



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

RELACIÓN DE FACTORES ECONÓMICOS DE
LOS HOGARES MEXICANOS Y LAS
ENFERMEDADES CRÓNICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaria

PRESENTA:

Celene Noemi León Alonso

DIRECTOR DE TESIS:

M. en C. José Salvador Zamora Muñoz

Ciudad Universitaria, CD. MX 2019





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO

1.Datos alumno

León Alonso Celene Noemi

59340262

Universidad Nacional Autonoma de México

Facultad de Ciencias

Actuaría

307274388

2.Datos Asesor

M. en C. Zamora Muñoz José Salvador

3.Datos Sinodal

Dra. Lizbeth Naranjo Albarrán

4.Datos Sinodal

Dra.Sofía Villers Gómez

5.Datos Sinodal

M en C. Jaime Vázquez Alamilla

6.Datos Sinodal

M en E. Leslie Alejandra Jiénez Rosas

Agradecimientos

Magos:

Te agradezco el apoyo ante todos los momentos difíciles que tuvimos que pasar juntas, y por todas las enseñanzas y consejos que me das y que han formado la mujer que soy hoy, quiero que sepas que eres el motor de mi vida y el regalo más grande que Dios me dio. Juntas seguiremos cumpliendo metas y disfrutando de la vida. Te amo!!!

Rosa:

Gracias por el amor incondicional que siempre me das, a pesar de que no soy una hija ejemplar, gracias por ser fuerte cuando más lo necesitamos, gracias por dar todo lo que teníamos para que yo tuviera tratamientos, porque gracias a eso logre terminar y mi suerte fue diferente a la de todas esas muchachas que veíamos, espero te sientas orgullosa de mí, quiero que sepas que siempre estás en mi mente y que mi deseo de ser mejor es para que tú siempre estés bien. Te amo mamá!!

Cesar:

Gracias por crecer conmigo, gracias por escucharme tantas veces de los sueños que tenía, gracias por que juntos pasamos los terrores de la carrera y las adversidades que se presentaron en nuestra contra para terminarla, y por que juntos tuvimos los mismo temores al salir de ella, pero al final todo salió mejor de lo que pensábamos, descubrir el mundo a tu lado hizo que todo fuera más sencillo. Gracias por seguir apoyándome hasta el último minuto. Lo logramos!!!!

Eduardo:

Nunca decías muchas cosas, solías pensar mucho y estoy segura que nosotras vivíamos en tu mente, las decisiones que tomabas eran en favor de nuestro bienestar, no siempre te salieron bien. Pero gracias a ello Magos y yo somos mujeres fuertes, consientes, inteligentes y de buenos valores. Es muy notorio que nos parecemos a tí, nuestra forma de pensar es por tí, si hoy somos profesionistas es por que estabas seguro de que lo seríamos. Te extraño mucho y espero que estés orgulloso de las decisiones que he tomado y aunque llegó un poco tarde lo logré. Te amo papá.

Abuelo:

Aunque no pude lograrlo cuando aun podías verlo, eres el primero al que se lo dedico, ya no debes preocuparte, ya lo logré y las tres estaremos más que bien, gracias por haber rezado todas las mañanas por nosotras.

Profesor Salvador:

Gracias por su apoyo y por compartirme sus conocimientos para que hoy pueda concluir una de las etapas más importantes de mi vida.

UNAM:

No podía faltar la Universidad y todas las personas que hacen posible su funcionamiento, ya que por tan poco, me dio educación, trabajo, amigos y siempre me ha respaldado su nombre a donde quiera que presento, gracias UNAM y Facultad de Ciencias.

Contenido

1. Introducción	7
2. Objetivos	9
3. Revisión de Literatura	10
3.1. Conceptos y definiciones	10
3.2. Causas de las enfermedades crónicas	14
3.2.1. Sobrepeso y obesidad	16
3.2.2. Diabetes mellitus	18
3.2.3. Hipertensión	20
3.3. Enfermedades crónicas en el mundo y el caso particular de México	21
3.4. Preliminares de la relación pobreza - salud	24
4. Análisis Exploratorio de los Datos	30
4.1. Estadística descriptiva de la encuesta	34
4.1.1. Peso	34
4.1.2. Diabetes	38
4.1.3. Hipertensión	41
5. Métodos Estadísticos	44
5.1. Preliminares regresión logística múltiple	44
5.2. Interpretación del modelo	47
5.3. Estimación de los parámetros	49
5.4. Pruebas de bondad de ajuste del modelo	54
5.4.1. Significancia de los parámetros	54
5.4.2. Diagnóstico del modelo logístico	56
5.4.3. Curva de ROC	58
5.4.4. Área bajo la curva de ROC	60
6. Construcción del Modelo	61
6.1. Construcción de la base de datos	62
6.2. Especificación de las variables	63
6.3. Definición del modelo	67
6.4. Regresión logística obesidad	69
6.4.1. Estimación de los parámetros	69
6.4.2. Interpretación del modelo	71

6.4.3. Bondad de ajuste del modelo	73
6.5. Regresión logística diabetes	81
6.5.1. Estimación de los parámetros	81
6.6. Regresión logística hipertensión	83
6.6.1. Estimación de los parámetros	83
7. Conclusiones	84
8. Bibliografía	86

1. Introducción

Segun el censo realizado por INEGI en 2015, México cuenta con 119,530,753 habitantes, dicha población representa el capital humano con que el país cuenta y trabaja cotidianamente. Como se ha comprobado a lo largo de la historia, la educación, la cultura, la salud y la nutrición de los habitantes es parte integral y fundamental de dicho capital humano para el fomento del desarrollo social y económico del país.

Desafíos como los cambios climáticos, los rápidos cambios tecnológicos, el crecimiento económico desigual, las transformaciones sociales y entre otros factores modelan los sistemas alimentarios y las dietas, planteándonos interrogantes acerca de la alimentación de la creciente población mundial.

Desde 1988 se han realizado diversas encuestas con el objetivo de documentar la situación nutricional en México. Las principales preocupaciones son la deficiencia de hierro o anemia y la desnutrición, sin embargo, como ya es bien conocido, en los últimos años el aumento de personas con sobrepeso y obesidad han alarmado, no solamente a nuestro país, sino al mundo. De acuerdo con el informe Global Nutrition Report 2016 indicó que los índices de obesidad y sobrepeso están en aumento en todas las regiones del mundo y en casi todos los países, asimismo, dicho reporte resalta cuán insuficientes son los avances realizados en la lucha contra todas las formas de la malnutrición.

La mayoría de los países están lejos de aportar los esfuerzos necesarios para reducir la anemia en las mujeres y para prevenir el aumento de los casos de diabetes. Al menos 57 países registran importantes niveles de desnutrición, sobrepeso y obesidad en los adultos, lo que representa una enorme presión sobre muchos sistemas sanitarios, de por sí muy frágiles.

México forma parte de estos 57 países, donde la malnutrición de sus habitantes es un problema serio para su capital humano, así como para las diferentes instituciones de salud gubernamentales, pues ya es bastante conocido que las enfermedades crónicas representan altos costos para estas instituciones gubernamentales de salud.

Las encuestas nutricionales nos ayudan a tener un panorama general de cómo se encuentra el estado nutricional en nuestro país, gracias a ellas se han realizado campañas y programas para la divulgación y prevención de las diferentes problemáticas que vive la alimentación de los mexicanos, sin embargo, a pesar de los esfuerzos, nos hemos dado cuenta que con el incremento en los índices de sobrepeso y obesidad, una campaña de cultura y concientización no es suficiente para frenar o erradicar tales problemas. Es por ello que debemos trabajar en encontrar las razones de raíz que nos llevan a los mexicanos a tener dietas deficientes así como una cultura de salud inadecuada.

Las razones por las cuales los problemas de malnutrición existen en el mundo son probablemente conocidas y discutidas, entre ellas están: la pobreza, la falta de información, los avances tecnológicos, así como la globalización. Sin embargo, debemos enfocarnos en medir los impactos que representan cada uno de estos factores para poder tomar decisiones y crear programas que ayuden a erradicar y prevenir las problemáticas derivadas de una mala situación nutricional.

El nivel socioeconómico juega un papel importante en el sobrepeso y obesidad, ya que las personas con mayor ingreso deberían padecer en mayor proporción estos problemas, al contar con mayores recursos y por ende disponibilidad de alimentos.

El objetivo del presente estudio es comprobar si factores económicos tales como el decil socioeconómico, el área donde viven, tamaños de vivienda, entre otros, son causa de enfermedades crónicas como obesidad y sobrepeso, hipertensión y diabetes.

La fuente de información utilizada para el presente estudio es la Encuesta Nacional de Salud y Nutrición 2012 (ENSANUT2012) ya que ésta incluye características de la vivienda y bienes en el hogar, seguridad social, aspectos sobre enfermedades crónicas, actividad física, diversidad de la dieta. Adicionalmente, se tomaron mediciones de antropometría de peso y talla, y muestras de sangre. Con ayuda de esta información se ajustará un modelo de regresión logística que, dependiendo de los percentiles de ingresos, proporcione evidencia suficiente para demostrar si el nivel socioeconómico y demás factores económicos son causa de presencia de enfermedades crónicas en los hogares mexicanos.

2. Objetivos

El objetivo primordial de este análisis es conocer si factores socioeconómicos tales como el decil socioeconómico, la marginalidad, área sociodemográfica y número de habitaciones para dormir son causa de la presencia de enfermedades crónicas como la obesidad y sobrepeso, hipertensión y diabetes en los hogares mexicanos.

Premisas:

- Los factores económicos son causa de enfermedades crónicas en los hogares mexicanos.
- A mayor nivel socioeconómico mayor probabilidad de presencia de enfermedades crónicas.

3. Revisión de Literatura

3.1. Conceptos y definiciones

Durante el presente trabajo se hablará y hará referencia con frecuencia a distintos conceptos, por lo que es necesario incorporar definiciones y contextos de los conceptos utilizados. Las definiciones están basadas en la Norma Oficial Mexicana PROY-NOM-043-SSA2-1999 ⁽¹¹⁾ y Organización Mundial de la Salud (OMS).

- **Adolescentes:** Son las personas del hogar de 10 a 19 años cumplidos.
- **Adultos:** Son las personas del hogar mayores de 20 años.
- **Alimentación correcta:** Dieta que, de acuerdo con los conocimientos reconocidos en la materia, cumple con las necesidades específicas de las diferentes etapas de la vida, promueve en los niños y las niñas el crecimiento y el desarrollo adecuados y en los adultos permite conservar o alcanzar el peso esperado para la talla y previene el desarrollo de enfermedades.
- **Alimentos densos en energía:** Contienen en promedio de 300 a 900 calorías por cada cien gramos o mililitros; el grupo se encuentra conformado por alimentos como azúcar, café, arroz, aceites y grasas vegetales, carnitas, chorizo, frijol, maíz, mantequilla, pan dulce, papas fritas, pasta para sopa, quesos amarillo, manchego y asadero, galletas y tocino, entre otros.
- **Alimentos procesados:** Son aquellos alimentos a los cuales les fueron agregados de forma industrializada distintos aditivos tales como conservadores, sal, saborizantes, azúcares, colorantes, etc.
- **Circunferencia de cintura (CC):** Es un indicador de adiposidad central muy útil para evaluar el riesgo cardiovascular. Al igual que el Índice de Masa Corporal (IMC) es de fácil aplicación, bajo costo y no invasivo. Aunque no ajusta para la estatura debido a la poca variabilidad en gran parte de la población, no afecta de forma importante su valor predictivo.

- **Diabetes mellitus:** Se caracteriza por la presencia en la sangre de niveles elevados de glucosa. En condiciones normales, la glucosa procedente de la metabolización de los alimentos se concentra en el torrente sanguíneo para llegar hasta las células de los diferentes tejidos entre ellos los músculos y transformarse en energía. Esta actividad se lleva a cabo gracias a la acción de la insulina que se produce en el páncreas. Si las células absorben menos glucosa y ésta se acumula en la sangre, sin que la insulina producida por el organismo sea suficiente para lograr que éste pueda absorber los azúcares y se mantengan en el torrente sanguíneo. De este modo se produce un círculo vicioso que es el origen de la diabetes mellitus. La obesidad es un factor predisponente.
- **Dieta:** Conjunto de alimentos y platillos que se consumen cada día, y constituye la unidad de la alimentación.
- **Dieta correcta:** La dieta que cumple con las siguientes características:
 - **Completa:** Que contenga todos los nutrimentos. Se recomienda incluir en cada comida alimentos de los 3 grupos (1.-Verduras y Frutas, 2.-Cereales y Tubérculos, 3.- Leguminosas y Alimentos de origen animal).
 - **Inocua:** Que su consumo habitual no implique riesgos para la salud porque está exenta de microorganismos patógenos, toxinas y contaminantes y se consuma con moderación.
 - **Suficiente:** Que cubra las necesidades de todos los nutrimentos, de tal manera que el sujeto adulto tenga una buena nutrición y un peso saludable y en el caso de los niños, que crezcan y se desarrollen de manera correcta.
 - **Variada:** Que incluya diferentes alimentos de cada grupo en las comidas.
- **Enfermedades crónicas:** Son enfermedades de larga duración y progresión generalmente lenta, son la principal causa de muerte e incapacidad en el mundo y las principales enfermedades crónicas son obesidad, diabetes e hipertensión.

- **Hábitos alimentarios:** Al conjunto de conductas adquiridas por un individuo, por la repetición de actos en cuanto a la selección, la preparación y el consumo de alimentos. Los hábitos alimentarios se relacionan principalmente con las características sociales, económicas y culturales de una población o región determinada. Los hábitos generalizados de una comunidad suelen llamarse costumbres.
- **Hipertensión:** La tensión arterial se mide en milímetros de mercurio (mm Hg) y se registra en forma de dos números separados por una barra. El primero corresponde a la tensión arterial sistólica, la más alta, que se produce cuando el corazón se contrae. El segundo corresponde a la tensión arterial diastólica, la más baja, que se produce cuando el músculo cardíaco se relaja entre un latido y otro. La tensión arterial normal en un adulto se define como una tensión sistólica de 120 mm Hg y una tensión diastólica de 80 mm Hg. La hipertensión se diagnostica como una tensión sistólica igual o superior a 140 mm Hg y una tensión diastólica igual o superior a 90 mm Hg.
- **Hogar:** Es el conjunto de personas relacionadas por algún parentesco o no, que habitualmente duermen en una vivienda bajo el mismo techo, beneficiándose de un ingreso común aportado por uno o más de los miembros del hogar.
- **Índice de masa corporal (IMC):** Al criterio diagnóstico que se obtiene dividiendo el peso entre la estatura elevada al cuadrado (se establece al dividir el peso corporal expresado en kilogramos, entre la estatura expresada en metros elevada al cuadrado). Es el indicador más útil para evaluar la composición corporal a nivel poblacional porque es de bajo costo, de fácil aplicación y la forma de calcularlo no varía en función del sexo ni de la edad en la población adulta.
- **Niños preescolares:** Son las personas del hogar con edades de 0 a 4 años cumplidos.
- **Niños escolares:** Son las personas del hogar de 5 a 9 años cumplidos.
- **Nutriente:** Toda sustancia presente en los alimentos que juega un papel metabólico en el organismo.

- Obesidad:** Se define como la acumulación en el cuerpo de masa grasa en exceso (cantidades superiores a las encontradas en individuos que mantienen el balance energético, es decir, que su consumo está equilibrado con el gasto de energía) el siguiente recuadro muestra los criterios para determinar si un individuo sufre de obesidad según el IMC y la Circunferencia de cintura.

CUADRO 1. Clasificación de obesidad por IMC y circunferencia de cintura.

Por IMC^a de acuerdo con la Organización Mundial de la Salud (OMS) y la norma oficial mexicana (008-SSA3-2010)

Fuente	Bajo peso	Normal	Sobrepeso	Obesidad		
OMS	<18.5	18.5 - 24.9	25.0-29.9	Grado I	Grado II	Grado III
				30.0-34.9	35.0-39.9	>40.0
NOM	-	-	25.0-29.9	≥ 30 o ≥ 25 en adultos de baja talla		
			o ≥ 23 y <25 en adultos de baja talla ^a			

Por el riesgo de desarrollar complicaciones metabólicas de acuerdo con la circunferencia de cintura^b

Riesgo de complicaciones metabólicas	Incrementado	Sustancialmente incrementado	
Hombres	≥ 94 cm	≥ 102 cm	
Mujeres	≥ 80 cm	≥ 88 cm	

Obesidad abdominal de acuerdo con los criterios de la International Diabetes Federation^c

Hombres	≥ 90 cm
Mujeres	≥ 80 cm

IMC = Peso actual (kg)/ estatura (m)².

IMC saludable^e < 24.

Peso saludable o IMC saludable = (24) * talla en (m)².

Rango de peso saludable: IMC saludable (escoger un IMC menor a 25) ejemplo: (24.9) * 1.60 m².

Peso saludable mínimo = 18.5 * 2.56 = 47.3.

Peso saludable máximo = 24.9 * 2.56 = 63.7.

^a El IMC saludable se puede definir en el ámbito clínico entre 20 y 25 unidades en hombres y entre 19 y 24 en mujeres.

El personal de salud debe determinar cuál es el IMC saludable dentro de este rango tomando como base la constitución, edad y expectativas de cada paciente. Si se toma el límite superior del rango, el peso despejado se puede denominar peso máximo normal.

^b Adaptado de: WHO (2000), Obesity: Preventing and Managing the Global Epidemic, Report of a WHO Consultation on Obesity.

^c Baja talla = estatura menor a 1.50 metros en la mujer adulta y menor de 1.60 metros para el hombre adulto.

