-0.398	1.045	1.383	0.668	2.836
educa_jefe2: jubilacionTRUE		0.012	1.145	0.010	0.992	-2.232	2.256	1.012	0.094	9.240
sexoTRUE: alfabetismTRUE	***	0.895	0.205	4.374	0.000	0.494	1.296	2.448	1.657	3.704

FUENTE: Elaboración propia.

NOTA: Códigos de significancia: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1

En la sección anterior se indica que la elección final del modelo es el más parsimonioso y con menor AIC; sin embargo, para fines ilustrativos, conviene introducirlo de nuevo formalmente. Entonces, el mejor modelo de regresión logística que explica la informalidad laboral en adultos mayores se encuentra conformado por un conjunto de variables sociodemográficas y económicas que se describen en la **TABLA 3.9**, en donde se muestra tanto el valor del coeficiente como el OR correspondiente.

Evaluar los coeficientes de cualquier modelo lineal permite discernir la calidad de las variables que explican el fenómeno. La prueba de Wald no representa ningún problema debido a que para elegir las covariables del modelo final debieron ser relevantes a distintos niveles de significancia (ver **TABLA 3.9**); resalta el hecho que algunas categorías de algunas variables no lo son, sin embargo, permanecen en el modelo porque contienen otras que si lo son y su exclusión aumenta el valor del AIC.

Se debe tener presente que el contraste de hipótesis de cada parámetro bajo el estadístico de Wald es cero frente a que no lo sea. A un nivel de significancia $\alpha = 0.05$ los parámetros asociados al tamaño de localidad (categoría 3), estatus socioeconómico (categorías 2 y 4), educación (categoría 2) y en “educa_jefe: jubilacionTRUE” (categorías 1, 2, 5, 6, 8, 9, 11) no son significativamente distintos de cero, al tener valores mayores en mayor absoluto al punto crítico $\frac{z_{\alpha}}{2} = 1.96$.

Lo anterior pudiera suponer que hay suficiente evidencia estadística para eliminarlas del modelo, pero conviene usar el contraste condicional de razón de verosimilitud ($-2LL$) para medir el impacto que supondría contemplar o no a las covariables en el modelo final, los resultados se muestran en la **TABLA 3.10**. De acuerdo

con ésta, las variables relacionadas al sexo y tamaño de localidad no pasan esta prueba, por lo que se podría sugerir eliminarlas del modelo.

Asimismo, los intervalos de confianza pueden suponerse como un análisis complementario para discernir que variables deben ser excluidas, si el intervalo de confianza incluye el cero, implica que con $\alpha = 0.05$ no se podría concluir que la variable es significativa; en esta situación se encuentran nuevamente las mismas categorías. Esto implicaría eliminar por completo a la variable e interacción en cuestión, y como se ha prestablecido, hacerlo aumenta el valor del AIC e indicar un ajuste global pobre.

Las pruebas de ajuste global que se hacen para validar al modelo se muestran en la **TABLA 3.6**, aquí destaca que la prueba Hosmer-Lemeshow no pasa la prueba de hipótesis, lo que infiere que no hay suficiente evidencia estadística para aceptar la consistencia del modelo; aun así, un *p-value* bajo es bastante usual cuando se trata de muestras grandes. Allison (2013) asegura que esta inconsistencia se puede resolver aumentando o disminuyendo la cantidad de grupos en los que se divide la muestra.

Para efectos de la muestra, la cantidad de grupos $11 \leq g \leq 13$ para que se rechace la hipótesis alternativa y se pueda asegurar que hay evidencia estadística para aprobar el ajuste global del modelo. Aun así, no hay evidencia teórica que señale en cuántos grupos se debe hacer la división para obtener el resultado deseado y, claramente, éstos no deben depender de cambios arbitrarios en las pruebas estadísticas que se usan para tener validez (Allison, 2013).

A su vez, al tratar con la Devianza, *R* muestra por defecto los resultados de la Devianza nula (solo el intercepto) y la Devianza residual (todas las covariables). Para que el modelo cuente con un buen ajuste, la Devianza residual (11 451) debe ser menor que

la nula (14 904), ya que valores menores en la Devianza indica que el modelo predice la variable dependiente con mayor precisión. Por otro lado, a través de las medidas R^2 , se puede averiguar que tan bien se puede predecir Y basado en las covariables.

TABLA 3.10 Contraste condicional de razón de verosimilitud

		LR Chisq	Df	Pr(>Chisq)
Parentesco	***	38.38	1	0.000
Sexo		1.54	1	0.215
Edad	***	20.63	1	0.000
Alfabetism	***	22.28	1	0.000
redsoc_2	*	6.44	1	0.011
hor_6	***	10.83	1	0.001
Segpop	*	5.15	1	0.023
Atemed	***	409.24	1	0.000
segvol_1	***	41.06	1	0.000
hijos_sob	**	10.57	1	0.001
tam_loc	.	7.8	3	0.050
est_socio	***	16.95	3	0.001
educa_jefe	***	218	10	0.000
Transfer	***	16.71	1	0.000
Jubilacion	***	18.4	1	0.000
trans_inst	**	7.58	1	0.006
hor_1	***	31.43	1	0.000
hor_4	***	35.65	1	0.000
tot_resid	**	10.08	1	0.001
educa_jefe: jubilacion	***	61.54	10	0.000
sexo: alfabetism	***	21.62	1	0.000

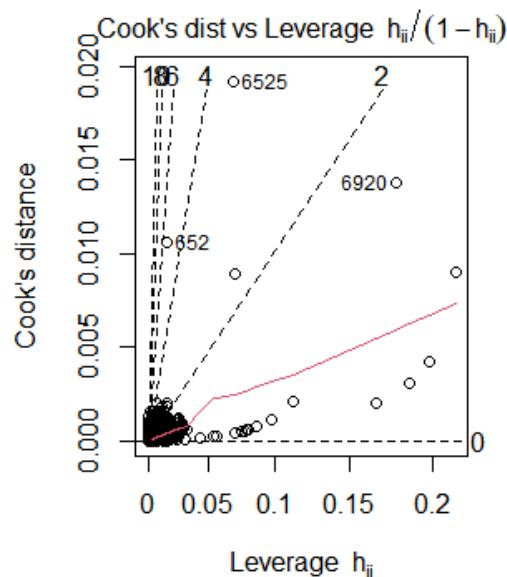
FUENTE: Elaboración propia.

NOTA: Códigos de significación: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1

Las pruebas estadísticas R^2 Cox – Shell y R^2 Mc Fadden muestran un ajuste pobre al tener valores oscilantes a 23% porque éstos suelen ser más efectivos cuando se trata con datos agrupados; en especial, porque R^2 Cox – Shell no tiene una escala que permita discernir si es un buen ajuste, es por eso por lo que la prueba R^2 Nagelkerke, al estar acotada entre cero y uno, permite hacer una valoración más efectiva. Indica un ajuste moderado (0.6772) de los datos.

El principal propósito de analizar los residuos en cualquier modelo lineal generalizado responde al de aislar los puntos en los que se tiene un mal ajuste o ejerzan una influencia excesiva en los resultados. En la **FIGURA 3.6**, se muestra que las observaciones 652, 6 525 y 6 920 pueden presentar una posible influencia; cuando se analiza individualmente cada distancia de Cook, se busca que ésta sea menor que uno. La salida computacional que ofrece *R* avala que ninguna distancia es significativa.

FIGURA 3.6 Distancias de Cook vs Leverage



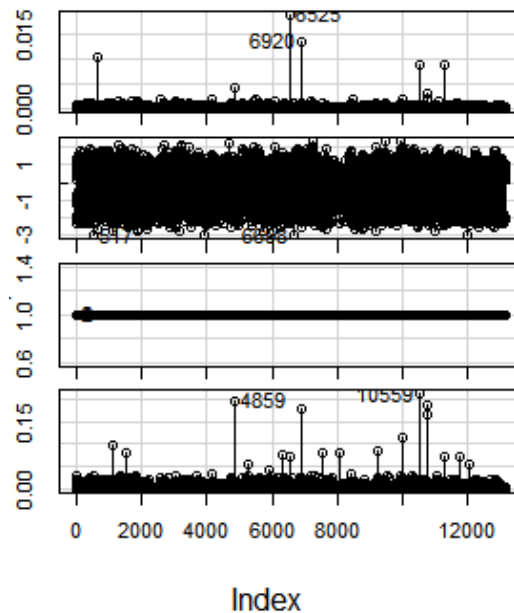
FUENTE: Elaboración propia.

En la **FIGURA 3.7** se muestra un resumen de gráficos importantes sobre las medidas de influencia. La cuarta gráfica muestra otra forma de averiguar si una observación es potencialmente influyente: a través del peso que tenga cada observación con su predicción, si éste llega a ser alto significa puede intervenir en su pronóstico. Las observaciones que pueden presentarse problemáticas son la 4 859 y 10 559; aun así, las distancias de Cook no indican que esto sea realmente un problema.

En el mismo sentido, en la gráfica de los residuos estudentizados se pueden identificar fácilmente a aquellos valores atípicos porque se espera que la mayoría de los

residuos oscilen entre $(-3,3)$, las observaciones que se encuentran por fuera de este rango son la 517 y 6 698. Todos estos valores atípicos (517, 652, 4 859, 6 525, 6 698, 6 920 y 10 559) explican el hecho que el porcentaje de residuos de Pearson (incluyendo los estandarizados) sea mayor a 5%. Véase **TABLA 3.8**.

FIGURA 3.7 Gráficas de diagnóstico



FUENTE: Elaboración propia.

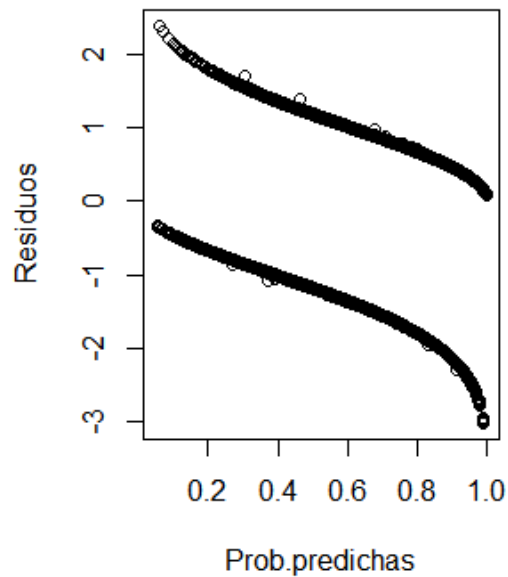
NOTA: Las gráficas, de arriba hacia abajo, son las distancias de Cook, residuos estudentizados, p-valores Bonferroni y *hat*-valores.

Asimismo, para hacer más completo el análisis, se incluye una gráfica que muestra la relación entre los residuos y los valores predichos (ver **FIGURA 3.8**), misma que no es fácil de interpretar porque Y solo puede ser cero o uno para cada observación por lo que la gráfica muestra un patrón definido; además, como los valores predichos se mueven en este rango, resulta en un grado de discrepancia más alta. Por eso se hace hincapié en estudiar datos atípicos, sobre todo al tratarse de una regresión logística.

Por último, otro resultado de interés en este rubro es comprobar la ausencia de la multicolinealidad en las covariables (ver **TABLA 3.11**); es menester hacer algunas

aclaraciones sobre los resultados. Se debe hacer notar que se espera encontrar una multicolinealidad moderada, sobre todo al tratarse con interacciones entre las variables independientes y con una muestra elevada.

FIGURA 3.8 Residuos de Devianza estandarizados vs probabilidades predichas



FUENTE: Elaboración propia.

Los valores normales de VIF tienen pequeñas variaciones superiores a uno, sin despegarse demasiado de ese valor, como se observa en la salida que ofrece *R* cuando se aplica la función de la librería *car* (tercera columna de la **TABLA 3.11**); estos valores cercanos a uno aluden a una variable poco correlacionada con las demás. Al tratarse con variables categóricas se usa el factor de inflación de la varianza generalizada ($GVIF^{1/2p}$) por ser más apropiada⁴³.

Entonces, para rechazar la presencia de multicolinealidad entre las covariables se deben presentar valores que también tiendan a la unidad. A pesar de esto, para comprobar la ausencia de multicolinealidad, se esperan valores menores a cinco, incluso

⁴³ Para más información al respecto consúltese (Cañadas-Reche, 2013).

diez, dependiendo de la literatura consultada; para efectos de la presente investigación se maneja un escenario más conservador. Las variables “sexo”, “alfabetism”, “hijos_sob”, “jubilacion” y las interacciones muestran valores altos, pero menores a cinco.

Lo anterior sucede debido a las interacciones de primer grado que hay en los factores principales que están involucradas y al considerar a las primeras, se introducen en el modelo un conjunto de regresores muy correlacionados. La otra variable, “hijos_sob”, tiene un valor de GVIF que supera tres, esto se justifica con que es una variable hecha para mujeres y a los hombres se le categoriza con “-999”, lo que podría explicar la multicolinealidad.

TABLA 3.11 Resultado de la prueba de multicolinealidad (VIF)

	GVIF	Df	$GVIF^{1/(2*Df)}$
Parentesco	1.302	1	1.141
Sexo	23.363	1	4.834
Edad	1.188	1	1.090
Alfabetism	5.406	1	2.325
redsoc_2	1.030	1	1.015
hor_6	1.751	1	1.323
Segpop	1.840	1	1.356
Atemed	1.931	1	1.389
segvol_1	1.014	1	1.007
hijos_sob	9.546	1	3.090
tam_loc	2.269	3	1.146
est_socio	2.622	3	1.174
educa_jefe	176.096	10	1.295
Transfer	1.375	1	1.173
Jubilacion	4.843	1	2.201
trans_inst	1.133	1	1.064
hor_1	1.194	1	1.093
hor_4	1.434	1	1.198
tot_resid	1.077	1	1.038
educa_jefe: jubilacion	303.419	10	1.331
sexo: alfabetism	17.524	1	4.186

FUENTE: Elaboración propia.