3.2. Causas de las enfermedades crónicas

Las enfermedades crónicas tales como las que se pretende estudiar (sobrepeso y obesidad, diabetes e hipertensión) son enfermedades de curso crónico que tiene como origen una cadena causal compleja multifactorial, donde interactúan factores genéticos, sociales y ambientales, incluyendo estilos de vida, así como determinantes sociales y económicos.

Si bien el objetivo primordial es verificar si las enfermedades crónicas son una consecuencia directa del presupuesto monetario de los mexicanos, es necesario comenzar por distinguir e identificar las causas y contextos de dichas enfermedades. Rivera y Gonzales⁽²⁾ hacen la siguiente clasificación:

Causas básicas	Causas Subyacentes	Causas Inmediatas
<p>Las causas básicas tienen que ver con la cultura, las condiciones sociales y el sistema económico</p> <p>Ejemplo</p> <ul style="list-style-type: none"> • Industrialización de alimentos • Nivel socioeconómico de los individuos 	<p>Las causas subyacentes son resultado de las causas básicas, tales como el entorno familiar, escolar, la disponibilidad y accesibilidad de los alimentos etc</p> <p>Ejemplo</p> <ul style="list-style-type: none"> • Cambio en la cultura alimentaria • Disponibilidad y accesibilidad de alimentos densos en energía • Disminución de los tiempos dedicados a realizar comidas 	<p>Las causas inmediatas se relacionan con las conductas individuales tales como la proporción de alimentos, los tipos de alimentos consumidos, las causas genéticas y fisiológicas.</p> <p>Ejemplo</p> <ul style="list-style-type: none"> • Dietas abundantes, poca actividad física • Mayor consumo de alimentos ricos en azúcares • En menor proporción genes que favorecen a los procesos digestivos y absorción de nutrientes

Las causas básicas están relacionadas con la industrialización, urbanización, globalización, los cambios tecnológicos en la producción y procesamiento de alimentos, también están las políticas de estado y las normas sociales, que favorecen u obstruyen la alimentación correcta de las poblaciones.

Sobre las causas subyacentes, están relacionadas con el cambio en la cultura alimentaria y hábitos alimenticios que incluye el tipo de bebidas, alimentos y platillos que se consumen, aumento en el tamaño de las porciones.

Otras de las causas subyacentes que están generando estos cambios son la elevada disponibilidad, accesibilidad y alto consumo de alimentos densos en energía y pobres en nutrientes y de bebidas con alto aporte calórico, el consumo cada vez mayor de alimentos procesados con menor contenido de granos enteros y fibra y con mayor contenido de harinas refinadas y azúcares.

Podemos notar que las distintas causas de las enfermedades crónicas tienen como base fundamental la alimentación. Las causas básicas por ejemplo se hacen presentes en la restricción de la compra de alimentos de calidad por cuestiones de una mala situación económica. Las causas subyacentes tienen que ver con la preferencia de alimentos altos en energía por una cuestión cultural y finalmente las causas inmediatas que tienen que ver con la frecuencia con la que se consumen los alimentos y el procesamiento físico de los mismos.

Como ya se ha mencionado, es una cadena de múltiples factores que finalizan en las causas denominadas inmediatas, entonces es de vital importancia conocer los factores de riesgo que tienen que ver con los tipos de alimentos que consumimos o los comportamientos que giran en torno a éstos para el desarrollo de enfermedades crónicas. Por lo que se describirán de manera particular y breve los principales factores de riesgo para cada una de las enfermedades crónicas en cuestión.

Para poder dar evidencia científica de los principales factores de riesgo de la obesidad, diabetes e hipertensión, instituciones como el World Cancer Research Found (WCRF) en la publicación *Food, Nutrition, Physical and the prevention of the Cáncer* y la Organización Mundial de la Salud (OMS) en el reporte técnico numero 916 - *Dieta Nutrición y la prevención de enfermedades crónicas*, hicieron revisiones sistemáticas de la literatura y meta-análisis para realizar asociaciones de acuerdo con la clasificación de las causas; mientras las causas inmediatas se basan en evidencia epidemiológica directa como ensayos clínicos aleatorizados o estudios de cohorte, las causas básicas se sustentan en evidencia indirecta, generalmente de tipo observacional en relación con los diversos factores que se asocian a estas causas básicas, y su estudio se apoya también en herramientas no epidemiológicas empleadas por disciplinas como la antropología, la psicología, la economía y la sociología.⁽³⁾

3.2.1. Sobrepeso y obesidad

La obesidad es el resultado de causas inmediatas tales como consumo excesivo de energía resultante de la cantidad, frecuencia y calidad de los alimentos consumidos aunados a la poca actividad física, sin embargo, estas causas inmediatas, características del ámbito individual, son a su vez el resultado de factores subyacentes y causas básicas.

La siguiente tabla muestra los factores de riesgo para el desarrollo de sobrepeso u obesidad de acuerdo a la clasificación de los estudios realizados por WCRF y OMS en los años 2007 y 2003, respectivamente.

EVIDENCIA	OMS 2003		WCRF 2007	
	Factores que disminuyen el riesgo	Factores que aumentan el riesgo	Factores protectores	Factores etiológicos
Convincente	Actividad física regular Consumo de fibra	Sedentarismo e inactividad física Consumo de alimentos de alta densidad energética	Actividad física	Inactividad física
Probable	Lactancia materna Ambientes que promuevan la elección de alimentos saludables	Consumo de refrescos y bebidas azucaradas Mercadotecnia de alimentos con alta densidad energética y de comida rápida Factores socioeconómicos adversos	Consumo de alimentos de baja densidad energética Lactancia materna	Consumo de alimentos de alta densidad energética Consumo de bebidas azucaradas Consumo de comida rápida Ver televisión
Posible	Consumo de alimentos de bajo índice glucémico	Aumento en el tamaño de la porción Consumo de alimentos fuera de casa		
Limitada (no concluyente)			Cereales refinados, vegetales con alto contenido de almidón, frutas, carne, leche y productos lácteos, pescado, jugos de frutas, café, bebidas alcohólicas, edulcorantes	

CUADRO 2. Clasificación de la evidencia existente sobre los factores de la dieta y de actividad física y su relación con la obesidad según la revisión de la Organización Mundial de la Salud (2003) y la revisión de la Fundación Mundial de la Investigación del Cáncer (2007)

Así las causas inmediatas principales de la obesidad y el sobrepeso son el consumo de alimentos con alta densidad energética, refrescos y bebidas azucaradas, los grandes tamaños de las porciones y la elevada frecuencia en la ingestión de alimentos que se consumen entre comidas y de los que acompañan platillos durante las comidas formales, en conjunto con la inactividad física y la prevalencia de estilos de vida sedentarios.

3.2.2. Diabetes mellitus

Al igual que la obesidad, la Diabetes Mellitus tiene múltiples causas básicas, subyacentes e inmediatas relacionadas con la alimentación, factores económicos, culturales, sociales y genéticos, se resumirán los principales factores de riesgos relacionados al desarrollo de diabetes tipo 2:

- **Presencia de obesidad:** La obesidad es un factor de riesgo y un potente predictor del desarrollo de diabetes tipo 2. Estudios han demostrado que la circunferencia de la cintura o el índice cintura-cadera pueden ser mejores indicadores de riesgo a desarrollar diabetes que el IMC. El menor riesgo de desarrollo de diabetes tipo dos se presentó en personas con un IMC menores a 21.⁽³⁾
- **Inactividad física:** Existen varios estudios que demuestran la asociación entre la Inactividad física y el desarrollo de diabetes tipo 2, ya que la realización de ejercicio intensivo (Cinco veces por semana, mínimo 20 minutos por sesión) mejora la sensibilidad a la insulina en sujetos con resistencia a la insulina normal.⁽³⁾
- **Consumo de bebidas azucaradas:** En un meta-análisis reciente se encontró la asociación entre un mayor consumo de bebidas azucaradas y refrescos y la incidencia de diabetes tipo 2. Aquellos individuos en el quintil más alto de consumo de estas bebidas (en general 1 - 2 raciones en el día) comparados con aquellos que no consumían o consumían 1 ración en el día, tuvieron 26 % más riesgo de desarrollar diabetes tipo 2.⁽³⁾

- **Dietas bajas en fibra y altas en carbohidratos:** En una revisión de la literatura, no se encontró una asociación entre el consumo total de carbohidratos y el riesgo de desarrollar diabetes tipo 2, concluyendo que el tipo y la fuente de los carbohidratos parecen ser factores más importantes que la cantidad en el desarrollo de diabetes tipo 2.

Algunos estudios grandes de cohorte han demostrado que un consumo de fibra relativamente bajo aumenta el riesgo de desarrollar diabetes tipo 2 significativamente, siendo la asociación más fuerte con fibra insoluble, como la que se encuentra en los cereales.⁽³⁾

En síntesis, los factores dietéticos y de actividad física que aumentan el riesgo de desarrollar diabetes tipo 2 son la existencia de obesidad y sobrepeso, consumo de refrescos y bebidas azucaradas, el sedentarismo y dietas bajas en fibra y altas en carbohidratos.

3.2.3. Hipertensión

La prevalencia creciente de la hipertensión se atribuye a factores de riesgo relacionados con el comportamiento, como la dieta malsana, el uso nocivo del alcohol, la inactividad física, el sobrepeso o la exposición prolongada al estrés⁽⁴⁾ algunos otros factores son:

- **Presencia de obesidad:** La obesidad es un factor de riesgo independiente para el desarrollo de hipertensión. Los datos indican que el peso corporal a los 18 años de edad se relaciona con riesgo a desarrollarla. Las recomendaciones son mantener un peso saludable a lo largo la vida (tener un IMC entre 18.5 y 24.9).⁽³⁾
- **Reducción de alimentos con sal:** Se ha documentado que el consumo de sodio y de potasio, por su papel en la regulación de la presión arterial, tiene un rol en la salud. En una revisión de intervenciones se concluyó que una reducción de 70-80 mmol en el consumo de sodio en la dieta se asoció con disminuciones en la presión arterial, en sujetos con y sin hipertensión. Estos estudios no han evaluado los efectos de suplementos de potasio a largo plazo, por lo que es recomendable asegurar un consumo adecuado de potasio que provenga de frutas y verduras en la dieta.⁽³⁾

Entonces factores como el sobrepeso y obesidad, el consumo elevado sodio, de alcohol y el estrés son causas inmediatas de la hipertensión. De manera general y concluyente, los factores que aumentan el riesgo de padecer obesidad y/o una o más de las enfermedades crónicas son dietas con consumo excesivo de refrescos, bebidas azucaradas y bebidas alcohólicas, de alimentos con alta densidad energética, o altos en sal, la inactividad física y el sedentarismo.

3.3. Enfermedades crónicas en el mundo y el caso particular de México

Las enfermedades crónicas como la obesidad, la diabetes mellitus y la hipertensión son causas cada vez más importantes de discapacidad y muerte prematura en los países tanto desarrollados como subdesarrollados, suponen una carga adicional para los presupuestos sanitarios de cada país.

La carga de enfermedades crónicas está aumentando rápidamente en todo el mundo, se ha calculado que, en 2001, las enfermedades crónicas causaron aproximadamente un 60 % del total de 56.5 millones de defunciones notificadas en el mundo y se ha previsto que para 2020 las enfermedades crónicas representarán casi las tres cuartas partes del total de defunciones; por ejemplo, el número de personas con diabetes en el mundo se multiplicará por más de 2.5 y pasará de 84 millones en 1995 a 228 millones en 2025; en cuanto al sobrepeso y la obesidad, no sólo la prevalencia actual ha alcanzado niveles sin precedentes, sino que la tasa de aumento anual es sustancial en la mayoría de las regiones en desarrollo, las implicaciones de este fenómeno para la salud pública son tremendas y ya están resultando evidentes.⁽⁵⁾

En lo que respecta a la hipertensión en 2008 en el mundo se habían diagnosticado aproximadamente el 40 % de los adultos mayores de 25 años y el número de personas afectadas aumentó de 600 millones en 1980 a 1000 millones en 2008.⁽⁴⁾

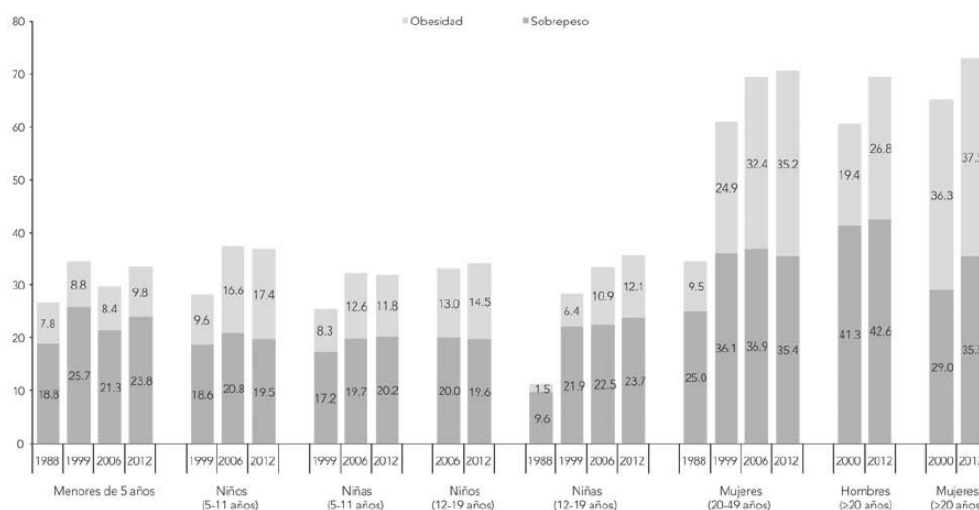
En países subdesarrollados como México, la transición epidemiológica de las enfermedades crónicas genera un problema de salud pública, pues el incremento en la esperanza de vida trae consigo un aumento en el riesgo de padecer enfermedades crónicas lo cual es una consecuencia inevitable.⁽⁶⁾

La obesidad es el principal factor de riesgo para el desarrollo de diabetes tipo 2, enfermedades cardiovasculares, hipertensión arterial y ciertos tipos de cáncer⁽³⁾. Con cada año aumenta la preocupación de que las más recientes generaciones tengan una menor calidad de vida por falta de actividad física y una ingesta dietética no saludable.

En nuestro país la prevalencia de sobrepeso y obesidad en niños y adultos se ha incrementado de manera alarmante en las últimas dos décadas, ha hecho que México se posicione como uno de los países con mayor aceleración en el crecimiento de la prevalencia de sobrepeso y obesidad y por ende de enfermedades crónicas

La siguiente gráfica presenta las prevalencias de sobrepeso y obesidad para menores de cinco años, niños en edad escolar (5 - 11 años), adolescentes (12 - 19 años), mujeres de 20 a 49 años y adultos (mayores a 20 años). A pesar de que la velocidad de incremento en la prevalencia de sobrepeso y obesidad en adultos ha disminuido en los últimos seis años, la tendencia mantenida entre 1994 y 2000 ha hecho que el aumento de 50 % en sobrepeso (24 % a 35 %) y 150 % en obesidad (de 9 % a 24 %) posiciona a México como uno de los países con mayor aceleración en el crecimiento de la prevalencia de sobrepeso y obesidad en el mundo. ⁽³⁾

GRÁFICA 1. Sobrepeso y obesidad en menores de cinco años, escolares (cinco-11 años), adolescentes (12-19 años), mujeres (20-49 años) y adultos (>20 años) de 1988 a 2012

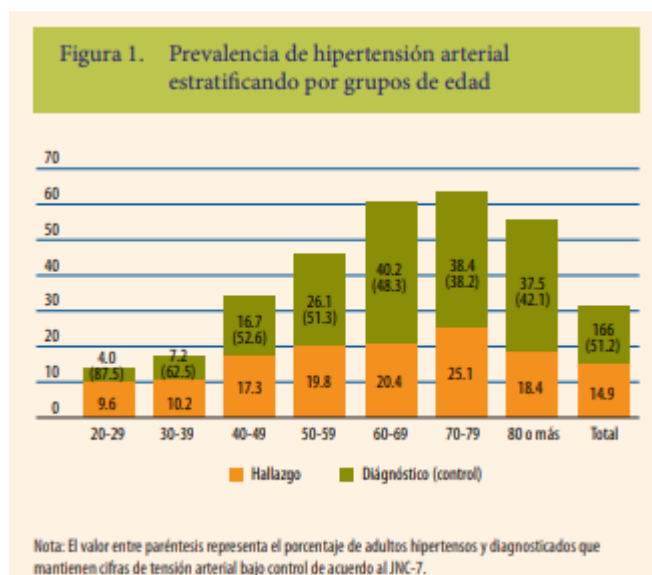


Fuente: Centro de Investigación en Nutrición y Salud (Instituto Nacional de Salud Pública).

En lo que respecta a la diabetes, ésta se posiciona como una de las primeras causas de muerte en México en 2010, esta enfermedad causó cerca de 83 000 muertes en el país. La proporción de adultos con diagnóstico previo de diabetes es de 9.2%* (ENSA 2000 fue de 4.6 %; ENSANUT 2006 fue de 7.3 %), los estados que presentan mayor prevalencia son Distrito Federal, Nuevo León, Veracruz, Tamaulipas, Durango y San Luis Potosí.⁽³⁾

El caso correspondiente a la hipertensión en México entre 2000 y 2006, la prevalencia de hipertensión incrementó 19.7% hasta afectar a 1 de cada 3 adultos, también durante la encuesta ENSANUT 2006 se mostró que la prevalencia es más alta en adultos con obesidad (42.3 %) que en adultos con índice de masa corporal normal (18.5 %), y en adultos con diabetes (65.6 %) que sin esta enfermedad (27.6 %).⁽⁸⁾

En la siguiente gráfica se observa la prevalencia de hipertensión arterial de la población mexicana por grupos de edad en el año 2006



El conocimiento sobre la magnitud, distribución y tendencia de las enfermedades crónicas (sobrepeso y obesidad, diabetes e hipertensión), sobre sus causas y los factores de riesgo en la población, permite generar recomendaciones en la alimentación a fin de diseñar políticas para su prevención y control.