Anteriormente se han mencionado las medidas R^2 para indicar la bondad de ajuste del modelo logístico. Sin embargo, la mejor forma de validar el ajuste es evaluar su capacidad de separar los grupos basadas en las probabilidades estimadas. Se observa la clasificación que aparece en la **TABLA 3.12**, cuyos resultados muestran que, de los 9 849 adultos clasificados como informales, el modelo acierta 8 900 de ellos, mientras que los que se catalogan como formales (3 332 personas), se acierta 1 544.

TABLA 3.12 Tabla de clasificación para el modelo logístico final

	Éxito	Fracaso
Éxito	8 900 (90.36%)	949 (9.63%)
Fracaso	1 788 (53.66%)	1 544 (46.33%)

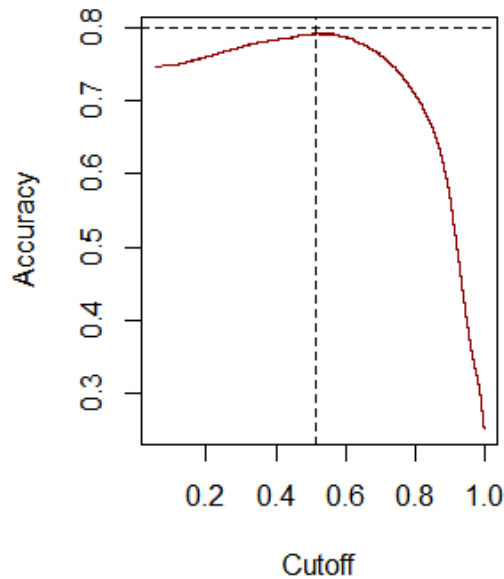
FUENTE: Elaboración propia.

NOTA: El punto de corte es 0.515

Entonces, el modelo tiene una tasa de clasificación correcta global de 79.23%, en otras palabras, la probabilidad de acertar en el estatus laboral de los adultos mayores con esta función logística es de 79.23%. El resumen de la tabla de clasificación del modelo es el siguiente: sensibilidad (VP) 90.36%, especificidad (VN) 46.33%; mientras que las tasas de Falsos Positivos (FP) y Falsos Negativos (FN) (véase **TABLA 3.7**) son, respectivamente, 53.66% y 9.63%.

El punto de corte que se usa para la construcción de la tabla de clasificación y área bajo la curva es aquel que maximiza la tasa de clasificados correctamente, tal como se muestra en la **FIGURA 3.9**, en que se muestra cómo cambia esta tasa según distintos puntos de corte. Cabe destacar que si se usa un punto de corte más tradicional (0.5), la tasa de clasificados correctamente no varía mucho (79.09%). Otra gráfica de interés es la curva ROC, la que representa la fracción de FP frente a la fracción de VP.

FIGURA 5.9 Tasa de clasificaciones correctas, dados distintos puntos de corte

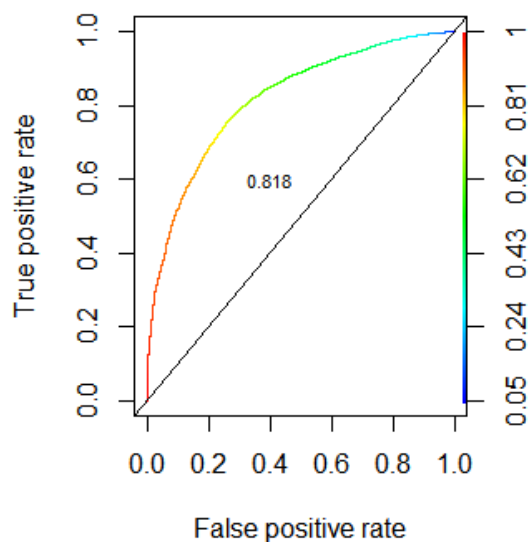


FUENTE: Elaboración propia.

En este gráfico (**FIGURA 3.10**) se traza una línea recta entre los puntos (0,0) y (1,1) como referencia para describir al diagnóstico incapaz de discriminar entre observaciones que presenten el suceso de interés y los que no, debido a que cada punto de esta recta tiene la misma especificidad y sensibilidad; entonces, se busca que la curva ROC se encuentre lo más alejada posible de la recta de 45° para que aumente su capacidad discriminativa.

Es así que se tiene un buen modelo cuando esta curva se acerca al borde superior izquierdo y acapara la mayor cantidad de área, la que sirve como parteaguas para discernir el nivel de discriminación. El modelo logístico final tiene una discriminación excelente al contar con un área bajo la curva de 0.818, la cual se entiende como la probabilidad de que a partir de un par de individuos, uno informal y otro formal, sean clasificados correctamente.

FIGURA 3.10 Curva ROC para el modelo final



FUENTE: Elaboración propia.

Para terminar el capítulo, se examinan los coeficientes y sus correspondientes OR; en líneas generales, los primeros indican una relación positiva o negativa con la variable dependiente, dependiendo de su signo y tamaño, tal como ocurre en los mínimos cuadrados ordinarios; mientras que, la también denominada razón de momios mide la cantidad de veces que ocurre una situación de interés en comparación con las veces que no sucede.

Sin embargo, cobra mayor importancia como las covariables afectan la probabilidad de que el adulto mayor se encuentre laborando en la informalidad laboral. Así, si el $OR > 1$ alude a una asociación positiva, mientras que si $OR < 1$ señala una inversa o negativa y, si es igual o tendiente a uno, es sinónimo de la ausencia de la relación entre las variables, véase **FIGURA 3.11** (Ramos, 2016). Los resultados de los OR se encuentran en la **TABLA 3.9**.

FIGURA 3.11 Interpretación gráfica de los OR



FUENTE: Elaboración propia con base en (Ramos, 2016).

El estatus de jefe de familia tiene una importante concepción cultural y le atribuye al individuo la obligación de contribuir económicamente al sustento económico y al liderazgo de su hogar; es así como tener esta responsabilidad (en relación con no tenerla) disminuye 1.51 veces la posibilidad de que el adulto mayor trabaje en la informalidad, esto podría indicar que, al tener impuesto este rol, se tenga preferencia a trabajos más estables.

Asimismo, los hombres de edad tienen 73.47% de probabilidad de no trabajar informalmente respecto a sus pares femeninos, cuya decisión de emplearse puede surgir de un abanico de causalidades que van desde la viudez, la jefatura del hogar hasta su propio sustento (Ramos, 2016); situación que las pone en especial desventaja al tener poca o nula trayectoria profesional, así como escaso nivel educativo (22.24% no cuenta con algún grado de instrucción).

Lo anterior no hace más que confirmar que los roles de género predisponen a los adultos mayores a cierto nivel de vida, estatus social y responsabilidades, los varones se ven forzados a ser el principal sustento económico tanto de su cónyuge como de su familia, incluso durante la senectud; mientras que las féminas sólo se encuentran en esta situación cuando su pareja se encuentra ausente.

En lo que respecta al nivel de instrucción y al alfabetismo, se encuentra que al saber leer y escribir un recado se hace 3.03 veces menos viable la informalidad; además,

los adultos mayores con educación superior tienen menos posibilidades de ser informales, por ejemplo, contar con un posgrado disminuye 5.10 veces tener un trabajo informal como principal fuente de ingresos, mientras que tener educación básica (escolaridad promedio de la población mayor) solo baja 1.4 veces.

Es importante destacar que tradicionalmente se infiere que contar con estudios universitarios se incrementan las oportunidades laborales, implica acceder a un estatus en donde hay mayores ingresos y una mejora en la calidad de vida; a su vez, se ven ampliadas las expectativas y existe una toma de decisiones más racional, en el que ya no importa solo el presente, sino como se viva el futuro.

Por otro lado, el tamaño de la localidad es una variable que sólo es significativa para la categoría dos (para las personas que vivan en zonas con 15 000 – 99 999 habitantes); es decir, residir en una localidad con esta cantidad de habitantes incrementa 54.95% la probabilidad de laborar en la informalidad, respecto a personas de la tercera edad que habitan en lugares con más de 100 000 habitantes. Esto no indica que en las zonas rurales no trabajen en actividades informales.

Concerniente a las características laborales, estar afiliado a alguna institución que proporcione atención médica disminuye 3.74 veces la posibilidad de emplearse en la informalidad. Lo mismo ocurre con la contratación voluntaria de un instrumento financiero como AFORE, SAR⁴⁴ o fondo de pensiones, decremента 2.72 veces o, en sentido contrario baja 73.11% la probabilidad. Estas covariables son las más significativas del modelo, lo que expresa la importancia de la seguridad social en la población mayor.

⁴⁴ Sistema de Ahorro para el Retiro.

El ingreso por concepto de jubilación disminuye 1.31 veces el suceso de interés, este OR prácticamente indiferente puede demostrar que el dinero que perciben los adultos mayores en este rubro es raquítico, por lo que se enfrentan a la desventura del mercado laboral. Lo mismo sucede con las transferencias institucionales, contar con un apoyo de este tipo apenas disminuye 53.48% la probabilidad de que la población de edad sea empleada en la informalidad laboral.

En general, contar con alguna transferencia de cualquier tipo, sólo aumenta 1.31 veces la posibilidad de la informalidad. De igual manera, el hecho de tener una fuerte red familiar (transferencias familiares) se torna indiferente debido a que los OR de las variables “redsoc_2” y “tot_resid” tienen valores tendientes a uno, aunado con que “transf_hog” e “hijos_sob” no se encuentren en el modelo final, puede sugerir que las redes sociales más importantes son las que provee el Estado.

En el mismo sentido, el tiempo que se dedique a las actividades económicas es indiferente, es decir, importa más el trabajo; esto indica que no existe tanta flexibilidad en los horarios laborales, característica sobresaliente de los empleos informales. Por último, el análisis de las interacciones arroja que el hecho de ser un hombre alfabeto aumenta 2.44 veces la posibilidad de laborar, así como ser jubilado con preparatoria incompleta (4.58 veces).

Entonces, ser hombre, jefe de familia, tener jubilación, AFORE/SAR o fondo de pensiones y poseer un alto nivel de estudios son los factores económicos y sociodemográficos que alejan a la población mayor de trabajar en un segmento de la economía tan desprotegido como lo es el informal, en donde es posible la necesidad de

reformular políticas públicas en materia de seguridad social para prevenir que las próximas generaciones tengan este desenlace.

En el siguiente capítulo se muestran las conclusiones y recomendaciones, haciendo énfasis en las condiciones laborales, de salud y bienestar físico, se sugieren algunas proposiciones en materia de envejecimiento activo que se ajustan a lo que propone la Organización Internacional del Trabajo.

CONCLUSIONES Y CONSIDERACIONES FINALES

“La vejez es la única cosa que llega sin tener que esforzarnos para conseguirla”

Cicerón

A la luz de la presente investigación se demuestra que la seguridad económica en la senectud tiene estrecha relación con la calidad del mercado laboral y la fortaleza del sistema de pensiones; si bien cualquier desventaja sociodemográfica de la población puede mermar su nivel de seguridad económica, los adultos mayores cuentan con características que plantean desafíos que solo pueden afrontar, pero no resolver.

El dilema se centra en que las fuentes formales de la seguridad económica (mercado y Estado) suelen excluirlos por enfocarse en otros sectores demográficos, sin darse cuenta con que las próximas generaciones también se enfrentarán a vicisitudes como pobreza, pocos ingresos y problemas de salud. Cada vez será más común encontrar adultos mayores en el mercado laboral en actividades poco remuneradas que sean su principal fuente monetaria de ingresos.

Esto se agrava si se considera la inserción al mercado laboral como una estrategia para compensar la desprotección social que enfrentan los adultos de edad. Con todo esto, se devela que los sistemas de protección social de países latinoamericanos como México se vuelven obsoletos, al no considerar factores como el incremento en la esperanza de vida, nuevas dinámicas familiares, la mayor empleabilidad que han experimentado las mujeres, así como mayor competitividad en la demanda laboral.

El trabajo remunerado es la base de las condiciones de vida, pero un porcentaje alto (74.72%⁴⁵) de adultos de edad trabaja en condiciones de informalidad, en donde sus

⁴⁵ Con base en las tres categorías que se mencionan en el apartado tres del capítulo anterior.

derechos laborales pasan a segundo plano. Este grupo etario es especialmente vulnerable y propenso a este tipo de empleos, por lo que la jubilación, pensión y trabajo formal y remunerado son las fuentes más importantes de ingreso en la vejez. Se hace notar la necesidad de crear empleos con valor social que garanticen seguridad económica.

Las pensiones no contributivas son una de las variables consideradas en el modelo, siendo las transferencias institucionales (“trans_inst” en el modelo) las que evitan que el adulto mayor trabaje en la informalidad; sin embargo, se encuentran un tanto limitadas porque el monto mensual que reciben es apenas \$600. En el mismo sentido, aspectos como ser mayor de 80 años, seguir laborando, tener baja escolaridad, ser mujer y estar afiliado al seguro popular son características entre quienes reciben este tipo de pensión, de acuerdo con CONSAR (2018).

Resultados similares encuentra Madrigal-Martínez (2010), quien destaca que las transferencias institucionales y familiares auxilian a adultos mayores con poco nivel de instrucción educativa y con características que pueden ser desventajosas como ser indígena, mujer o vivir en una zona rural, mismas que se relacionan con ausencia de ingresos por pensión o trabajo; es decir, estas características limitan el acceso a una pensión y oportunidades de emplearse a estos adultos mayores.

Sin embargo, en ciertas zonas geográficas este tipo de apoyo institucional carece de importancia, tal como lo señala Millán-León (2010), ya que los apoyos por programas de gobierno no son estadísticamente importantes dentro de su modelo de regresión binaria, por lo que no condicionan la participación económica en la vejez en el Estado de

México; en el lado opuesto se encuentra el beneficio por pensión: en mayor medida están activos quienes no cuentan con este ingreso.