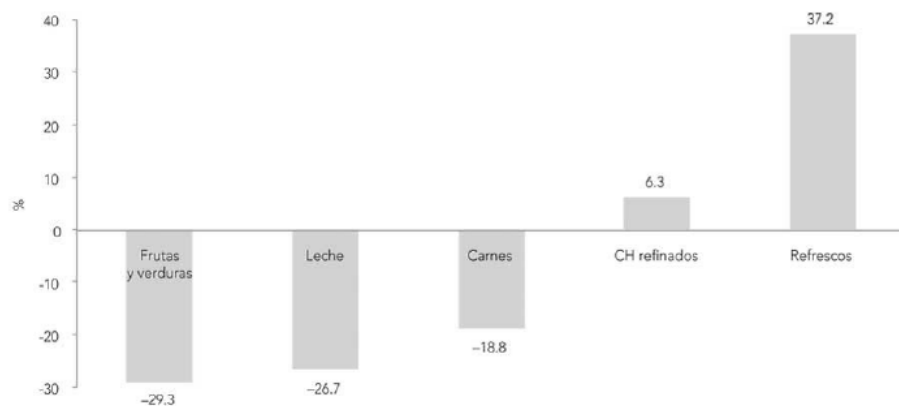
3.4. Preliminares de la relación pobreza - salud

La estabilidad de la salud física de un individuo tiene como base la alimentación y la actividad física, teniendo la primera un mayor peso, si se tiene una alimentación inadecuada la estabilidad de la salud puede verse alterada. Factores como la educación, distribución del dinero y los recursos son la causa de una correcta o mala alimentación. Las poblaciones pobres eligen alimentos más baratos que generan dietas pobres en nutrientes, surgiendo así afectaciones a la calidad del consumo que inciden en problemas de salud pública tales como el desarrollo de obesidad, diabetes e hipertensión.

El incremento de las enfermedades crónicas no es tanto una falla en los sistemas biológicos sino más bien un fenómeno social y económico, pues los problemas financieros impiden accesos a dietas más sanas, malas dietas conllevan a que existas altos niveles de obesidad y diabetes entre las minorías y los trabajadores pobres, situación que puede estarse presentando en México.⁽⁹⁾

Los aumentos en los precios de los alimentos afectan las decisiones tomadas por los hogares con respecto a la sustitución en el consumo de los productos de mayor calidad nutricional, por aquellos productos de menor costo y menor calidad nutricional pero con mayor contenido energético. Un estudio relativo a las encuestas de ingreso y gasto en los hogares realizadas por el Instituto Nacional de Estadística Geografía e Informática (INEGI) entre 1984 y 1998, se analizaron los cambios en la media de compra de alimentos por adulto equivalente durante el periodo de estudio y se encontró que la cantidad de alimentos comprados disminuyó para frutas y verduras (29.3 %), lácteos (26.7 %) y carnes (18.8 %), mientras que se registraron aumentos para productos altos en carbohidratos refinados y azúcares (6.3 %) y para refrescos (37.2 %).⁽³⁾

GRÁFICA 5. Cambios en la media de compras de alimentos por adulto equivalente en 1998 en relación con 1984 (%) por grupos de alimentos en México

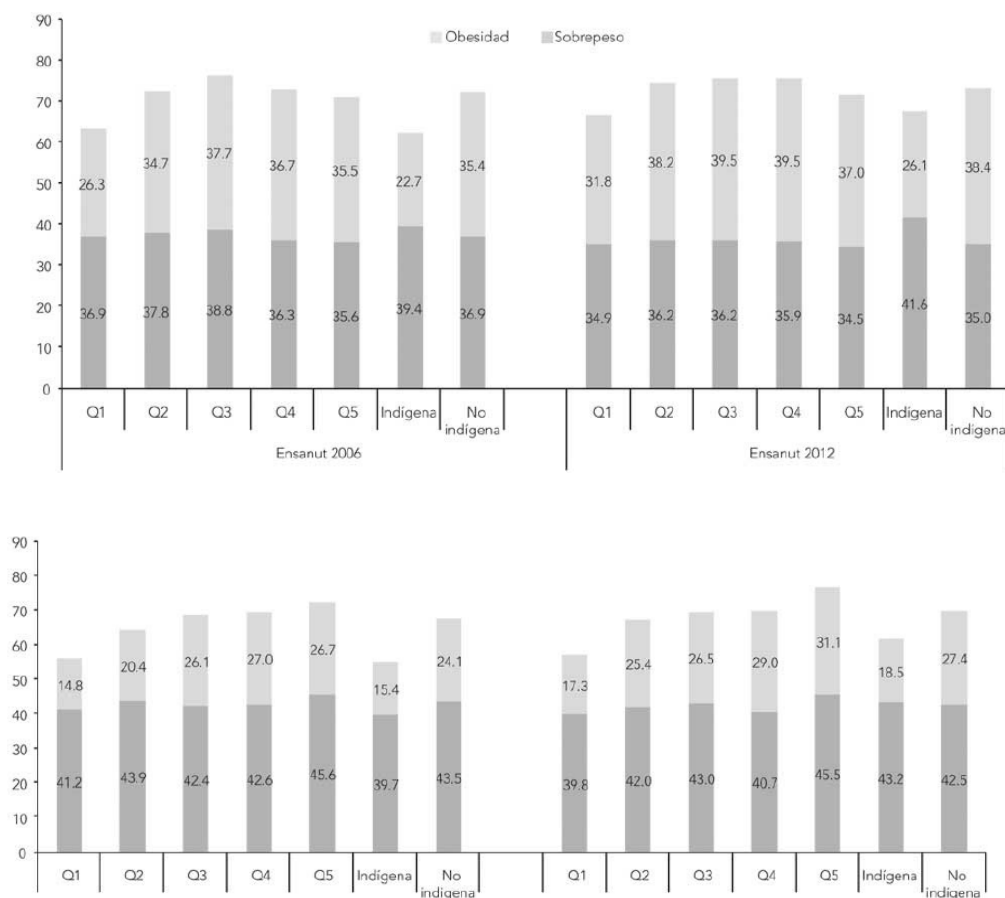


Fuente: Rivera JA y Col. Epidemiological and nutritional transition in Mexico: Rapid increase of non-communicable chronic diseases and obesity. *Public Health Nutrition*, 5 (1A) 113-122, 2002.

El ingreso no define por sí solo el patrón alimentario, pero influye en su estratificación y determina los niveles de consumo, permitiendo comprender hasta dónde un individuo o grupo familiar alcanzan su tope de gasto y las circunstancias económicas que definen el perfil de consumo; de aquí se derivan las preferencias, cambios alimentarios subsecuentes y las condiciones nutricionales de la estructura alimentaria. ⁽¹⁰⁾

La información sobre la distribución de las prevalencias en México parece indicar que los grupos en mayor desventaja socioeconómica tienen menores prevalencias de peso excesivo. Sin embargo, un análisis de los incrementos de sobrepeso y obesidad combinada entre 1988 y 1999 indica una mayor velocidad de aumento entre la población más marginada socialmente; por ejemplo, los aumentos más grandes (mayor que el 30 % por año) se observaron en zonas rurales y la población indígena, mientras que los mayores aumentos de obesidad entre 1999 y 2006 (mayor que el 15 % por año) también se observaron en la población indígena. ⁽³⁾

GRÁFICA 4. Sobrepeso y obesidad en adultos de 20 años o más por quintil socioeconómico (Q) e indigenismo. Ensanut-2006 y Ensanut-2012



Los patrones de consumo de los hogares son diferentes según el nivel de ingreso y varían en el tiempo. Los hogares con menores niveles de ingreso toman decisiones de consumo que les permiten obtener un mayor nivel de calorías a un precio menor, situación que se mantiene con el paso del tiempo dejando claro que la distribución del ingreso en México representa un problema estructural que se traduce en mejores o peores condiciones de acceso a la alimentación.

El estudio más reciente acerca del ingreso y gasto en la dieta realizado por el CONEVAL con información de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 1992, 2000 y 2010, nos muestra que existe una relación inversa entre niveles de pobreza y el costo con el cual los individuos pueden satisfacer su requerimiento mínimo de calorías: los hogares consumen calorías de menor precio cuando existen condiciones que hacen que vivan en condiciones de pobreza. Los hogares con menores niveles de ingreso pudieran estar en condiciones de aversión al riesgo al determinar sus elecciones de consumo de alimentos favoreciendo el contenido calórico y de saciedad en contraste de un contenido nutricional más adecuado.⁽⁵⁾

CUADRO 1. Porcentaje del gasto total en alimentos consumidos dentro del hogar por quintil de ingreso y nivel de densidad energética, México, 1992, 2000, 2010

Nivel de densidad energética	Quintil I	Quintil II	Quintil III	Quintil IV	Quintil V	Nacional
1992						
Baja	23.3*	27.1*	29.7*	31.6	32.8*	30.2*
Media	35.8*	42.9*	47.1	47.8*	48.8*	46.2
Alta	40.9*	30*	23.2	20.6	18.4*	23.6
Total	100.0	100.0	100.0	100.0	100.0	100.0
2000						
Baja	26.4	31.3*	33.6*	35.3*	36.3*	33.8*
Media	35.2*	43.9*	45.7	44.5*	43.6	43.5*
Alta	38.3*	24.7	20.6*	20.1*	20.1	22.7
Total	100.0	100.0	100.0	100.0	100.0	100.0
2010						
Baja	26.5	29.2	30.2	32.0	35.1	31.5
Media	41.4	45.7	46.7	46.5	44.9	45.3
Alta	32.1	25.1	23.1	21.5	20	23.2
Total	100.0	100.0	100.0	100.0	100.0	100.0

* Existe una diferencia estadísticamente significativa al 95% de confianza con respecto al valor de 2010.

Fuente: estimaciones propias utilizando las ENIGH 1992, 2000, 2010.

En el cuadro anterior el porcentaje del gasto por quintil de ingreso corriente indica que a menor nivel de ingreso, mayor es el porcentaje de consumo de alimentos con alta densidad energética, relación que es la contraria si se toma en cuenta alimentos con bajo o medio nivel de densidad energética.

CUADRO 5. Costo de mil calorías a pesos de agosto de 2010 por quintil de ingreso y nivel de densidad energética, México, 1992, 2000, 2010

Nivel de densidad energética	Quintil I	Quintil II	Quintil III	Quintil IV	Quintil V	Nacional
1992						
Baja	33.60*	28.70*	29.50*	31.20*	35.10	32.00*
Media	11.30*	12.30*	14.00*	17.10*	24.20*	16.60
Alta	3.40*	4.70*	5.50	6.70	8.80	5.50*
Total	6.40*	9.20*	11.60*	14.40*	19.80*	12.50*
2000						
Baja	26.80*	24.90*	25.40*	26.40*	29.40*	26.90*
Media	9.00*	10.00*	11.30	12.50	17.30*	12.50*
Alta	2.70*	4.10*	5.40*	6.50	7.70*	4.90*
Total	5.20*	8.60*	10.90	12.50	15.80	10.70*
2010						
Baja	26.50	25.80	25.80	27.40	30.90	27.80
Media	9.60	10.10	11.10	12.50	15.90	12.10
Alta	3.60	5.10	5.90	6.50	8.10	5.60
Total	7.00	9.40	10.70	12.20	15.60	11.10

* Existe una diferencia estadísticamente significativa al 95% de confianza con respecto al valor de 2010.

Fuente: estimaciones propias utilizando las ENIGH 1992, 2000, 2010.

Este mismo estudio también muestra la evolución del costo de satisfacer el requerimiento mínimo de 2135 calorías por día por persona comparada con la evolución del porcentaje de la población que vive en condiciones de pobreza alimentaria, ambas para el periodo que abarca de 1992 a 2010.

GRÁFICA 2. Evolución del costo de satisfacer el requerimiento mínimo de 2 135 calorías (precios de agosto de 2010) y porcentaje de población en pobreza alimentaria 1992 -2010



Fuente: Coneval (2011) y estimaciones propias utilizando las ENIGH 1992-2010.

Se observa que ambas series tienen un comportamiento de espejo a lo largo del tiempo; cuando los niveles de pobreza aumentan, los hogares consumen más calorías a un menor costo, lo que implica que pueden satisfacer el total de las calorías de su requerimiento a un menor costo. Existe una alta correlación entre ambas series de 1992 a 2010. ⁽³⁾

Estas marcadas diferencias en la estructura de consumo de los mexicanos de acuerdo con su nivel de ingreso, se muestra también en que la alimentación de los individuos más pobres está más orientada a alimentos con alto contenido energético, a diferencia de los individuos con mayores ingresos; los más pobres concentran su consumo energético en pan, tortillas y cereales, frutas y hortalizas, aceites y grasas vegetales comestibles, mientras que los individuos con mayor ingreso destinan un mayor porcentaje de consumo energético en grupos alimenticios como las carnes, leche, derivados de leche y huevo, azúcares y otros alimentos. ⁽³⁾

La ingesta excesiva de energía, en combinación con la baja actividad física, produce un elevado riesgo de la aparición de enfermedades crónicas relacionadas con la nutrición, complicaciones y muerte. Los aumentos de las prevalencias de la obesidad en México se encuentran entre los más rápidos documentados en el plano mundial, las enfermedades crónicas relacionadas con la nutrición son ahora el mayor problema en México.

Los distintos estudios revelan que existe una relación entre la salud y los ingresos, y que los sectores más pobres de la población son los más vulnerables considerando la frecuencia de enfermedades crónicas, las personas pobres están expuestas a un mayor riesgo y menos acceso a tratamientos.

4. Análisis Exploratorio de los Datos

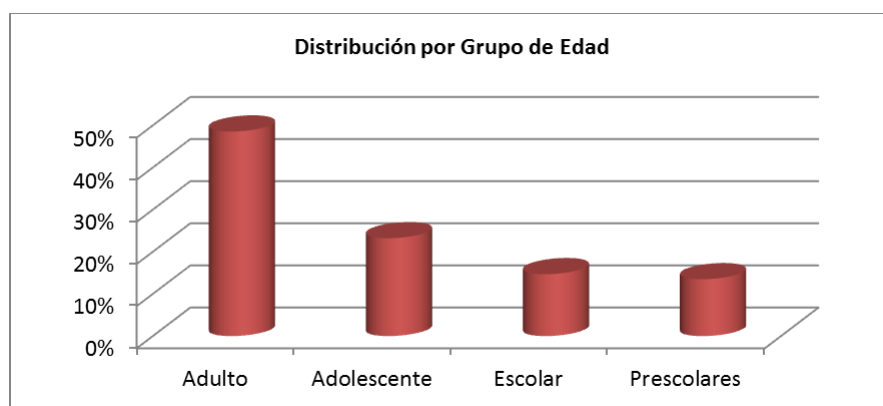
La Encuesta Nacional de Salud y Nutrición 2012 (ENSANUT 2012), permite contar con información detallada sobre el estado de salud y nutrición de la población, a partir de una muestra representativa de los mexicanos, tanto en lo que se refiere a su distribución geográfica, como por niveles socioeconómicos.

El tamaño de la muestra efectiva de la ENSANUT 2012 es de 50,528 hogares de las 32 entidades. Dicha encuesta ofrece información sobre la situación particular en las entidades federativas del país para los ámbitos urbano y rural.

De los 50,528 hogares se eligieron algunos integrantes de cada hogar para realizar pruebas antropométricas, el total de individuos fue de 85,291 con la siguiente distribución por edad:

Distribución		
Grupo Edad	Frecuencia	Proporción
Adulto	41,418	49 %
Aolescentes	19,814	23 %
Escolar	12,539	15 %
Prescolar	11,520	14 %
Total	85,291	100 %

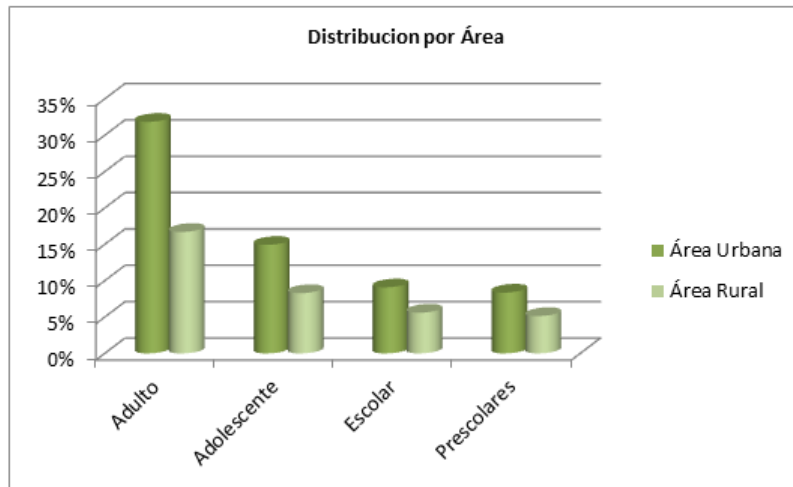
Los individuos que conforman la muestra fueron clasificados en cuatro grupos de edad: adultos que abarcan los individuos de 20 años en adelante que son 49 % de la muestra, adolescentes que son los individuos de 10 años hasta 19 con un porcentaje de 23 %, los escolares que son niños de 5 hasta 9 años con una participación en la muestra de 15 % y finalmente los prescolares que son individuos de 0 hasta 4 años con una participación de 14 %.



La distribución de la muestra por estado se puede observar en la siguiente tabla:

Entidad Federativa	Adulto	Adolescentes	Escolar	Prescolares	Total
Aguascalientes	1,284	650	413	412	2,759
Baja California	1,146	545	347	347	2,385
Baja California Sur	1,168	486	326	307	2,287
Campeche	1,420	645	405	387	2,857
Coahuila de Zaragoza	1,249	597	376	331	2,553
Colima	1,282	603	322	303	2,510
Chiapas	1,475	806	551	520	3,352
Chihuahua	1,085	520	307	279	2,191
Distrito Federal	1,206	500	293	285	2,284
Durango	1,227	558	328	316	2,429
Guanajuato	1,201	633	408	393	2,635
Guerrero	1,459	739	492	433	3,123
Hidalgo	1,302	603	371	363	2,639
Jalisco	1,289	611	422	375	2,697
México	1,297	647	426	366	2,736
Michoacan de Ocampo	1,329	620	417	377	2,743
Morelos	1,335	588	362	327	2,612
Nayarit	1,233	564	364	314	2,475
Nuevo León	1,267	615	347	317	2,540
Oaxaca	1,553	747	512	468	3,280
Puebla	1,273	660	445	416	2,794
Querétaro	1,278	599	400	369	2,646
Quintana Roo	1,353	651	400	380	2,784
San Luis Potosi	1,324	735	493	392	2,944
Sinaloa	1,095	505	298	305	2,203
Sonora	1,196	570	338	310	2,414
Tabasco	1,444	704	422	436	3,006
Tamaulipas	1,271	597	372	348	2,588
Tlaxcala	1,397	687	422	357	2,863
Veracruz de Ignacio de la Llave	1,431	673	406	335	2,845
Yucatán	1,398	624	387	313	2,395
Zacatecas	1,157	532	367	339	2,395
Total	41,418	19,814	12,539	11520	85,291

Donde para cada estado y cada grupo de edad le corresponde el 3 % de la muestra.



Donde de la población adulta le corresponde el 17 % para el área rural y el 32 % en el área urbana, los adolescentes tienen un 15 % para el área urbana y un 8 % de participación por parte del área rural, escolares un 9 % corresponde al área urbana y el 6 % al área rural asimismo el grupo de edad para prescolares cuenta con 8 % y 5 %, respectivamente.

4.1. Estadística descriptiva de la encuesta

El área de interés para este estudio es la medición de los individuos que cuentan con alguna enfermedad crónica tal como son el sobrepeso y la obesidad, diabetes o la hipertensión. En la encuesta de ENSANUT se realizaron algunos estudios (Pruebas antropométricas) a los 85,291 integrantes donde se tomaron muestras de peso y talla, presión arterial, muestras de sangre, mediciones de cintura, entre otras. A continuación se muestran las frecuencias:

4.1.1. Peso

Durante las pruebas antropométricas se tomo el peso y la talla de cada integrante, con el fin de obtener el Índice de Masa Corporal (IMC) y a través de la tabla del IMC realizada por la OMS hacer la clasificación de sobrepeso y obesidad, el IMC en la encuesta ENSANUT fue medido para adultos como:

$$IMC = \frac{Peso}{Talla^2}$$

Para los grupos de edad adolescentes, preescolares y escolares el IMC fue corregido utilizando la categorización de la puntuación z para el IMC para la edad, debido a que los niños no han alcanzado ni su máxima estatura ni el peso adecuado y continuará su crecimiento, la puntuación z consiste en calcular el IMC, como se marca en la ecuación anterior, el número del IMC se registra en las tablas de crecimiento de Centers for Disease Control and Prevention (CDC) para el IMC por edad (para niños o niñas) para obtener la categoría del percentil.

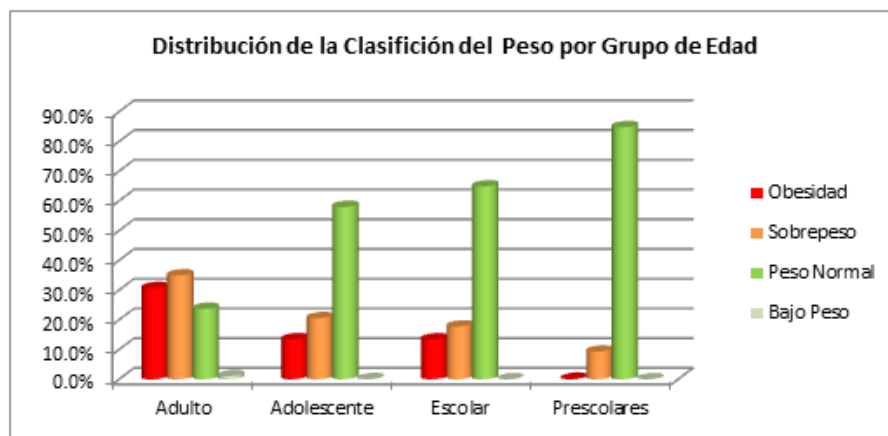
Los percentiles son el indicador que se utiliza con más frecuencia para evaluar el tamaño y los patrones de crecimiento de cada niño. El percentil indica la posición relativa del IMC del niño entre niños del mismo sexo y edad. Las tablas de crecimiento muestran las categorías del nivel de peso que se usan con niños y adolescentes (bajo peso, peso saludable, sobrepeso y obeso).

Las categorías del nivel de peso del IMC por edad y sus percentiles correspondientes se muestran en la siguiente tabla.

Categoría de nivel de peso	Rango del percentil
Bajo peso	Menos del percentil 5
Peso saludable	Percentil 5 hasta por debajo del percentil 85
Sobrepeso	Percentil 85 hasta por debajo del percentil 95
Obeso	Igual o mayor al percentil 95

Los resultados de las mediciones en la encuesta ENSANUT 2012 fueron:

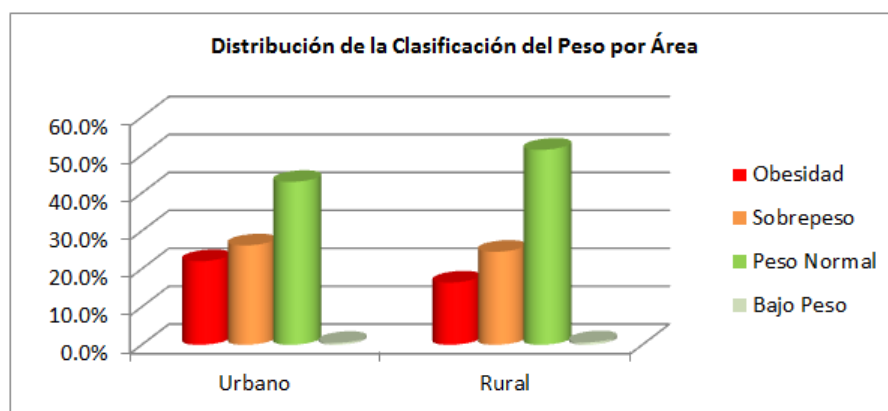
Peso Edad	Obesidad	Sobrepeso	Normal	Bajo Peso	Nulos
Adulto	30.7 %	34.9 %	23.7 %	1.1 %	9.6 %
Adolescentes	13.5 %	20.5 %	58 %	0.1 %	7.9 %
Escolar	13.4 %	17.7 %	64.9 %	0 %	3.9 %
Prescolares	0.2 %	9.3 %	84.8 %	0 %	5.7 %
Total	20 %	25.6 %	46 %	0.5 %	7.9 %



En el cuadro de frecuencias y la gráfica anterior se puede observar que el 45.6 % del total de la muestra tiene un problema de sobrepeso u obesidad, información alarmante. Por otro lado, para el caso específico de los adultos el 65.6 % presenta algún problema con su peso, si bien para los grupos de adolescentes y escolares la cifra no es tan estratosférica, ronda alrededor del 30 % para quienes presentan obesidad o sobrepeso, en el rango donde se observa el menor problema son para el grupo de prescolares.

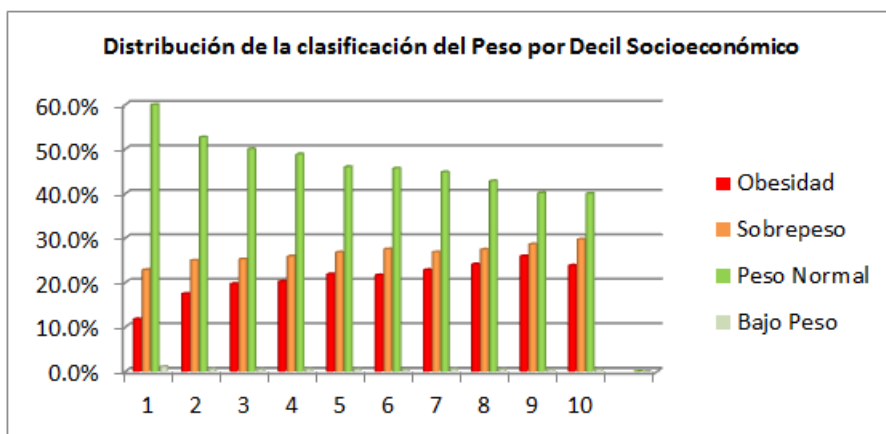
La distribución correspondiente por área rural o urbana, tenemos que el 64 % de nuestra muestra proviene de un ambiente urbano de los cuales 48 % presentan un problema de peso contra un 36 % de personas que se desarrollan en un ambiente rural con un porcentaje del 40.9 %.

Total	Peso Área	Obesidad	Sobrepeso	Normal	Bajo Peso	Nulos
64 %	Urbano	22 %	26.20 %	42.90 %	0.50 %	8.32 %
36 %	Rural	16.40 %	24.50 %	51.40 %	0.63 %	7.05 %



Adicionalmente, se realizó una distribución por decil de ingresos, es decir, para cada hogar de acuerdo a su ingreso total se le asignó un decil, siendo el decil uno el de menor ingreso y el decil 10 el de mayor ingreso.

La siguiente gráfica y cuadro de frecuencia muestran los resultados.



Decil Socioeconómico	Obesidad	Sobrepeso	Normal	Bajo Peso	Total
1	11.8 %	22.8 %	59.9 %	1.1 %	10.7 %
2	17.5 %	24.9 %	52.7 %	0.6 %	10.4 %
3	19.7 %	25.2 %	50 %	0.5 %	10.2 %
4	20.2 %	25.9 %	48.8 %	0.5 %	10.1 %
5	21.9 %	26.8 %	46 %	0.5 %	9.9 %
6	21.7 %	27.5 %	45.7 %	0.4 %	9.8 %
7	22.8 %	26.8 %	44.8 %	0.6 %	9.7 %
8	24.1 %	27.4 %	42.8 %	0.5 %	8.9 %
9	25.9 %	28.6 %	40.1 %	0.4 %	8.7 %
10	23.8 %	29.7 %	40 %	0.5 %	8.3 %

Podemos observar que a mayor decil de ingresos crece el número de personas con problemas en el peso, consistente con la disminución de personas con pesos estables.

4.1.2. Diabetes

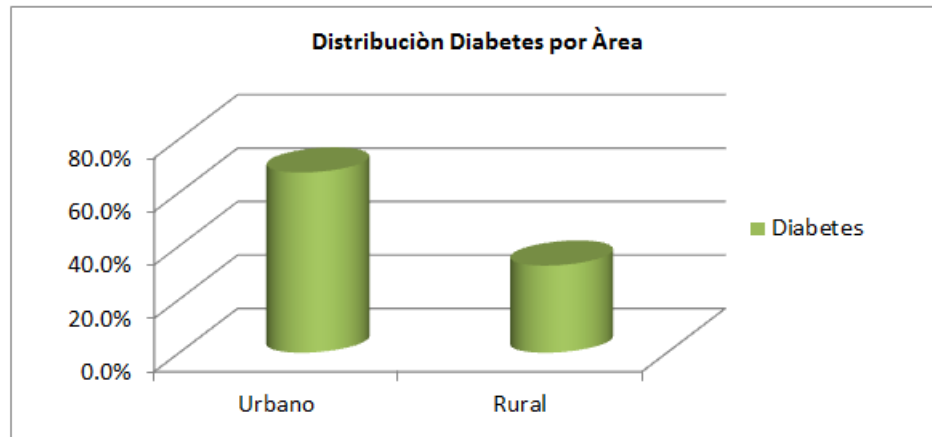
Para el diagnóstico de diabetes se utilizó el siguiente criterio en las pruebas de sangre: Si la glucosa es mayor 200 mg/dl en cualquier momento del día se diagnosticó Diabetes Mellitus.

Los resultados arrojados en la encuesta fueron:

Resultado Diabetes	Frecuencia	Proporción
No	11,635	13.64 %
No sabe	397	0.47 %
Si	930	1.09 %
No contesto	72,329	84.80 %
Total	85,291	100

Se observa que los individuos que presentan diabetes bajo las pruebas antropométricas, representan tan solo el 1 % de la muestra y corresponden únicamente al grupo de edad adultos, la distribución por área rural o urbana es la siguiente:

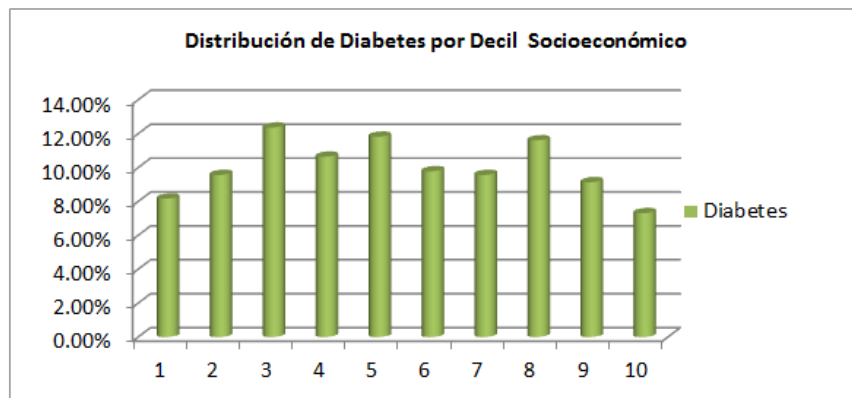
Área	Diabéticos	Proporción
Urbano	627	67.4 %
Rural	303	32.6 %
Total	930	100 %



Observamos que el 67% de la población que presenta diabetes se desarrolla en un área urbana contra un 32% que se desarrolla en áreas rurales.

La distribución correspondiente para los deciles de ingreso son:

Decil Socioeconómico	Diabetes	Proporción
1	76	8.17 %
2	89	9.57 %
3	115	12.3 %
4	99	10.65 %
5	110	11.83 %
6	91	9.78 %
7	89	9.57 %
8	108	11.61 %
9	85	9.14 %
10	68	7.31 %
Total	930	100 %



A diferencia de la Obesidad y Sobrepeso, la distribución de los individuos que presentan diabetes no es tan clara en los distintos deciles de ingreso pues podemos observar en el gráfico anterior que la mayor frecuencia se encuentra en el decil 3 seguida del 5 y 8 por lo que aún no podríamos decir que existe evidencia para relacionar que a mayor ingreso mayor presencia de individuos con diabetes.

4.1.3. Hipertensión

La hipertensión se diagnostica como una tensión sistólica igual o superior a 140 mm Hg y una tensión diastólica igual o superior a 90 mm Hg. Los resultados de las pruebas antropométricas fueron:

Resultado Presión	Frecuencia	Porcentaje
Null	64,840	75.948 %
No Cooperó	482	0.6 %
No se encontró	1,428	1.7 %
Problema Físico	44	0.052 %
Se negó	1,006	1.2 %
Sin problema	17,491	20.5 %
Total	85,291	100 %

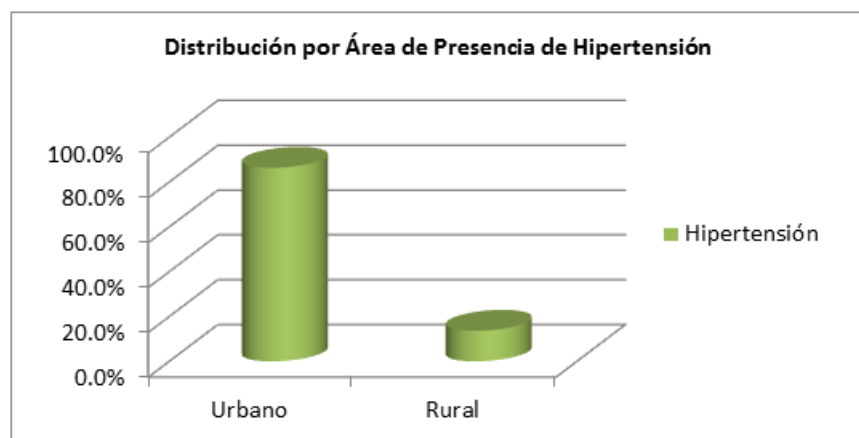
Como se puede observar en el cuadro anterior, no se cuenta con información suficiente para hacer el estudio pues la mayor parte de los resultados no fueron registrados.

Para poder realizar el estudio se tomaron los datos de diagnóstico, es decir, durante la encuesta que realizó ENSANUT se les preguntó a los individuos de cada hogar si se les había diagnosticado hipertensión en el último año o bien si tenían conocimiento de padecerla, las frecuencias de esta pregunta son:

Edad	Hipertensión	Porcentaje
Adultos	1,466	98.7 %
Adolescentes	20	1.3 %
Escolares		0 %
Prescolares		0 %
Total	1486	100 %

Contamos con 1486 diagnósticos de los cuales el 98 % pertenece al grupo de edad adultos y sólo el 1 % corresponde a adolescentes, la distribución por área rural o urbana es la siguiente:

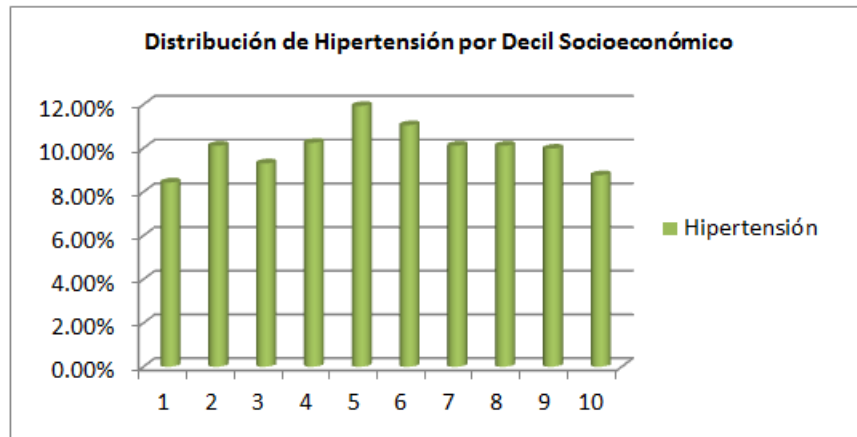
Área	Hipertensión	Porcentaje
Urbano	979	65.9 %
Rural	507	34.1 %
Total	1,486	100 %



Vemos una marcada presencia de la hipertensión en los habitantes que se desarrollan en áreas urbanas.