En líneas generales, Ramos (2016) encuentra que las pensiones y jubilaciones tienen tal importancia en la estabilidad económica en los adultos de edad que tener ingresos de este rubro significa que esta población tiene menores posibilidades de incorporarse en el mercado laboral (ya sea formal o informal o que sea hombre o mujer), lo que, a su vez, implica que la seguridad social es relevante en todos los puntos del ciclo vital de los individuos.

Es así como los factores sociodemográficos como ser hombre, tener un mayor nivel educativo y contar con la afiliación a algún sistema de salud institucional son cualidades que poseen las personas de edad avanzada con pensión contributiva según datos de la CONSAR (2018), mismos aspectos que coinciden con el perfil individual del adulto mayor con menores posibilidades de trabajar en la informalidad de acuerdo con los resultados del modelo logístico binario empleado en este trabajo.

La regresión logística es especialmente útil para analizar fenómenos dicotómicos, como es clasificar en informal o formal a los adultos mayores, debido a que con su algoritmo fácilmente interpretable se puede identificar por qué se obtienen los resultados a los que se llegan de una forma eficiente. Además, no sólo indica cuan relevante puede llegar a ser una característica, sino también la dirección de su asociación, lo que puede ser especialmente útil en contextos en donde se analicen múltiples covariables.

Otra de las bondades de la regresión *logit* es que determina la importancia que tiene cada variable para el modelo final, es decir, se toman las decisiones con base en la existencia o ausencia de alguna de estas, lo que inclusive puede llegar a ser más

relevante que el modelo *per se*. También, al centrarse en predicciones del comportamiento del fenómeno estudiado con ayuda de probabilidades estimadas, es posible un pronóstico de clasificación.

Los resultados que arroja esta investigación revelan que las causas de que el grupo etario analizado esté activo dentro del mercado laboral informal se deben esencialmente a que, en general, sus condiciones son poco ventajosas, siendo la edad avanzada, pertenencia al estrato socioeconómico medio bajo, contar con seguro popular y por supuesto, un nivel educativo bajo el caldo de cultivo que obliga al adulto mayor a esta clase de trabajos.

Respecto a este último factor, prevalece la educación primaria como el común denominador como el máximo grado de estudios desde el 2010, cuando Madrigal-Martínez estudiaba la configuración de la seguridad económica en la vejez, a pesar de que la mayoría de la población mayor de esa época tenía este grado educativo inconcluso. El trabajo académico más reciente sobre el tema, el de Ramos en 2016, coincide en que la población de 60 años y más tiene seis años de instrucción formal.

La permanencia de este grado educativo es una marca generacional, en donde imperaba la necesidad de generar ingresos a edades más tempranas en vez de formar un capital humano, pero, sobre todo, porque el sistema educativo sólo favoreció a unos pocos. Contar con educación formal, aparte de posibilitar a las personas a mejores empleos y mayores ingresos, es una condición que las integra a la sociedad y favorece su independencia.

Conforme a las variables analizadas en el modelo logístico, contar con jubilación, AFORE o Fondo de Pensiones y atención médica influye de forma significativa para no

estar en el mercado informal. Estos resultados no hacen más que confirmar la importancia que siguen teniendo la seguridad social y los instrumentos financieros que ofrece; sin embargo, este sistema no está adecuado para resolver los problemas de exclusión social que se originan en el mercado laboral.

Es importante explorar el perfil ocupacional del adulto mayor, sobre todo en materia de la calidad del empleo y previsión social; si bien, ya están en el último tramo del ciclo vital, todavía son individuos con derechos y necesidades y, aunado al crecimiento en la esperanza de vida y nuevas tendencias en la oferta de empleos, parecen ser un sector demográfico especialmente olvidado. Aunque tiene bastante peso el Estado, también tiene que ver tiene el individuo, siendo quien forja el curso de su vida.

Asimismo, un área de oportunidad para complementar esta investigación es incluir otras técnicas evaluar la informalidad, porque a pesar de contemplar varios aspectos imprescindibles al estudiar este rubro, el lector puede notar la exclusión de la ganadería y agricultura de subsistencia, por lo que puede existir cierto subregistro, en especial con la parte de la población mayor que se considera PNEA, pero cuya principal fuente de ingresos es fruto del trabajo de campo.

En el mismo sentido, una propuesta de valor puede ser realizar una variante dependiendo de la zona geográfica, debido a que la evidencia empírica señala que existen desigualdades en lugares más rurales y el efecto de la migración es relevante sobre todo para las mujeres de edad; tal como demuestran los resultados de la investigación de Millán-León (2010), en donde favorece no tener una trayectoria laboral, no ser jefe del hogar, reportar un estado de salud deteriorado y un ingreso pensionario para ser inactivo económicamente en la tercera edad de los adultos mexiquenses.

Cabe destacar que algunos de los resultados de esta autora son conexos en esencia a los resultados encontrados en esta investigación: al ser la jefatura de hogar una figura responsable del bienestar de todos los integrantes, Millán-León (2010) encuentra que ser jefe de familia incide de forma positiva en la actividad laboral, aunque no va más allá sobre en qué tipo de actividad están ocupados, mientras que los resultados de esta investigación indican que aquellos con este estatus tienen predilección a trabajos en el mercado formal, al tener toda la carga financiera propia como de sus dependientes.

Por otro lado, la misma autora encuentra que el estado civil es relevante únicamente para la inactividad económica en las mujeres, lo que pone de relieve dos situaciones; primero, que el espectro de la viudez pesa más en ellas, ya que al morir el cónyuge pasan a vivir con sus hijos si son incapaces de sostenerse económicamente y segundo, que las redes de apoyo familiar tienen menos presencia entre los hombres. Estos aspectos no son condicionantes en lo que refiere este trabajo académico, pues importa más la fuente de ingreso que proporcionan la seguridad social y el Estado.

Es por estas razones que no se puede hacer una configuración homogénea en la seguridad económica de los adultos mayores ya que, a pesar de haber características clave que pueden llegar a definirla, hay otras que marcan una disparidad que prevalece en este sector poblacional principalmente porque esta seguridad económica tiene una fuerte carga de características demográficas, mismas que no son idénticas dentro de un mismo grupo, a pesar de ser generalizables.

Por otra parte, una alternativa adicional que se puede considerar es la propuesta de un modelo logístico para cada sexo, para develar diferencias sociales y económicas, producto de sus características generacionales y cómo encaran la informalidad laboral;

por ejemplo, Millán-León (2010) y Ramos (2016), en sus respectivos trabajos, coinciden que el factor determinante que aleja a las mujeres del mercado de trabajo es residir en zonas urbanas, más allá de su nivel de protección social o situación conyugal.

Finalmente, una cuestión importante es regular la empleabilidad y actividades económicas de la población mayor, donde se les provea de protección social y no tengan que recurrir a trabajos sin derechos laborales como única alternativa de sustento, sino por el contrario, que decidan permanecer en la PEA como lo que son, individuos productivos, en donde imperen como personas llenas de experiencias, en condiciones equitativas y con incentivos para la actualización de nuevas herramientas de trabajo.