La distribución correspondiente a los deciles de ingreso es la siguiente:

Decil Socioeconómico	Hipertensión	Porcentaje
1	125	8.42 %
2	150	10.11 %
3	138	9.30 %
4	152	10.24 %
5	177	11.93 %
6	164	11.05 %
7	150	10.11 %
8	150	10.11 %
9	148	9.97 %
10	130	8.76 %
Total	1,484	100 %



De igual forma que para la Diabetes, tenemos una distribución por decil de ingresos para la hipertensión muy similar para todos los deciles de ingreso por lo que no podríamos garantizar un patrón de a mayor ingreso mayor presencia de individuos con hipertensión.

5. Métodos Estadísticos

5.1. Preliminares regresión logística múltiple

La regresión logística es un modelo diseñado específicamente para variables dependientes binarias, es decir aquellas donde la variable de respuesta no es numérica y corresponde a una designación de dos posibles categorías por ejemplo, vivo o muerto, enfermo o sano, hombre o mujer, fuma o no fuma, etc. El objetivo general de la investigación cuya modelación es la regresión logística es encontrar la probabilidad de que un objeto o individuo en la investigación tenga o no la respuesta dadas otras características de este individuo.

Si denominamos Y a nuestra variable de respuesta (dependiente) la cual solo puede tomar dos valores $Y = \textit{Éxito}$ o $Y = \textit{Fracaso}$ según hayamos categorizado el fenómeno de investigación. Por ejemplo si queremos saber si alguien tiene o no diabetes podríamos llamar éxito a tener diabetes y fracaso a no tenerlo.

En el lenguaje matemático Y es una variable Bernoulli donde $P[Y = \textit{Éxito}] = \Pi$ y $P[Y = \textit{Fracaso}] = 1 - \Pi$ en toda investigación contamos con un conjunto de variables que explican a nuestra variable de respuesta que comúnmente llamamos variables independientes, explicativas o regresores y las denotamos por $X_1, X_2, X_3, \dots, X_p$ la intención es modelar la probabilidad de respuesta como función de estos conjuntos de variables explicativas de la siguiente manera:

$$P[Y = \textit{Éxito} \mid X_1, X_2, \dots, X_p] = \Pi(X)$$

En el ejemplo de la diabetes podríamos pensar que nuestras variables explicativas podrían ser, el peso del individuo, su consumo de azúcar, la actividad física, etc.

El planteamiento del modelo inmediatamente nos lleva a pensar en un modelo de regresión lineal que es una de las herramientas más usadas por la estadística para explicar y relacionar fenómenos o variables, si partimos de esta idea tenemos lo siguiente:

$$P[Y = \textit{Éxito} \mid X_1, X_2, \dots, X_p] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p = \Pi(X)$$

La función que nos permite ajustar el rango de las variables es la función logística la cual es:

$$f(x) = \frac{\exp(x)}{1 + \exp(x)} \text{ con } x \in R$$

Si aplicamos la función logística al modelo anterior tenemos:

$$\begin{aligned} P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p] &= \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \\ &= \Pi(X) \end{aligned}$$

Con esto garantizamos que el valor de Π efectivamente sea una probabilidad. Sin embargo, nuestro modelo no es lineal y por lo tanto la interpretación es más complicada, por lo que vamos a buscar construir un modelo lineal.

Si construimos el momio correspondiente tenemos:

$$\begin{aligned} \frac{\Pi(X)}{1 - \Pi(X)} &= \frac{P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p]}{1 - P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p]} \\ &= \frac{\frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}{1 - \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} \\ &= \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \end{aligned}$$

Si aplicamos logaritmo de ambos lados tenemos:

$$\begin{aligned} \log\left(\frac{\Pi(X)}{1 - \Pi(X)}\right) &= \log\left(\frac{P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p]}{1 - P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p]}\right) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \end{aligned}$$

La ecuación anterior ya es un modelo lineal para el logaritmo del cociente de momios de la variable respuesta Y a la que suele llamarse *Logit* y este es el modelo de regresión logística para p regresores.

La transformación arriba descrita logra que los dominios de ambas partes del modelo sean compatibles, es decir, el modelo lineal liga una transformación de la variable respuesta con el predictor lineal $(X_i\beta_i)$.

Sin embargo, debemos tener en cuenta que aquí estamos trabajando con los momios de la variable respuesta Y y no necesariamente con la probabilidad de éxito de la variable respuesta, veamos la interpretación del modelo.

5.2. Interpretación del modelo

Para realizar la interpretación de los parámetros situémonos en el modelo más simple y después realizamos la generalización al modelo con más de un regresor.

Supongamos que tenemos solo un regresor y éste es dicotómico:

$$\log \left(\frac{P(Y = 1 | X = 1)}{1 - P(Y = 1 | X = 1)} \right) = \beta_0 + (1)\beta_1 = \beta_0 + \beta_1$$

y

$$\log \left(\frac{P(Y = 1 | X = 0)}{1 - P(Y = 1 | X = 0)} \right) = \beta_0 + (0)\beta_1 = \beta_0$$

Restando ambas ecuaciones tenemos

$$\begin{aligned} \log \left(\frac{P(Y = 1 | X = 1)}{1 - P(Y = 1 | X = 1)} \right) - \log \left(\frac{P(Y = 1 | X = 0)}{1 - P(Y = 1 | X = 0)} \right) &= \beta_0 + \beta_1 - \beta_0 \\ \log \left(\frac{\frac{P(Y=1 | X=1)}{1-P(Y=1 | X=1)}}{\frac{P(Y=1 | X=0)}{1-P(Y=1 | X=0)}} \right) &= \beta_1 \end{aligned}$$

Aplicando exponencial tenemos:

$$\frac{\frac{P(Y=1 | X=1)}{1-P(Y=1 | X=1)}}{\frac{P(Y=1 | X=0)}{1-P(Y=1 | X=0)}} = \exp(\beta_1)$$

La expresión anterior es el cociente de los momios cuando $X = 0$ y cuando $X = 1$ por lo que $\exp(\beta_1)$ se interpreta como el cambio entre el momio de la respuesta cuando $X = 1$ con respecto al momio de la respuesta cuando $X = 0$; es decir, si el resultado fuera 5 se dirá que es 5 veces más grande el momio de la respuesta cuando $X = 1$ que el momio de la respuesta cuando $X = 0$.

Al estimar β_1 , es decir, $\hat{\beta}_1$ la interpretaremos como el cambio promedio o cambio esperado en el momio de respuesta para $X = 1$ con respecto del momio de respuesta cuando $X = 0$.

Cuando la variable X es discreta o continua se tomarán dos valores consecutivos de la misma, $X = x$ y $X = x + 1$ y se realizara el mismo proceso, por lo que la expresión será igual a la anterior sólo que ahora el momio corresponde a dos valores consecutivos de la variable. Entonces la interpretación es el cambio en el momio de respuesta por unidad de cambio en X la variable explicativa o regresor.

La regresión logística recurre a los momios porque son medidas estandarizadas que permite comparar el nivel de influencia o fortaleza de las variables independientes o regresores sobre la variable dependiente, cuanto más se aleja de uno más fuerte es la relación entre las dos variables.

Debemos resaltar que la interpretación se hace como una comparación de los momios que generan los regresores y no en términos de probabilidades, es importante recordar que uno de los objetivos del modelo es proporcionar la probabilidad de respuesta $Y = \acute{E}xito$ para un conjunto determinado de regresores o variables explicativas, las cuales se obtienen de la siguiente manera:

$$\begin{aligned}\Pi(X) &= P[Y = \acute{E}xito \mid X_1, X_2, \dots, X_p] \\ &= \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}\end{aligned}$$

Un inconveniente de la obtención de las probabilidades es cuando en el modelo hay al menos una variable continua, entonces podemos tener tantas probabilidades diferentes como individuos en el estudio.

5.3. Estimación de los parámetros

Los parámetros $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ de la regresión logística se estiman a través del método de Máxima Verosimilitud.

La idea de este método es la de encontrar primero la función de densidad conjunta de todas las observaciones a la que llamamos función de verosimilitud.

Supóngase que se tiene una muestra $x_1, x_2, x_3, \dots, x_n$ de n sujetos con variable de respuesta Y y con p variables explicativas o regresores (X_1, X_2, \dots, X_p) con función de probabilidad $f(x; \theta)$. Entonces la función de verosimilitud se define como:

$$L(\theta, x) = \prod_{i=1}^n f(x; \theta)$$

Ya que Y se distribuye Bernoulli con probabilidad $\Pi(X)$, entonces su función de densidad es:

$$f(Y) = \Pi(X)^y * (1 - \Pi(X))^{1-y}$$

Sustituyendo en la función de verosimilitud tenemos:

$$L(\beta_0, \beta_1, \beta_2, \dots, \beta_p | \mathbf{y}, \mathbf{x}) = L(\theta, x) = \prod_{i=1}^n \Pi(x_i)^{y_i} * (1 - \Pi(x_i))^{1-y_i}$$

Aplicando logaritmo a ambos lados de la ecuación de verosimilitud tenemos:

$$\log(L(\beta_0, \beta_1, \beta_2, \dots, \beta_p | \mathbf{y}, \mathbf{x})) = \log\left(\prod_{i=1}^n \Pi(x_i)^{y_i} * (1 - \Pi(x_i))^{1-y_i}\right)$$

Usando propiedades de logaritmo tenemos:

$$\begin{aligned}
\log(L(\beta_0, \beta_1, \beta_2, \dots, \beta_p | \mathbf{y}, \mathbf{x})) &= \sum_{i=1}^n y_i \log(\Pi(x_i)) + (1 - y_i) \log(1 - \Pi(x_i)) \\
&= \sum_{i=1}^n y_i \log(\Pi(x_i)) - y_i \log(1 - \Pi(x_i)) + \log(1 - \Pi(x_i)) \\
&= \sum_{i=1}^n y_i \log\left(\frac{\Pi(x_i)}{1 - \Pi(x_i)}\right) + \log(1 - \Pi(x_i))
\end{aligned}$$

pero

$$\begin{aligned}
\Pi(x_i) &= \frac{\exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)} \\
1 - \Pi(x_i) &= 1 - \frac{\exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)} \\
&= \frac{1 + \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right) - \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)} \\
&= \frac{1}{1 + \exp\left(\sum_{j=1}^n \beta_j x_{ij}\right)}
\end{aligned}$$

Suatiuyendo en la ecuación el valor de $\Pi(x_i)$ tenemos:

$$\begin{aligned}
&\sum_{i=1}^n y_i \log\left(\frac{\exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}\right) + \log\left(\frac{1}{\exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}\right) \\
&= \sum_{i=1}^n y_i \log\left(\frac{(1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)) \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}\right) + \log\left(\frac{1}{\exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)}\right) \\
&= \sum_{i=1}^n y_i \left(\sum_{j=0}^n \beta_j x_{ij}\right) + \log(1) - \log\left(1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)\right) \\
&= \sum_{i=1}^n \left(\sum_{j=0}^n y_i \beta_j x_{ij}\right) - \log\left(1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)\right) \\
&= \sum_{i=1}^n \left(\sum_{j=0}^n y_i \beta_j x_{ij}\right) - \sum_{i=1}^n \log\left(1 + \exp\left(\sum_{j=0}^n \beta_j x_{ij}\right)\right)
\end{aligned}$$

Para encontrar cada β_j tenemos que derivar con respecto a β_j e igualar a cero.

$$\begin{aligned}
\frac{\partial}{\partial \beta_j} (\log(L(\beta_0, \beta_1, \dots, \beta_p \mid \mathbf{y}, \mathbf{x}))) &= \frac{\partial}{\partial \beta_j} \left(\sum_{i=1}^n \left(\sum_{j=0}^n y_i \beta_j x_{ij} \right) - \sum_{i=1}^n \log \left(1 + \exp \left(\sum_{j=0}^n \beta_j x_{ij} \right) \right) \right) \\
&= \frac{\partial}{\partial \beta_j} \left(\sum_{i=1}^n \left(\sum_{j=0}^n y_i \beta_j x_{ij} \right) \right) - \frac{\partial}{\partial \beta_j} \left(\sum_{i=1}^n \log \left(1 + \exp \left(\sum_{j=0}^n \beta_j x_{ij} \right) \right) \right) \\
&= \sum_{i=1}^n (y_i x_{ij}) - \sum_{i=1}^n \left(\frac{1}{1 + \exp(\sum_{j=0}^n \beta_j x_{ij})} * \exp \left(\sum_{j=0}^n \beta_j x_{ij} \right) * x_{ij} \right) \\
&= \sum_{i=1}^n (y_i x_{ij}) - \sum_{i=1}^n \left(\frac{\exp(\sum_{j=0}^n \beta_j x_{ij}) * x_{ij}}{1 + \exp(\sum_{j=0}^n \beta_j x_{ij})} \right) \\
&= \sum_{i=1}^n x_{ij} \left(y_i - \frac{\exp(\sum_{j=0}^n \beta_j x_{ij})}{1 + \exp(\sum_{j=0}^n \beta_j x_{ij})} \right) = \sum_{i=1}^n x_{ij} (y_i - \Pi(x_i)) = 0
\end{aligned}$$

Así obtenemos

$$\frac{\partial}{\partial \beta_j} (\log(L(\beta_0, \beta_1, \dots, \beta_p \mid \mathbf{y}, \mathbf{x}))) = \sum_{i=1}^n x_{ij} (y_i - \Pi(x_i)) = 0$$

Así que obtenemos lo siguiente:

$$\begin{aligned}
\frac{\partial}{\partial \beta_j} (\log(L(\beta_0, \beta_1, \dots, \beta_p \mid \mathbf{y}, \mathbf{x}))) &= \begin{pmatrix} \frac{\partial \log(\beta_0, \beta_1, \dots, \beta_p \mid \mathbf{y}, \mathbf{x})}{\partial \beta_0} \\ \vdots \\ \frac{\partial \log(\beta_0, \beta_1, \dots, \beta_p \mid \mathbf{y}, \mathbf{x})}{\partial \beta_p} \end{pmatrix} \\
\begin{pmatrix} \sum_{i=1}^n x_{i0} (y_i - \Pi(x_i)) \\ \vdots \\ \sum_{i=1}^n x_{ip} (y_i - \Pi(x_i)) \end{pmatrix} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}
\end{aligned}$$

Para esto tendríamos una derivada diferente para cada β_i , así tendríamos un sistemas de ecuaciones de tamaño $p + 1 \times n$, el cual resolveremos por métodos numéricos.

Vamos a escribir la ecuación en forma vectorial.

$$\begin{aligned}
\begin{pmatrix} \sum_{i=1}^n x_{i0}(y_i - \Pi(x_i)) \\ \vdots \\ \sum_{i=1}^n x_{ip}(y_i - \Pi(x_i)) \end{pmatrix} &= \begin{pmatrix} x_{10}(y_1 - \Pi(x_1)) + x_{20}(y_2 - \Pi(x_2)) + \dots + x_{n0}(y_1 - \Pi(x_n)) \\ \vdots \\ x_{10}(y_1 - \Pi(x_1)) + x_{20}(y_2 - \Pi(x_2)) + \dots + x_{n0}(y_1 - \Pi(x_n)) \end{pmatrix} \\
&= \begin{pmatrix} x_{10} & x_{20} & \dots & x_{n0} \\ \vdots & \vdots & \dots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} y_i - \Pi(x_1) \\ \vdots \\ y_n - \Pi(x_n) \end{pmatrix} \\
&= \begin{pmatrix} x_{10} & x_{20} & \dots & x_{n0} \\ \vdots & \vdots & \dots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \left[\begin{pmatrix} y_i \\ \vdots \\ y_2 \end{pmatrix} - \begin{pmatrix} \Pi(x_1) \\ \vdots \\ \Pi(x_n) \end{pmatrix} \right] = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}
\end{aligned}$$

Consideremos lo siguiente:

$$\mathbf{X} = \begin{pmatrix} x_{10} & x_{20} & \dots & x_{n0} \\ \vdots & \vdots & \dots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} y_i \\ \vdots \\ y_2 \end{pmatrix} \quad y \quad \Pi(\mathbf{X}, \boldsymbol{\beta}) = \begin{pmatrix} \Pi(x_1) \\ \vdots \\ \Pi(x_n) \end{pmatrix}$$

Donde \mathbf{X} es de tamaño $p+1 \times n$, que será la matriz de diseño que contiene los valores de las variables independientes o regresores de los individuos en la muestra. Las ecuaciones de verosimilitudes anteriores, pueden escribirse como:

$$\mathbf{f}(\boldsymbol{\beta}) = \mathbf{X}(\mathbf{Y} - \Pi(\mathbf{X}, \boldsymbol{\beta})) = \mathbf{0}$$

El estimador de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, corresponden a la solución del sistema anterior. Para resolverlo se utiliza el algoritmo de Newton-Raphson que requiere la matriz de segundas derivadas, que viene dada por $\frac{\partial^2 L(\beta_0, \beta_1, \beta_2, \dots, \beta_p \mid \mathbf{y}, \mathbf{x})}{\partial \beta \partial \beta'}$ los elementos de esta matriz tienen la forma.

$$\begin{aligned}
\frac{\partial^2 L(\beta_0, \beta_1, \beta_2, \dots, \beta_p \mid \mathbf{y}, \mathbf{x})}{\partial \beta_i \partial \beta_k} &= - \sum_{i=1}^n \frac{\partial}{\partial \beta_k} [x_{ij}(y_i - \Pi(x_i))] \\
&= - \sum_{i=1}^n x_{ij}(y_i - \Pi(x_i))(1 - x_{ij}(y_i - \Pi(x_i)))x_{ik} \\
&= \sum_{i=1}^n \Pi(\mathbf{x}_i, \boldsymbol{\beta})(1 - \Pi(\mathbf{x}_i, \boldsymbol{\beta}))x_{ij}x_{ik}
\end{aligned}$$

Si redefinimos $\mathbf{W}_{ii}(\boldsymbol{\beta}) = \Pi(\mathbf{x}_i, \boldsymbol{\beta})(1 - \Pi(\mathbf{x}_i, \boldsymbol{\beta}))$ y $\mathbf{W}_{ik} = 0$.
 Así tendremos que resolver un sistema de ecuaciones de verosimilitud con

$$\mathbf{U}(\boldsymbol{\beta}) = -\mathbf{X}^t \mathbf{W}_{ii}(\boldsymbol{\beta}) \mathbf{X}$$

El algoritmo de Newton-Raphson opera de manera iterativa hasta que se logre la convergencia, iterando del paso $r - 1$ al r , a través de la ecuación

$$\boldsymbol{\beta}^{(r)} = \boldsymbol{\beta}^{(r-1)} - \mathbf{U}(\boldsymbol{\beta}^{(r-1)})^{-1} \mathbf{U}(\boldsymbol{\beta}^{(r-1)})$$

Donde empezaremos con un valor inicial $\boldsymbol{\beta}^{(0)}$ del estimador, $r - \text{ésimo}$, paso del proceso iterativo es:

$$\hat{\boldsymbol{\beta}}^{(r)} = \hat{\boldsymbol{\beta}}^{(r-1)} + \left(\mathbf{X}^t \mathbf{W}_{ii}(\hat{\boldsymbol{\beta}}^{(r-1)}) \mathbf{X} \right)^{-1} \mathbf{X}^t \left(\mathbf{y} - \Pi(\mathbf{X}, \hat{\boldsymbol{\beta}}^{(r-1)}) \right)$$

5.4. Pruebas de bondad de ajuste del modelo

5.4.1. Significancia de los parámetros

Una vez estimados los parámetros del modelo, se realizan pruebas para determinar que el modelo ajusta a los datos observados. La primera prueba es para determinar la significancia o importancia de nuestras variables explicativas o regresores, es decir qué tanto cada una de ellas explica la variable respuesta o variable independiente.

Para ello debemos plantear nuestra prueba de hipótesis

$$H_0 : \beta_1 = \dots = \beta_p = 0 \quad \text{vs} \quad H_1 : \exists \text{ al menos } \beta_j \neq 0$$

Con ello tratamos de corroborar que al menos uno de nuestros regresores o variables explicativas, explica la variable respuesta o variable dependiente por lo que al menos existe un $\beta_j \neq 0$, es decir, lo que buscamos es rechazar la hipótesis nula, sin embargo se debe hacer una prueba individual para cada β_j para saber cuáles y cuántos de nuestros regresores o variables explicativas son importantes para explicar la respuesta.

Debemos recordar que el vector de parámetros estimados $\hat{\beta}$ fue estimado por máxima verosimilitud y los estimadores de máxima verosimilitud tienen las siguientes propiedades:

- En caso de existir un estimador insesgado y eficiente (varianza igual a la cota de Cramér-Rao), dicho estimador es el máximo verosímil.
- Si el estimador máximo verosímil es sesgado, su sesgo tiende a cero al aumentar el tamaño muestral.
- La distribución del estimador máximo verosímil es asintóticamente Normal con esperanza igual al valor del parámetro estimado y de varianza igual a la cota de Cramér-Rao.
- En caso de existir un estadístico suficiente para un parámetro, el estimador máximo verosímil es función de dicho estadístico suficiente.
- Son invariantes: si U es el estimador máximo verosímil de θ y g es una función biunívoca, entonces $g(U)$ es el estimador máximo verosímil de $g(\theta)$.

Haciendo uso de la propiedad 3 para cada $\hat{\beta}_j$ con $j \in (1, \dots, p)$ tiene una distribución Normal con esperanza $\hat{\beta}_j$ y varianza igual a la cota inferior de Cramér-Rao) es decir:

$$\hat{\beta}_j \sim N\left(\hat{\beta}_j, \hat{V}(\hat{\beta}_j)\right)$$

Donde

$$\hat{V}(\hat{\beta}_j) = \frac{1}{E\left[\left(\frac{\partial}{\partial \beta_j} \ln(L(X, \hat{\beta}_j))\right)^2\right]}$$

Por lo que el estadístico de prueba es una t-Student con $n - p$ grados de libertad:

$$T\text{Estadístico} = \frac{\hat{\beta}_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \approx t_{n-p} \text{ con } n \text{ tamaño de la muestra}$$

De tal forma que si $T\text{Estadístico} > t_{n-p}$ se rechaza hipótesis nula H_0 , que es equivalente a que si $p - \text{value} < 0.05$ se rechaza la hipótesis nula.

Si el tamaño de muestra es lo suficientemente grande entonces t converge a z por lo que también la prueba se puede realizar con el estadístico.

$$\frac{\hat{\beta}_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \approx N(0, 1)$$

5.4.2. Diagnóstico del modelo logístico

Una vez que se ha encontrado un modelo que se cree ajusta nuestros datos, es necesario determinar si el modelo los ajusta adecuadamente, es decir, realizar la prueba de hipótesis:

H_0 : *El modelo ajusta los datos vs* H_1 : *El modelo no ajusta los datos*

Para ello se utiliza la Estadística de Hosmer-Lemeshow, esta prueba de bondad de ajuste es una prueba donde se evalúa la distancia entre lo observado en los datos que tenemos de la realidad y lo esperado bajo el modelo.

La prueba de bondad de ajuste de Hosmer-Lemeshow está basada en las probabilidades estimadas por el modelo ($\hat{\Pi}_1, \hat{\Pi}_2, \dots, \hat{\Pi}_n$) la idea es agrupar a los sujetos u observaciones mediante sus probabilidades estimadas para tener g grupos de ellos y realizar una prueba tipo χ^2 ya que este es el estadístico que sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias.

Si denotamos por:

$$OE_g = \sum_{i=1}^{n_g} y_i, OF_g = \sum_{i=1}^{n_g} (1 - y_i), EE_g = \sum_{i=1}^{n_g} \hat{\Pi}_i, EF_g = \sum_{i=1}^{n_g} (1 - \hat{\Pi}_i)$$

- OE_g La suma de los valores observados de éxitos en el grupo g
- OF_g La suma de los valores observados de fracasos en el grupo g
- EE_g La suma de los valores esperados de éxitos en el grupo g
- EF_g La suma de los valores esperados de fracasos en el grupo g
- n_g El número de sujetos u observaciones en el grupo g

Esto puede ser representado en una tabla de $2 \times g$ como sigue:

variable Independiente		Grupos					
		1	2	3	g	Total
y=1	Observados	OE_{11}	OE_{12}	OE_{13}	OE_{1g}	OE_1
	Esperados	EE_{11}	EE_{12}	EE_{13}	EE_{1g}	EE_1
y=0	Observados	OE_{01}	OE_{02}	OE_{03}	OE_{0g}	OE_0
	Esperados	EE_{01}	EE_{02}	EE_{03}	EE_{0g}	EE_0
Total		$\approx n/g$	$\approx n/g$	$\approx n/g$	$\approx n/g$	n

Realizando una prueba χ^2 con la tabla de valores esperados y observados tenemos que la estadística de prueba por Hosmer-Lemeshow

$$\chi_{HL}^2 = \sum_{j=1}^g \frac{(OE_j - EE_j)^2}{EE_j} + \sum_{j=1}^g \frac{(OF_j - EF_j)^2}{EF_j} \sim \chi_{(g-2)}^2$$

De forma que si $\chi_{HL}^2 < \chi_{(g-2)}^2$ entonces no se rechaza H_0 que es equivalente a que el $p - value$ sea mayor que 0.05

5.4.3. Curva de ROC

Una vez que hemos comprobado que las variables independientes o regresores del modelo realmente contribuyen a explicar la variable respuesta o independiente y que hemos visto que el modelo está bien ajustado, entonces debemos diagnosticar si el modelo tiene una buena capacidad predictiva, es decir, si puede pronosticar los éxitos y los fracasos. La siguiente tabla muestra la estructura de clasificación donde los renglones son los valores de la variable independiente observados y las columnas los valores predichos por el modelo.

Clasificación Observada	Clasificación Predictiva	
	Éxitos	Fracasos
Éxitos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Fracasos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Donde:

VP: Casos que fueron predichos por el modelo como éxitos y que realmente se observó eran éxitos

FP: Casos que fueron predichos como éxitos y que realmente se observaron como fracasos

FN: Casos que fueron predichos como fracasos y que realmente fueron observados como éxitos

VN: Casos que fueron predichos como fracasos y que realmente fueron observados como fracasos

Para determinar si el modelo es un buen predictor se reflejará a través de:

$$Sensitividad = \frac{Verdaderos\ Positivos}{Total\ de\ Positivos}$$

$$Especificidad = \frac{Verdaderos\ Negativos}{Total\ de\ Negativos}$$

Es decir para modelos que realizan una buena predicción los porcentajes de sensibilidad y especificidad deben ser muy altos.

La forma en que el modelo clasifica las observaciones es a través del punto de corte en el rango de las probabilidades, de tal forma que, si la probabilidad predicha de un sujeto por el modelo es mayor que el punto de corte, se clasifica como respuesta exitosa y si la probabilidad predicha de un sujeto por el modelo es menor que el punto de corte, se clasifica como negativa.

La curva de ROC es una gráfica de la sensibilidad (tasa de VP , eje y) contra la especificidad (tasa de FP , eje x) en una serie de umbrales o puntos de corte para la probabilidad de respuesta positivas. Cuanto más cercana esté la curva al borde izquierdo de la gráfica y al borde superior de la misma más adecuado y preciso es el modelo, cuanto más cercano esté la curva a la diagonal menos preciso, ya que la diagonal representa la situación en la que la predicción del modelo es totalmente al azar.

5.4.4. Área bajo la curva de ROC

La curva de ROC puede resumirse a través de su área bajo la curva (AUC), que se calcula por medio de la regla del trapecio:

$$AUC = \sum_{i=1}^n \frac{(x_{i+1} - x_i)(y_{i+1} + y_i)}{2}$$

Donde y_{i+1} y y_i son las bases del trapecio y $x_{i+1} - x_i$ es la altura con i particiones donde se calcula el área bajo la curva. Hay que tener en cuenta que el área bajo la diagonal es de 0.5 por lo que la curva de ROC debe definir un área por lo menos tan grande como este valor.

Entonces, el área bajo la curva de ROC (AUC) mide la capacidad de discriminación del modelo: la capacidad del modelo para pronosticar y clasificar correctamente a las respuestas positivas ($Y = 1$) y las respuestas negativas ($Y = 0$). No existe una prueba estadística de las AUC , cuyo juicio depende del campo de aplicación. Por regla general, el área puede ser calificada en una escala de 0 a 10, como 6 es bueno, 8 es muy bueno y 9 es excelente.

6. Construcción del Modelo

Como se mencionó al principio de este trabajo, el objetivo es conocer si factores socioeconómicos tales como el decil socioeconómico, la marginalidad, características de vivienda, estructura sociodemográfica, son causa de la presencia de enfermedades crónicas como la obesidad y sobrepeso, hipertensión y diabetes en los hogares mexicanos.

Esta relación se verificó a través de aplicar una Regresión Logística a los datos obtenidos de la Encuesta Nacional de Salud y Nutrición 2012 (ENSA-NUT2012).

Hipótesis:

1. Las características socioeconómicas son causa de enfermedades crónicas degenerativas en los hogares mexicanos.
2. A mayor nivel socioeconómico mayor probabilidad de presencia de enfermedades crónicas degenerativas.

6.1. Construcción de la base de datos

La fuente de información es ENSANUT2012 la cual es una base de datos que está constituida por diferentes tablas. Se realizaron una serie de cruces entre las tablas para crear una base de datos únicamente con las variables o regresores en el modelo.

La información se extrajo de las siguientes tablas:

Variable	Tabla de extracción de la información
Folio Consecutivo	Hogar
Parentesco	Integrantes
Grupo Edad	Construida
Área	Hogar
Marginalidad	Hogar
Obesidad	Antropometría trabajadas para cada grupo de edad
Hipertensión	Integrantes
Diabetes	Integrantes
Cuartos Dormir	Hogar
Decil Socioeconómico	Antropometría trabajada para cada grupo de edad

- **Hogar:** La tabla “Hogar” contiene información general de los hogares tales como gastos, características de la vivienda, seguridad social, características sociodemográficas, etc.
- **Integrantes:** Contiene la información de todos los integrantes que conforman cada hogar tal como la edad, ingresos, nivel de educación, empleo, parentesco, estado civil, etc.
- **Antropometría trabajada por cada grupo de edad:** Para cada grupo de edad ENSANUT trabajó la información obtenida de las pruebas antropométricas para definir variables tales como IMC, presencia de diabetes, talla, peso, etc.

Se tomó una reemuestra aleatoria del 5% de los individuos que conforman la base de integrantes que participaron en las pruebas antropométricas de la encuesta de ENSANUT2012, debido a que el volumen de la información desvía los resultados, así como los problemas del manejo de la base de datos para el paquete estadístico.

6.2. Especificación de las variables

- **Obesidad:** Es una variable dicotómica que indica 1 si el Individuo fue clasificado con problemas de obesidad o sobrepeso y 0 en el caso opuesto.
- **Diabetes:** Es una variable dicotómica que indica 1 si el individuo fue clasificado con problemas de diabetes y 0 en el caso opuesto.
- **Hipertensión:** Es una variable dicotómica que indica 1 si el Individuo fue clasificado con problemas de hipertensión y 0 en el caso opuesto.
- **Marginalidad:** Está clasificada en Alta o Baja, se refiere a la situación de desventaja que tiene un individuo en el ámbito social y económico.
- **Área:**
 - **Urbano:** Localidades con 2500 o más habitantes. Se etiquetaron bajo este rubro a la categoría metropolitana marcada en las bases de ENSANUT2012.
 - **Rural:** Localidades con menos de 2500 habitantes.
- **Cuartos Dormir:** La cantidad de habitaciones que se declararon son utilizadas para dormir en el hogar.
- **Decilsocioeconómico:** El decil 1 se refiere a los más bajos ingresos mientras que el decil 10 corresponde a los más altos ingresos. A cada individuo se le asignó el decilsocioeconómico con el que fue clasificado en la base de pruebas antropométricas trabajadas por grupo de edad por ENSANUT2012, mismo que le fue asignado a todo el hogar.

En la publicación de los resultados de la encuesta de ENSANUT2012 se hace referencia al artículo Clasificación Socioeconómica de los hogares de la ENSANUT2012 escrito por Juan Pablo Gutierrez donde se describe que para la asignación del decil socioeconómico se utilizaron indicadores comúnmente asociados con el nivel socioeconómico, tales como el nivel educativo y la propiedad de bienes en el hogar; asimismo, se consideró el ingreso de los hogares en la forma limitada que lo obtiene ENSANUT 2012, se estimaron las condiciones de pobreza multidimensional de los hogares aproximando el modelo propuesto por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, que utiliza igualmente la ENIGH 2010) con la información disponible en ENSANUT 2012.

De forma resumida, el proceso consiste en la estimación de un modelo predictor del decil del ingreso por adulto, equivalente de los hogares incluidos en la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH 2010) y, a partir de los puntos de corte estimados, imputar la clasificación del Nivel Socioeconómico a los hogares de ENSANUT 2012.

Para identificar las variables se siguieron dos criterios: por un lado, identificar la información sobre características de los hogares que pudieran reflejar diferencias en el nivel de ingreso presentes en ENSANUT 2012 y, por el otro, identificar el conjunto mínimo de estas características que permitieran una predicción del nivel de ingreso en la ENIGH de forma parsimoniosa.

La identificación en la ENIGH 2010 partió de un ejercicio previo, realizado con la ENIGH 2004, en el cual se utilizó el método de regresión por pasos (stepwise) para identificar el conjunto de indicadores que permitieran predecir el decil de ingreso de los hogares. A partir de este conjunto de variables, se identificaron en ENSANUT 2012 aquellas disponibles para estimar el modelo con el mismo conjunto de variables. La revisión de las variables se ha hecho en ejercicios con las encuestas 2006, 2008 y 2010 de la ENIGH con resultados similares.

Para implementar el proceso con el conjunto de variables, se estimó un modelo logit ordenado que permite establecer la distribución de probabilidad de que un hogar sea clasificado en un determinado decil de ingreso por adulto equivalente en la ENIGH 2010. Las variables incluidas en el modelo para identificar estas probabilidades incluyen las condiciones estructurales de la vivienda, los bienes del hogar, los patrones de consumo.