BIBLIOGRAFÍA

- Alderete, A. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Revista Evaluar*, 6(1), 52-67. doi: <https://doi.org/10.35670/1667-4545.v6.n1.534>
- Allison, P. (2013). R Pubs by RStudio. Obtenido de Why I Don't Trust the Hosmer-Lemeshow Test for Logistic Regression: <https://rpubs.com/aniuxa/SFP11>
- AMAFORE. (2015). *Vejez y pensiones en México*. Ciudad de México: Publicación de AMAFORE.
- Aranibar, P. (2001). Acercamiento conceptual a la situación del adulto mayor en América Latina (21). Santiago: Publicación de las Naciones Unidas. Obtenido de <https://fiapam.org/wp-content/uploads/2012/10/lcl1656p.pdf>
- Arriaga, R. (2010). Efecto riqueza inmobiliario: análisis comparado para España, Estados Unidos y Reino Unido (Tesis doctoral). Universidad Autónoma de Madrid, Madrid.
- Aubone, A., & Wöhler, O. (2000). Aplicación del método de máxima verosimilitud a la estimación de parámetros y comparación de curvas de crecimiento de von Bertalanffy. Publicación de INIDEP, informe técnico, (37), 1-21. Obtenido de <https://biblat.unam.mx/es/revista/inidep-informe-tecnico/articulo/aplicacion-del-metodo-de-maxima-verosimilitud-a-la-estimacion-de-parametros-y-comparacion-de-curvas-de-crecimiento-de-von-bertalanffy>
- Bertranou, F., & Velasco, J. (2006). Tendencias en indicadores de empleo y protección social en América Latina. En F. Bertranou, *Envejecimiento, empleo y protección social en América Latina* (págs. 45-57). Santiago: Publicación de la OIT.

Bertranou, F. (2005). Restricciones, problemas y dilemas en la protección social en América Latina: enfrentando los desafíos del envejecimiento y la seguridad de los ingresos. *Bienestar y Política Social*, 1(1), 35-58. Obtenido de <https://biblat.unam.mx/es/revista/bienestar-y-politica-social/articulo/restricciones-problemas-y-dilemas-de-la-proteccion-social-en-america-latina-enfrentando-los-desafios-del-envejecimiento-y-la-seguridad-de-los-ingresos>

Boudiny, K. (2013). 'Envejecimiento activo': de la retórica vacía a una herramienta política eficaz. *Envejecimiento y sociedad*, 33(6), 1077–1098. doi:10.1017 / S0144686X1200030X

Cañadas-Reche, J. (2013). Regresión logística. Tratamiento computacional con R (Tesis de maestría). Universidad de Granada, Departamento de Estadística e Investigación Operativa, Granada, España.

CEPAL/CELADE. (2003). Las personas mayores en América Latina y El Caribe: diagnóstico sobre la situación y las políticas. Conferencia regional intergubernamental sobre envejecimiento: hacia una estrategia regional de implementación para América Latina y el Caribe del Plan de Acción Internacional de Madrid sobre el Envejecimiento (págs. 11-16). Santiago: Publicación de las Naciones Unidas. Obtenido de http://www.sld.cu/galerias/pdf/sitios/gericuba/adultos_mayores_en_america_latina_y_el_caribe.pdf

Choueiry, G. (2020). Understand Forward and Backward Stepwise Regression. Obtenido de Quantifying Health: <https://quantifyinghealth.com/stepwise-selection/>

Comisión Nacional del Sistema de Ahorro para el Retiro (CONSAR). (2018). Ingresos y gastos del adulto mayor en México: la importancia de las pensiones. Obtenido de

<https://www.gob.mx/consar/prensa/nuevo-documento-de-trabajo-ingresos-y-gastos-del-adulto-mayor-en-mexico-la-importancia-de-las-pensiones?idiom=es>

CONAPO. (2020). Proyecciones de la población de México y de las Entidades Federativas, 2016-2050. Ciudad de México: Publicación de CONAPO.

CONEVAL (Consejo Nacional de Evaluación de la Política de Desarrollo Social). (2016). Medición de la pobreza. Recuperado el 14 de enero de 2021, de Glosario: <http://www.coneval.org.mx/Medicion/Paginas/Glosario.aspx>
<http://www.coneval.org.mx/Medicion/Paginas/Glosario.aspx>

Cuentas, S. (2002). Propuestas e iniciativas para la seguridad económica de los adultos mayores en América Latina y El Caribe. Madrid: Presentación en Seminario del BID Sonia Cuentas. Centro de Orientación Socio-Legal del Adulto Mayor.

Cuevas-Rodríguez, E., De la Torre-Ruíz, H., & Regla-Dávila, S. (2016). Características y determinantes de la informalidad laboral en México. Estudios Regionales en Economía, Población y Desarrollo. Cuadernos de Trabajo de la UACJ, 35(6), 3-26. Obtenido de <https://erevistas.uacj.mx/ojs/index.php/estudiosregionales/article/view/1536>

De la Fuente-Fernández, S. (2011). Regresión Logística. Recuperado el 16 de noviembre de 2020, de Facultad de Ciencias Económicas y Empresariales, UAM: <https://docplayer.es/21085069-Santiago-de-la-fuente-fernandez-regresion-logistica.html>

Dobson, A. (2002). An introduction to generalized linear models. Boca Raton: Chapman & Hall/CRC.

Domínguez-Alonso, E., & Aldana-Padilla, D. (2001). Regresión logística. Un ejemplo de uso en Endocrinología. Revista Cubana Endocrinología, 12(1). Recuperado el 16 de noviembre

de 2020, de http://scielo.sld.cu/scielo.php?pid=S1561-29532001000100007&script=sci_arttext&lng=en

ENIGH. (2018). Nota Técnica. Ciudad de México: Publicación de INEGI. Obtenido de <https://www.inegi.org.mx/programas/enigh/nc/2018/>

Enríquez, A., & Galindo, M. (2015). Empleo, México ¿cómo vamos? Serie de Estudios Económicos. 1, 1-9. Obtenido de https://scholar.harvard.edu/files/vrios/files/201508_mexicoemployment.pdf?m=1453513189

Federación Internacional de Administradoras de Fondos de Pensiones (FIAP). (2016). Reformas a los Sistemas de Pensiones. Obtenido de Países que han incorporado el régimen de capitalización individual de los ahorros de manera obligatoria: <http://www.fiapinternacional.org/reformas-a-los-sistemas-de-pensiones/>

Ferrán, M. (2001). SPSS para Windows. Análisis estadístico. Madrid: McGraw-Hill.