La estimación se realizó considerando tanto el diseño de la ENIGH como los pesos relativos de los hogares, es decir, los ponderadores. Los coeficientes de las variables en el modelo fueron entonces aplicados a la información de ENSANUT 2012; utilizando los puntos de corte generados por el modelo para identificar los deciles, cada hogar de ENSANUT 2012 fue asignado al decil en el que tenía mayor probabilidad. ⁽¹³⁾

Cuadro I
PESO DE LAS VARIABLES UTILIZADAS PARA PREDECIR EL
DECIL DE INGRESO. HOGARES MEXICANOS, MÉXICO,
ENSANUT 2012

<i>Variables</i>	<i>Coficiente</i>
<i>Características del jefe de familia</i>	
Años de educación	0.03171
Sexo	-0.13249
Si trabajó el mes pasado	-0.16775
<i>Características de estructura sociodemográfica</i>	
Total de integrantes del hogar	-0.39902
Total de personas que trabajan	0.35429
Número de adultos que trabajan	0.35544
Número de niños que trabajan	0.05949
Índice de dependencia económica	0.06795

Características de la vivienda	
Cocina exclusiva	0.17542
Baño	0.14192
Uso de leña o carbón como combustible	-0.43841
Número de focos	0.03061
Piso de tierra	0.01584
Bienes del hogar	
TV a color	-0.01777
TV de paga	0.24983
Microondas	0.21307
Lavadora	0.22423
Computadora	0.00762
Vehículo de motor	0.61762
Estéreo	0.03340
Internet	0.08414
Celular	0.06634
Línea telefónica	0.01165
Patrones familiares de consumo	
Paga renta por la vivienda	0.37093
Logaritmo del gasto total del hogar	1.81751

Fuente: Gutiérrez Juan Pablo, Clasificación Socioeconómica de los hogares en la ENSANUT2012. Salud Publica Mex 2013. (13)

6.3. Definición del modelo

Una vez listas las variables o regresores a utilizar se propone el modelo para el ajuste de la regresión logística a los datos para todas las enfermedades crónica degenerativas (Obesidad y Sobrepeso, Diabetes e Hipertensión). El paquete estadístico utilizado es *R* versión 3.5.3.

Las variables que se han elegido para el desarrollo de este modelo fueron basadas en artículos donde se han realizado investigaciones similares en nuestro país.

La teoría detrás de estos hallazgos, es que menos países desarrollados, con una mayor proporción de los ingresos gastados en alimentos tienen una relación positiva entre el ingreso y el peso corporal, mientras países desarrollados con un nivel de ingresos relativamente alto con respecto a sus precios de alimentos tendrían una relación negativa entre el ingreso y el peso corporal, sin embargo, en países de medianos ingresos como México, los resultados de los estudios sobre la asociación entre el ingreso y el peso son mixtos. ⁽¹²⁾

En México Fernald en 2007 en su estudio “Socio-economic status and body mass index in low-income Mexican adults” encontró una positiva asociación entre el peso corporal y la educación, la ocupación, la calidad de las condiciones de la vivienda y los bienes del hogar, lo que sugiere que los cambios rápidos en los estilos de vida y el acceso a alimentos industrializados están asociados con la riqueza y el aumento de la prevalencia del sobrepeso y la obesidad. Es por ello que se incluyó la variable decil socioeconómico ya que esta variable como lo he mencionado en la sección 6.2 fue trabajada por ENSANUT2012, donde para determinar el decil socioeconómico al que pertenece un hogar mexicano fueron tomados en cuenta las características del jefe de familia tal como su ingreso, educación y sexo, también se consideró el número de integrantes en el hogar, característica de la vivienda, es decir, si cuenta con baño, cocina, material de construcción etc, bienes del hogar y patrones de consumo.

La investigación de M.A Colchero “Heterogeneity of income and lifestyle determinants of body weight among adult women in Mexico” en 2006 sugiere que los ingresos y estilos de vida de las mujeres adultas en México son factores determinantes de su IMC. En esta investigación se realizó una regresión lineal en dos pasos que contemplaban los ingresos, el número de cuartos utilizados para dormir, gasto en ropa y calzado, actividad física, gasto en bebidas azucaradas, hábitos de fumar entre otras. Para este modelo al igual que en la investigación de Colchero fue considerado el número de habitaciones bajo el supuesto de que las familias más ricas generalmente viven en casas más grandes con más habitaciones.

En esta misma investigación se hace una diferencia entre las mujeres indígenas, rurales y urbanas ya que se asevera los ingresos y estilos de vida son distintos, es por ello que en este modelo se incluyó la variable marginalidad que es un indicador de pobreza de acuerdo al lugar de vivienda y desventaja económica, así como se incluyó la variable sociodemográfica de urbanidad y ruralidad.

Para el caso de Hipertensión y Diabetes son pocos los estudios e investigaciones realizadas (Fernald 2008 Cash component of conditional cash transfer program is associated with higher body mass index and blood pressure in adults) por lo que el modelo se hará extensivo para todas las enfermedades crónicas bajo este análisis.

6.4. Regresión logística obesidad

Se definen las siguientes variables:

X : Decil Socioeconómico donde $X \in \{1, 2, \dots, 10\}$

Y : Marginalidad donde $Y \in \{0 = \text{Alta marginalidad}, 1 = \text{Baja marginalidad}\}$

Z : Área donde $Z \in \{0 = \text{Rural}, 1 = \text{Urbano}\}$

W : Cuartos para Dormir donde $W \in N$

$$\text{Logit Obesidad} = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 Z + \beta_4 W$$

6.4.1. Estimación de los parámetros

	Coefficients			
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.45396	0.10806	-4.201	2.66E-05 ***
factor(DecilSocioeconomico)2	0.30804	0.13062	2.358	0.01836 *
factor(DecilSocioeconomico)3	0.37486	0.12971	2.89	0.003852 **
factor(DecilSocioeconomico)4	0.56939	0.13384	4.254	2.10E-05 ***
factor(DecilSocioeconomico)5	0.76278	0.13399	5.693	1.25E-08 ***
factor(DecilSocioeconomico)6	0.61565	0.13904	4.428	9.52E-06 ***
factor(DecilSocioeconomico)7	0.70489	0.14078	5.007	5.53E-07 ***
factor(DecilSocioeconomico)8	0.90474	0.14913	6.067	1.31E-09 ***
factor(DecilSocioeconomico)9	0.84537	0.15176	5.571	2.54E-08 ***
factor(DecilSocioeconomico)10	1.04298	0.16507	6.318	2.65E-10 ***
Marginalidad	0.13785	0.07135	1.932	0.053362 .
Área	0.02593	0.0744	0.348	0.72747 .
Cuartos Dormir	-0.13697	0.03955	-3.463	0.000534 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 ' ' 0.1 ' ' 1

En los resultados arrojados por el paquete estadístico R , que se observan en el cuadro anterior los parámetros en la regresión en su mayoría son altamente significativos ya que estos tienen un $p - value$ menor al 0.001, lo que nos indica que las variables aportan información al modelo:

$$\begin{aligned} \text{Logit } \textit{Obesidad} = & -0.4539 + 0.3080X_2 + 0.3748X_3 + 0.5693X_4 \\ & + 0.7627X_5 + 0.6156X_6 + 0.7048X_7 + 0.9047X_8 + 0.8453X_9 \\ & + 1.0429X_{10} + 0.1378Y + 0.0259Z - 0.1369W \end{aligned}$$

6.4.2. Interpretación del modelo

Una vez obtenidos los parámetros de los regresores del modelo obtenemos, los cocientes de momios a través de aplicar a cada uno de ellos la función exponencial:

	Parámetros	Cociente de Momios (exp(β))
(Intercept)	-0.45396	0.6351082
factor(DecilSocioeconomico)2	0.30804	1.3607601
factor(DecilSocioeconomico)3	0.37486	1.454788
factor(DecilSocioeconomico)4	0.56939	1.7671939
factor(DecilSocioeconomico)5	0.76278	2.1442375
factor(DecilSocioeconomico)6	0.61565	1.850864
factor(DecilSocioeconomico)7	0.70489	2.0236303
factor(DecilSocioeconomico)8	0.90474	2.4712953
factor(DecilSocioeconomico)9	0.84537	2.3288294
factor(DecilSocioeconomico)10	1.04298	2.8376489
Marginalidad	0.13785	1.147807
Área	0.02593	1.0262684
Cuartos Dormir	-0.13697	0.8719995

En el capítulo de Métodos estadísticos se mencionó que los momios se pueden interpretar de manera general cuando la variable es dicotómica, como el cambio promedio o esperado en el momio de respuesta cuando nuestra variable regresora toma el valor de uno respecto al valor cero.

En nuestro modelo tenemos dos variables dicotómicas:

La marginalidad que toma el valor de 0 si es alta y 1 si es baja, por lo que el momio de respuesta de tener obesidad cuando la marginalidad es baja es 1.14 veces más grande que para cuando la marginalidad es alta.

De manera similar la interpretación del regresor de área, que toma el valor uno si es urbano y cero si es rural, es que el cambio promedio de respuesta en el momio es 1.02 veces más grande de tener obesidad si se vive en el área urbana que si se vive en el área rural.

El decil socioeconómico podemos interpretarlo como que el momio de respuesta para tener obesidad es 1.36 veces más grande de tener obesidad si se pertenece al decil dos que al decil uno conforme sube el nivel socioeconómico podemos observar el momio también aumenta.

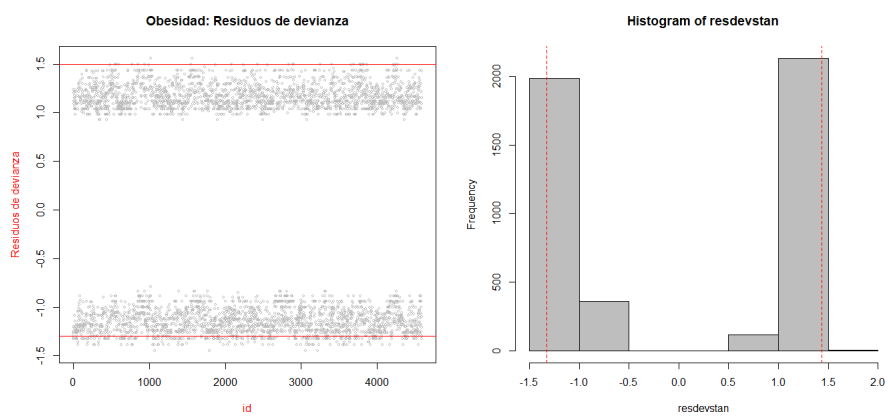
Este momio de respuesta de cada decil socioeconómico ya nos indica que a mayor nivel socioeconómico hay más posibilidades de tener obesidad.

La otra variable es la de cuartos para dormir en la cual el cambio en el momio de respuesta de tener obesidad es 0.87, como es menor que 1 indica que un aumento en nuestra variable regresora disminuye el momio de tener obesidad conforme se tengan menos cuartos para dormir, esto puede indicarnos que a lo largo del tiempo los modos de vida han cambiado y si bien en estudios anteriores esta variable era un indicador de suficiencia económica en la actualidad puede deberse a que las personas con mayores posibilidades económicas suelen tener menos familia y por ende destinar menor cantidad de cuartos para dormir.

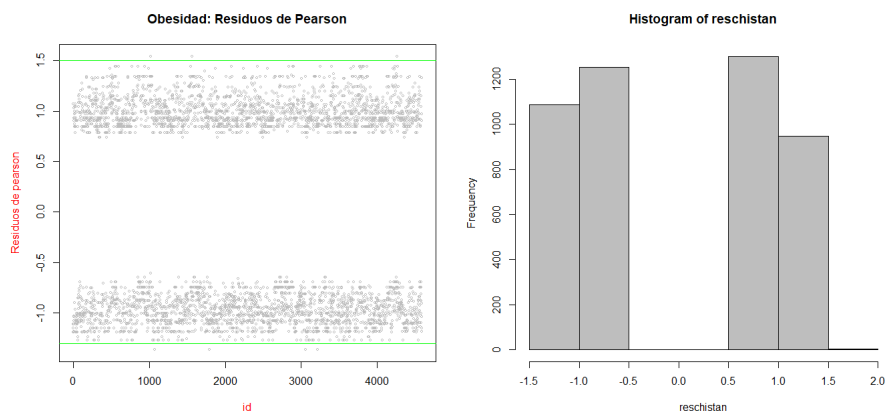
6.4.3. Bondad de ajuste del modelo

Una vez que hemos observado que las variables aportan información al modelo y que son significativas, entonces tenemos que ver si el modelo ajusta correctamente a los datos. Lo primero que debemos ver son los residuos de Pearson y de Devianza que juegan un papel preponderante para determinar el ajuste del modelo.

Veamos cómo se ven los residuos de Devianza.



Veamos los residuos de Pearson



Como vimos en las gráficas anteriores nuestros residuos de devianza y los residuos de Pearson se encuentran en un rango de diferencia contra los datos observados de 1 a 1.5 unidades por arriba o por debajo, por lo que podemos decir que el error de nuestro modelo contra los datos observados es pequeño.

Sin embargo, necesitamos realizar una prueba de bondad de ajuste que nos mida si el modelo ajusta correctamente a los datos o no, esta prueba de bondad de ajuste como hemos visto en el capítulo de métodos estadísticos se realiza con la estadística de Hosmer-Lemeshow y contrasta la siguiente prueba de hipótesis:

$$H_0 : \text{El modelo ajusta los datos vs } H_1 : \text{El modelo no ajusta los datos}$$

Con ayuda del paquete estadístico *R* hemos realizado la prueba de bondad de ajuste y los resultados son:

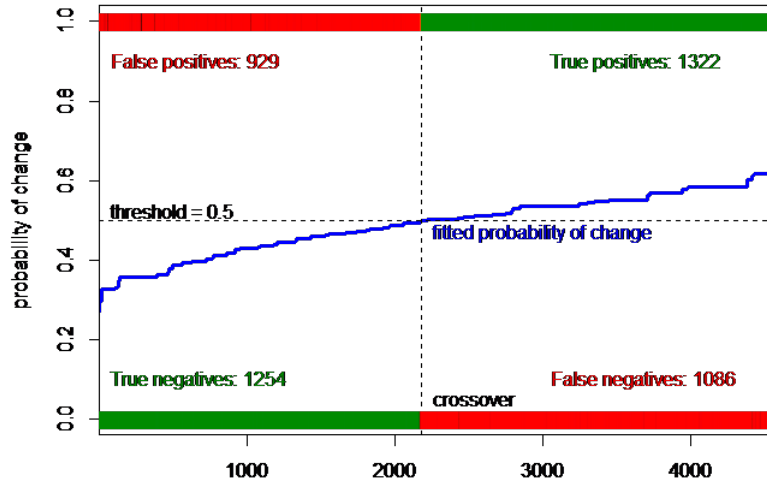
	Hosmer-Lemeshow C statistics	Hosmer-Lemeshow H statistics
X-squared	2.456041	9.3560889
p.value	0.9637493	0.3131527

Donde podemos aceptar la hipótesis nula ya que el p-value es mayor que 0.05 por lo que el modelo propuesto es un buen modelo para explicar la presencia de obesidad dadas ciertas características económicas.

Ya hemos visto que el modelo ajusta correctamente los datos, por lo que podemos decir que los factores económicos tales como el decil socioeconómico, la marginalidad social y económica, el área urbana o rural y los cuartos utilizados para dormir son factores que explican la presencia de obesidad en individuos con cierto rasgos económicos.

Sin embargo, es importante evaluar la capacidad predictiva del modelo, es decir, que éste pueda predecir si una persona tendrá obesidad dadas sus características económicas, para ello utilizaremos la estructura de la tabla de clasificación de la curva de ROC mostrada en el capítulo de Métodos Estadísticos para un punto de corte de 0.5.

Model success



Sensitivity: 0.5873 ; Specificity: 0.5359

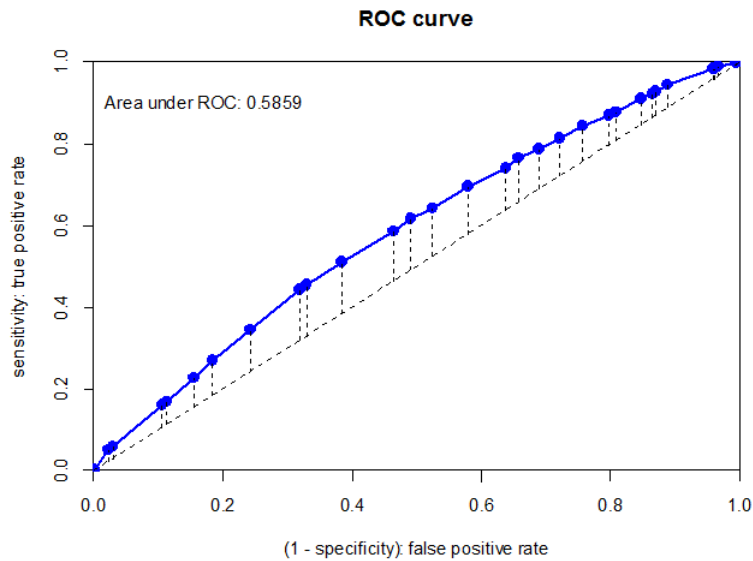
Observados	Clasificación del modelo		TOTAL
	Presenta Obesidad	No Presenta Obesidad	
Presenta Obesidad	1322	929	2251
No Presenta Obesidad	1086	1254	2340
TOTAL	2408	2183	4591

$$Sensitividad = \frac{Verdaderos\ Positivos}{Total\ de\ Positivos} = \frac{1322}{2251} = 0.5873$$

$$Especificidad = \frac{Verdaderos\ Negativos}{Total\ de\ Negativos} = \frac{1254}{2340} = 0.5359$$

Como se observó en la gráfica anterior, con un punto de corte de 0.5 se obtiene una sensibilidad de 58 % y un 53 % de especificidad y una tasa global de predicción del modelo de 56. %

Veamos el comportamiento de la curva de ROC para poder localizar el punto de corte óptimo



Punto de corte	Sensitividad	Especificidad	Tasa Global
0.38	92.09 %	13.59 %	52.08 %
0.39	91.07 %	15.30 %	52.45 %
0.4	87.65 %	19.15 %	52.73 %
0.41	86.94 %	20.30 %	52.97 %
0.42	84.41 %	24.32 %	53.78 %
0.43	81.34 %	27.86 %	54.08 %
0.44	78.90 %	30.98 %	54.48 %
0.45	76.41 %	34.23 %	54.91 %
0.46	74.14 %	36.24 %	54.82 %
0.47	69.61 %	42.09 %	55.59 %
0.48	64.33 %	47.56 %	55.78 %
0.49	61.66 %	50.94 %	56.20 %
0.5	58.73 %	53.59 %	56.11 %
0.51	51.13 %	61.50 %	56.41 %
0.52	45.58 %	67.05 %	56.52 %

En el cuadro anterior podemos ver que el punto de corte que favorece tanto a la sensibilidad como a la especificidad es 0.49, dándole prioridad a la sencitividad, debido a que la primordial del modelo es la predicción de los verdaderos positivos.