Fiuza-Pérez, M., & Rodríguez-Pérez, J. (2000). La regresión logística: una herramienta versátil. Nefrología, 20(6), 477-565. Obtenido de <https://www.revistanefrologia.com/es-la-regresion-logistica-una-herramienta-versatil-articulo-X0211699500035664>

Garay, S., & Montes de Oca, V. (2006). La vejez en México: una mirada general sobre la situación socioeconómica y familiar de los hombres y mujeres adultos mayores. Perspectivas sociales, 13(1), 143-165. Obtenido de <http://eprints.uanl.mx/8789/>

Guzmán, J. (2002). Envejecimiento y desarrollo en América Latina y El Caribe (28). Santiago: Publicación de las Naciones Unidas, CELADE. Obtenido de <https://digitallibrary.un.org/record/472461?ln=es>

- Guzmán, J., Huenchuan, S., & Montes de Oca, V. (2003). Marco teórico conceptual sobre redes de apoyo social de las personas mayores. En CELADE, Redes de apoyo social de las personas mayores en América Latina y El Caribe (págs. 24-32). Santiago: Publicación de las Naciones Unidas.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). Análisis multivariante. Madrid: Prentice Hall Iberia.
- Hosmer, D., & Lemeshow, S. (2000). Applied logistic regression. USA: Wiley-Interscience.
- Huamaní-Miranda, M. A. (2014). Identificación de un modelo explicativo de retención de clientes con riesgo de fuga para una entidad bancaria aplicando regresión logística y árboles de clasificación CART (Trabajo monográfico). Universidad Nacional Agraria La Molina, Lima, Perú.
- Huenchuan, S. (2005a). Pobreza en la vejez y desigualdades de género en América Latina. *Perspectivas*, (15), 57-67. Obtenido de <http://ediciones.ucsh.cl/ojs/index.php/Perspectivas/issue/view/129>
- Huenchuan, S. (2005b). Pobreza y redes de apoyo social en la vejez. Acercamiento desde las diferencias de género. *Revista de Trabajo Social*, (13), 20-29. Obtenido de <https://biblat.unam.mx/es/revista/revista-trabajo-social-mexico-d-f/articulo/pobreza-y-redes-de-apoyo-en-la-vejez-acercamiento-desde-las-diferencias-de-genero>
- Huenchuan, S., & Guzmán, J. (2007). Seguridad económica y pobreza en la vejez: tensiones, expresiones y desafíos para políticas. *Notas de Población*, año XXXIII (83), 99-125. Obtenido de <https://fiapam.org/wp-content/uploads/2012/10/Pobreza.pdf>

Ibarra, M., & Michalus, J. (2010). Análisis del rendimiento académico mediante un modelo logit. Revista Ingeniería Industrial 9(2), 47-56. Obtenido de <http://revistas.ubiobio.cl/index.php/RI/article/view/56>

Instituto Nacional de Estadística y Geografía (INEGI). (2000). XII Censo General de Población y Vivienda 2000. Obtenido de <https://www.inegi.org.mx/temas/estructura/#Tabulados>

INEGI. (2005). II Censo de Población y Vivienda 2005. Obtenido de <https://www.inegi.org.mx/temas/estructura/#Tabulados>

INEGI. (2010). Censo de Población y Vivienda 2010. Obtenido de <https://www.inegi.org.mx/temas/estructura/#Tabulados>

INEGI. (2014). La informalidad laboral ENOE: marco conceptual y metodológico. c2014. México: Publicación de INEGI. Obtenido de http://internet.contenidos.inegi.org.mx/contenidos/productos//prod_serv/contenidos/espanol/bvinegi/productos/metodologias/ENOE/ENOE2014/informal_laboral/702825060459.pdf

INEGI. (2015). Encuesta Intercensal 2015. Obtenido de <https://www.inegi.org.mx/temas/estructura/#Tabulados>

INEGI. (2017). Encuesta Nacional de Empleo y Seguridad Social (ENESS) 2017. Obtenido de <https://www.inegi.org.mx/programas/eness/2017/#Tabulados>

INEGI. (2018). Encuestas en Hogares. Obtenido de Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). 2018 Nueva serie: <https://www.inegi.org.mx/programas/enigh/nc/2018/>

INEGI. (2018). Estadísticas a propósito del día internacional de las personas de edad (adultos mayores). Datos nacionales. Comunicado de prensa Núm. 432/18 (págs. 1-6). Ciudad de México: Publicación de INEGI.

INEGI. (2020). Censo de Población y Vivienda 2020. Obtenido de <https://www.inegi.org.mx/temas/estructura/#Tabulados>

INEGI. (2020). Encuesta Nacional de Ocupación y Empleo (ENOE) (nueva versión) 2020. Obtenido de <https://www.inegi.org.mx/temas/empleo/#Tabulados>

INEGI. (2020). Estadísticas a propósito del día del trabajo. Datos nacionales. Comunicado de Prensa Núm. 166/20 (págs. 1-2). Ciudad de México: Publicación de INEGI. Obtenido de <https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2020/trabajoNal.pdf>

Ivonet-Munder, M., & Estrada-Hernández, J. A. (2017). Análisis teórico por conglomerados de la seguridad económica de las pensiones mínimas de vejez. Anuario Facultad de Ciencias Económicas y Empresariales. 8, 3-16.

Lassen, A. (2015). The biopolitics of old age. How knowledge on ageing forms active ageing policies. Sociología Histórica, (5), 331-362. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5304412>

Lera-López, F. (1997). Insuficiencias de la teoría del ciclo vital en el comportamiento ahorrador. El caso de la comunidad Foral de Navarra. Documentos de Trabajo (Universidad Pública de Navarra. Departamento de Economía), 5, 1-26. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5519488>

López-Roldán, P., & Fachelli, S. (2015). Metodología de la investigación social cuantitativa. Barcelona: Creative Commons.

Madrigal-Martínez, M. (Marzo de 2010). Ingresos y bienes en la vejez, un acercamiento a la configuración de la seguridad económica de los adultos mayores mexiquenses. *Papeles de Población*, 16(63), 117-153. Obtenido de Ingresos y bienes en la vejez, un acercamiento a la configuración de la seguridad económica de los adultos mayores mexiquenses: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252010000100005

Maldonado, C., & Yáñez, M. (2014). Una aproximación al estudio del empleo en la tercera edad. *Cuadernos del CENDES*, 31(86), 95-110. Obtenido de <https://www.redalyc.org/pdf/403/40332804006.pdf>

Martínez-Restrepo, S., Enríquez, E., Pertuz, M. C., & Alzate-Meza, J. (2015). *El mercado laboral y las personas mayores*. Bogotá: Fundación Saldarriaga Concha.

Millán-León, B. (2010). Factores asociados a la participación laboral de los adultos mayores mexiquenses. *Papeles de la Población*, 16(64), 93-121. Obtenido de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252010000200005

Montes de Oca, V., & Hebrero, M. (2006). Turning points and advanced family cycles: aging effect in Mexican homes. *Papeles de la Población*, 12(50), 82-100. Obtenido de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5790122/>

Moreno, P., Catalán, H., & Murillo, R. (2000). La reforma de la seguridad social y la promesa de fortalecer el ahorro nacional. *Revista Trabajo*, 2(3), 139-146. Obtenido de <http://www2.izt.uam.mx/sotraem/Documentos/Trabaja2n32000.pdf>

Muñoz-Cobos, F., & Espinosa-Almen, J. (2008). Envejecimiento activo y desigualdades de género. *Atención Primaria*, 40(6), 305-309. Obtenido de

<https://www.semanticscholar.org/paper/Envejecimiento-activo-y-desigualdades-de-g%C3%A9nero-Cobos-Almendro/cebf2a5a7f5d595b3a071d371eb377d5660c1166?p2df>

Nava-Bolaños, I., & Ham-Chande, R. (2014). Determinantes de la participación laboral de la población de 60 años o más en México. *Papeles de la Población*, 20(81), 59-87. Obtenido de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252014000300004

Nava-Bolaños, I., Ham-Chande, R., & Ramírez-López, B. (2016). Seguridad económica y vejez en México. *Revista Latinoamericana de Población*, 10(19), 169-190. Obtenido de <https://www.redalyc.org/pdf/3238/323849595009.pdf>

OISS. (2008). Seguridad económica, salud, atención de la dependencia y participación de los adultos mayores en los países del cono sur. Proyecto sobre personas mayores, dependencia y servicios sociales en los países Iberoamericanos. Publicación de Secretaría General OISS. Obtenido de <https://oiss.org/proyecto-personas-mayores4750/>

OIT. (2002). Una sociedad inclusiva para una población que envejece: el desafío del empleo y la protección social. *Revista Española de Geriatria y Gerontología*, 37(52), 106-122. Obtenido de <https://www.elsevier.es/es-revista-revista-espanola-geriatria-gerontologia-124-articulo-una-sociedad-inclusiva-una-poblacion-13035695>

Omari, C., Nyambura, S., & Mwangi, J. (2018). Modeling the Frequency and Severity of Auto Insurance Claims Using Statistical Distributions. *Journal of Mathematical Finance*, (8), 137-160. doi: <https://doi.org/10.4236/jmf.2018.81012>

Pedrero-Nieto, M. (1999). Situación económica en la tercera edad. *Papeles de población*, 5(19), 77-101. Obtenido de <https://www.redalyc.org/pdf/112/11201906.pdf>

Pérez-Ortiz, L. (1996). La posición económica de los ancianos mayores. *Revista Española de Investigaciones Sociológicas*, 73(96), 149-176. Obtenido de <https://dialnet.unirioja.es/ejemplar/82651>

Pessanha, B., & Alves, V. (2019). Economía plateada: estrategias organizacionales para lidiar con el envejecimiento de los profesionales en el trabajo en el escenario brasileño. *Revista Contribuciones a las Ciencias Sociales*. Obtenido de <https://www.eumed.net/rev/cccss/2019/06/economia-plateada.html>

Ramos, E. (2016). Análisis de la participación laboral de los adultos mayores con base en un modelo logit. En CONAPO, *La situación demográfica de México 2016* (págs. 87-108). Ciudad de México: Publicación de CONAPO.

Rofman, R. (2005). *Social Security Coverage in Latin America*. Washington, D.C.: The World Bank.

Romero-Suárez, N. (2012). La revolución en la toma de decisiones estadísticas: el p-valor. *Revista Telos* 14, (3), 439-446. Obtenido de <https://www.redalyc.org/pdf/993/99324907004.pdf>

Rubalcava-Montero, D. (2019). *Historias laborales y seguridad económicas en la vejez: el caso México y Chile* (Tesis de maestría). Universidad de Chile. Facultad de Ciencias Físicas y Matemáticas, Santiago.

Rubio-Hurtado, M., & Vilá-Baños, R. (2016). El análisis de conglomerados bietápico o en dos fases con SPSS. *Revista de Innovación*, 10(1), 118-126. Obtenido de <http://diposit.ub.edu/dspace/bitstream/2445/110711/1/667750.pdf>

Ruiz-Durán, C. (2003). Modigliani: sesenta años de teoría económica. *Economía informal*, (321), 21-29. Obtenido de <http://www.economia.unam.mx/publicaciones/reseconinforma/pdfs/321/03%20Modigliani.pdf>

Sagaró del Campo, N., & Zamora-Matamoros, L. (2019). Análisis estadístico implicative versus Regresión logística binaria para el estudio de la causalidad en salud. *Revista Multimed*, 23(6), 1416-1440. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1028-48182019000601416

Sánchez-Vera, P. (2000). Sociología de la vejez versus economía de la vejez. *Papers. Revista de Sociología*, (61), 39-88. doi: <http://dx.doi.org/10.5565/rev/papers/v61n0.1053>

Serrano-Espinosa, J., & León-Andrade, M. (2015). Modelo de gestión de la cultura del ahorro para una vejez digna. *Jóvenes en la ciencia*, 1(2), 786-791. Obtenido de <http://www.jovenesenlaciencia.ugto.mx/index.php/jovenesenlaciencia/article/view/356/pdf1>

Silva-Aycaguer, L. (1995). *Excursión a la regresión logística en ciencias de la salud*. Madrid: Díaz de Santos.

Smith, T., & McKenna, C. (2013). A Comparison of Logistic Regression Pseudo R² Indices. *Multiple Linear Regression Viewpoints*, 39 (2), 17-26. Obtenido de https://www.glmj.org/archives/articles/Smith_v39n2.pdf

Solano-Dávila, O., Ramírez, A., Bartolo, F., Giraldo, O., & Salinas, A. (2007). Análisis de Diagnóstico en el Modelo de Regresión Logística: Una aplicación. *PESQUIMAT Revista de la Fac. CC. MM. de la Universidad Nacional Mayor de San Marcos*, 10(1), 55-70.

Obtenido

de

<https://revistasinvestigacion.unmsm.edu.pe/index.php/matema/article/view/9431/8253>

Stoltzfus, J. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, 18, 1099-1104. doi: <https://doi.org/10.1111/j.1553-2712.2011.01185.x>

van Gameren, E. (2008). Labor Force Participation of Mexican Elderly: the Importance of Health. *Estudios Económicos (México)*, 23(1), 89-127. Obtenido de <https://pubmed.ncbi.nlm.nih.gov/29375171/>

Vega, E., & Ayala, A. (2015). Envejecimiento activo y saludable. Ciudad de México: Ponencia en Encuentro México-Japón sobre Envejecimiento Activo y Saludable Organizado por la SIDSS, el Instituto Nacional de Geriátría, la Asociación de Ex Becarios Nikkei de México y la Japan International Agency.

Vilar del Hoyo, L., Martín Isabel, M., & Martínez-Vega, J. (2008). Empleo de regresión logística para la obtención de modelos de riesgo humano de incendios forestales. *Boletín de la Asociación de Geógrafos Españoles*, 47, 5-29. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=3093337>

Weller, J. (2018). La inserción laboral de las personas mayores en América Latina. *Boletín de Envejecimiento y Derechos de las Personas Mayores*. Núm. 16, CEPAL. Obtenido de https://crm.cepal.org/civicrm/mailling/view?id=378#articulo_central

Zuluaga-Callejas, M., & Robledo-Marín, C. (2016). Alternativas para garantizar la seguridad económica en la vejez: desafíos para Colombia. *CES Derecho*, 7(1), 39-50. Obtenido de <http://www.scielo.org.co/pdf/cesd/v7n1/v7n1a05.pdf>