Observados	Clasificación del modelo		TOTAL
	Presenta Obesidad	No Presenta Obesidad	
Presenta Obesidad	1388	863	2251
No Presenta Obesidad	1148	1192	2340
TOTAL	2536	2055	4591

$$Sensitividad = \frac{Verdaderos\ Positivos}{Total\ de\ Positivos} = \frac{1388}{2251} = 0.616$$

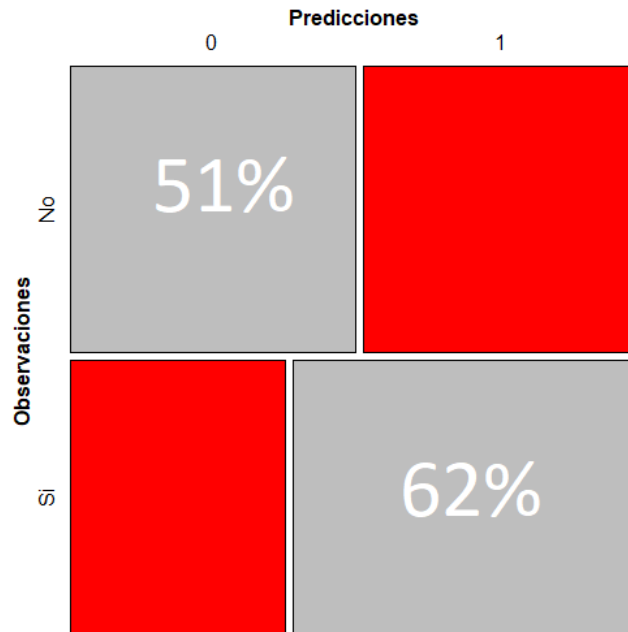
$$Especificidad = \frac{Verdaderos\ Negativos}{Total\ de\ Negativos} = \frac{1192}{2340} = 0.509$$

Podemos decir que el modelo predice el 62 % de las respuestas positivas observadas y el 51 % de las respuestas negativas observadas.

Por lo que la tasa total de buena clasificación es

$$TBC = \frac{1388 + 1192}{4591} = 0.562$$

En la gráfica de la Curva de ROC también se puede ver que el área bajo la curva es de casi 60 % por lo que nuestro modelo solo puede ser clasificado como bueno de acuerdo a la escala mostrada en el capítulo de Métodos Estadísticos.



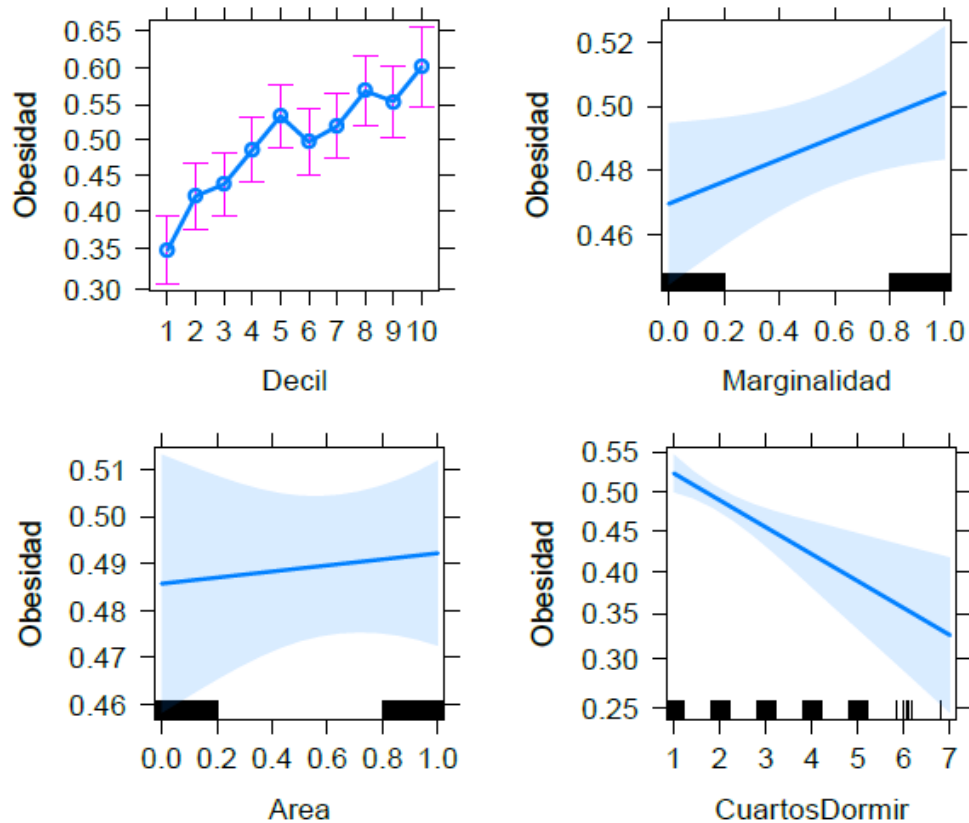
Es decir el modelo global predice correctamente 56% de las respuestas observadas. Si bien la tasa de predicción del modelo no es muy alta, predice en casi un 62% a los verdaderos positivos.

Por lo que podemos decir que los factores económicos que definen a las familias son una causa de la presencia de obesidad en los hogares mexicanos y con base en esos factores podemos predecir el 56% de los casos que bajo estos factores económicos tendrán un caso de obesidad.

Una vez comprobado que existe una relación causal de la obesidad versus factores económicos, queremos verificar que a mayor decil socioeconómico, mayor probabilidad de presentar Obesidad, para ello se recuperó las probabilidades en R bajo la mecánica explicada en el capítulo Métodos Estadísticos.

$$\Pi(X) = P(Y = \acute{E}xito) | X_1, X_2, \dots, X_p) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}$$

GRÁFICAS MARGINALES



En la gráfica anterior se observa las distribuciones marginales para cada regresor, en donde se agrupó a los individuos fijando las características económicas de acuerdo a los valores tomado por cada regresor.

En la marginal respecto al Decil socioeconómico podemos ver de manera general que las probabilidades van de manera ascendentes según aumenta el decil socioeconómico aumenta la probabilidad de tener obesidad.

Con respecto al regresor Marginalidad vemos que aquellos marcados con 1 que representan a los individuos con baja marginalidad, es decir, más favorecidos económicamente, tienen mayor probabilidad de tener obesidad que los marcados con cero que representan a la población más marginada.

En la variable Área observamos que ligeramente las probabilidades para los individuos marcados con 1, aquellos que viven en áreas urbanas, son más grandes que aquellos que viven en áreas rurales marcados con cero.

Finalmente, para el regresor Cuartos para Dormir observamos un comportamiento inesperado, se observa que a mayor cantidad de habitaciones utilizadas para dormir la probabilidad de tener obesidad disminuye, lo cual se contrapone con los estudios realizados con anterioridad, sin embargo, esto puede estar pasando debido a los cambios culturales, ya que las parejas actuales con mejores posiciones económicas tiene menos cantidad de hijos o ausencia de los mismos, por ende sus casas tienen menos cuartos destinados para dormir.

6.5. Regresión logística diabetes

Se definen las siguientes variables:

X : Decil Socieconómico donde $X \in \{1, 2, \dots, 10\}$

Y : Marginalidad donde $Y \in \{0 = \text{Alta marginalidad}, 1 = \text{Baja marginalidad}\}$

Z : Área donde $Z \in \{0 = \text{Rural}, 1 = \text{Urbano}\}$

W : Cuartos para Dormir donde $W \in N$

$$\text{Logit Diabetes} = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 Z + \beta_4 W$$

6.5.1. Estimación de los parámetros

Usando la misma metodología del modelo de obesidad, estimamos los parámetros para el modelo de diabetes con el paquete estadístico R

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.8826	0.585	-4.928	8.33E-07	***
factor(Decilsocioeconómico)2	0.601	0.6732	0.893	0.372	
factor(Decilsocioeconómico)3	0.2763	0.7032	0.393	0.6944	
factor(Decilsocioeconómico)4	1.1435	0.6274	1.823	0.0684	.
factor(Decilsocioeconómico)5	0.4618	0.677	0.682	0.4952	
factor(Decilsocioeconómico)6	-0.5831	0.9152	-0.637	0.524	
factor(Decilsocioeconómico)7	1.0013	0.6992	1.432	0.1521	
factor(Decilsocioeconómico)8	0.2985	0.7786	0.383	0.7015	
factor(Decilsocioeconómico)9	0.459	0.7979	0.575	0.5651	
factor(Decilsocioeconómico)10	0.3683	0.8803	0.418	0.6757	
Marginalidad	-0.2766	0.3415	-0.81	0.4179	
Área	0.6605	0.3827	1.726	0.0844	.
Cuartos Dormir	-0.2601	0.2011	-1.294	0.1957	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En los resultados arrojados por el paquete estadístico R que se observan en el cuadro anterior los parámetros en la regresión no son significativos, lo que nos dice que las variables independientes o regresores no aportan información al modelo, es decir los factores socioeconómicos tales como el decil socioeconómico, la marginalidad, la región sociodemográfica y el número de habitaciones destinados para dormir que definen a un individuo no son suficientes para relacionar la causa de la presencia de diabetes.

Esto probablemente se deba a que existan factores más fuertes que expliquen la causa de la diabetes tales como la genética dejando en un menor grado los factores económicos que lleven a tener una mala alimentación que desencadene el desarrollo de diabetes.

Por lo que este modelo no es funcional para concluir que existe relación de los factores económicos y la presencia de diabetes en los hogares mexicanos.

6.6. Regresión logística hipertensión

Se definen las siguientes variables:

X : Decil Socioeconómico donde $X \in \{1, 2, \dots, 10\}$

Y : Marginalidad donde $Y \in \{0 = \text{Alta marginalidad}, 1 = \text{Baja marginalidad}\}$

Z : Área donde $Z \in \{0 = \text{Rural}, 1 = \text{Urbano}\}$

W : Cuartos para Dormir donde $W \in N$

$$\text{Logit Hipertensión} = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 Z + \beta_4 W$$

Siguiendo la misma metodología para los modelos anteriores estimamos los parámetros del modelo de hipertensión.

6.6.1. Estimación de los parámetros

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.444856	0.355624	-9.687	<2e-16 ***
factor(Decilsocioeconómico)2	-0.006058	0.412151	-0.015	0.9883
factor(Decilsocioeconómico)3	-0.355784	0.447962	-0.794	0.4271
factor(Decilsocioeconómico)4	-0.995159	0.546521	-1.821	0.0686
factor(Decilsocioeconómico)5	0.204861	0.402108	0.509	0.6104
factor(Decilsocioeconómico)6	-0.407765	0.485528	-0.84	0.401
factor(Decilsocioeconómico)7	-0.201685	0.451628	-0.447	0.6552
factor(Decilsocioeconómico)8	-1.255452	0.618888	-2.029	0.0425 *
factor(Decilsocioeconómico)9	-0.813579	0.556178	-1.463	0.1435
factor(Decilsocioeconómico)10	-0.336843	0.533446	-0.631	0.5277
Marginalidad	0.103846	0.249271	0.417	0.677
Área	0.361371	0.267898	1.349	0.1774
Cuartos Dormir	-0.067102	0.140277	-0.478	0.6324

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De manera análoga con el modelo de diabetes los parámetros arrojados por el paquete estadístico R no son significativos, lo que nos dice que las variables independientes o regresores no aportan información al modelo, es decir los factores socioeconómicos tales como el decil socioeconómico, la marginalidad, la región sociodemográfica y el número de habitaciones destinados para dormir que definen a un individuo no son suficientes para relacionar la causa de la presencia de hipertensión.

Por lo que este modelo no es funcional para concluir que existe relación de los factores económicos y la presencia de hipertensión en los hogares mexicanos.

7. Conclusiones

De acuerdo a los resultados obtenidos en este estudio y bajo las premisas de estudios realizados con anterioridad, se confirma que la situación económica de los hogares mexicanos es causa de obesidad entre sus integrantes, sin embargo en este estudio se observa que la relación no es inversamente proporcional entre obesidad y pobreza pues existe una tendencia de que a mayor decil socioeconómico mayor probabilidad de presencia de obesidad, ya que el nivel socioeconómico de los hogares permite definir estilos de vida, es decir, si se pertenece a deciles socioeconómicos más altos se tienen niveles de ingresos superiores por ende acceso a productos alimenticios procesados con mayor satisfacción en el sabor y con menos valor nutricional a un mayor costo dejándolos fuera de la canasta básica, si bien los hogares pertenecientes a deciles más bajos y con ingresos menos favorecedores también llevan una dieta deficiente, se inclinan por productos con mayor aportación calórica a un costo menor los cuales se orientan más a las harinas y no a los azúcares los cuales desencadenan un aumento de peso en una menor velocidad que alimentos altos en azúcar.

También se observó que las personas que viven en áreas urbanas, se posicionan en deciles socioeconómicos más altos y con ello una mayor exposición al riesgo de presencia de obesidad en comparación de las áreas rurales.

De tal manera que si bien en los hogares que tienen características económicas más favorecedoras se podría pensar que tiene más acceso a la información y educación alimenticia y de salud observamos que el poder adquisitivo de alimentos que satisfacen el sabor pesa más en la cultura alimentaria mexicana pese a los esfuerzos de programas de prevención de obesidad de la seguridad social.

En lo que respecta a las enfermedades como diabetes e hipertensión el modelo no permitió ver que existe la relación entre los factores socioeconómicos con el riesgo de presencia de estas enfermedades, si bien están ligadas a la forma de alimentación y está a los niveles de ingreso, se puede pensar que existen factores más fuertes que desencadenen las causas de mencionadas enfermedades crónicas tales como lo son la genética, y la presencia de enfermedades preexistentes como la obesidad.

Finalmente el valor de conocer y medir la exposición al riesgo de presentar obesidad dadas ciertas características económicas, podría ser el principio de identificar grupos y regiones del país que requieran un tratamiento y enfoque de educación alimentaria y financiera para la prevención de obesidad en niños, y dar pie al origen de seguros o programas que cubren distintos problemas de salud que pueden estar ligados a la obesidad.

8. Bibliografía

- (1).Grupo de Expertos Independientes (GEI). (2016).Global Nutrition Report from promise to impact:ending malnutrition by 2030. Estados Unidos.
- (2).Juan Angel Rivera Don Marco, M. H. (2013). Obesidad en México:Recomendaciones para una política de Estado. Mexico: Direccion General de Publicaciones y Fomento Editorial.
- (3).Rivera J. A, G. C. (2008). Pobreza, Nutrición y Salud, en Informe México sobre determinantes sociales de la salud. México: Fondo de Cultura Económica.
- (4).Salud, O. M. (2013). Información General sobre la Hipertensión en el mundo. Ginebra, Suiza: WHO Document Production Services.
- (5).The World Health Organization Technical Report Series. (2013). Diet, nutrition and the prevention of chronic diseases: Report of a Joint WHO/FAO Expert Consultation. Geneva, Switzerland: World Health.
- (6). Óscar Velázquez-Monroy, Martín Rosas Peralta, Agustín Lara Esqueda, Gustavo Pastelín Hernández, Grupo ENSA 2000,* Claudia Sánchez Castillo, Fause Attie y Roberto Tapia “Prevalencia e interrelación de enfermedades crónicas no transmisibles y factores de riesgo cardiovascular en México: Resultados finales de la Encuesta Nacional de Salud (ENSA) 2000”
- (7). Diabetes. Federación Mexicana de Diabetes
(<http://fmdiabetes.org/estadisticas-diabetes-inegi-2013/>)
- (8).Pública, I. N. (2012). Encuesta Nacional y Nutrición 2012 Evidencia para la política pública en salud. Mexico.
- (9). Drewnoswski A, Specter SE, ”Proverty and Obesity:The Role of Energy, Density an Energy Cost”. The American Journal of Clinical Nutrition, Volumen 79, Número 1, 2004, Páginas 6-16.
- (10). Torres Felipe, Jose Gasca, ”Ingreso y Alimentacion de la Poblacion en México siglo XX”, Texto Breves de economía Instituto de Investigaciones Economicas UNAM. México 2001.

- (11). NORMA Oficial Mexicana NOM-043-SSA2-2005, Servicios básicos de salud. Promoción y educación para la salud en materia alimentaria. Criterios para brindar orientación.
- (12).M. A. Colchero, S. G. (2006). Heterogeneity of income and lifestyle determinants of body weight among adult women in Mexico. *Social Science Medicine*, 120 a 128.
- (13). Gutiérrez Juan Pablo, "Clasificación Socioeconómica de los hogares en la ENSANUT2012". *Salud Pública Mex* 2013;55 supl 2:S341-S346.
- (14). Fernald, L. C. (2007). Socio-economic status and body mass index in low-income Mexican adults. *Social Science Medicine*, 64(10), 2030e2042.