

**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

**FACULTAD DE ECONOMÍA**

**“Factores de Riesgo que determinan la obesidad en México, a través de un Análisis de Regresión Logística Binaria basados en la información de la ENSANUT 2012”**

**TESIS**

**Que para obtener el Título de Licenciado en Actuaría**

**Presenta**

**P.L.A Leticia Betzabeth Díaz Fuentes**

**Asesora: M en MA Verónica Ángeles Morales**

**Noviembre 2016**

## ÍNDICE

Prólogo	I
Introducción	1
<b>Capítulo I: Marco Teórico y Referencial</b>	
1.1 Marco Teórico	6
1.1.1 Obesidad en la antigüedad	6
1.1.1.1 Prehistoria	6
1.1.1.2 Egipto	7
1.1.1.3 Grecia	8
1.1.1.4 Roma: Galeno	8
1.1.1.5 Cultura cristiana: ideas sobre la glotonería	9
1.2 Obesidad Moderna	9
1.2.1 Siglo XV- Obesidad en la Cultura Cristiana	9
1.2.2 Siglos XVI-XVII	10
1.2.3 Siglo XVIII-primera mitad	10
1.2.4 Siglo XIX	11
1.2.5 Siglo XX	11
1.3 Marco Referencial	12
1.3.1 Artículos académicos que abordan problemas de salud a través de un análisis multivariado.	12
1.3.2 Aplicación de análisis multivariado para la diferenciación de individuos sanos según su contenido sérico de minerales.	12
1.3.3 Análisis Multivariado Discriminante en trabajadores con úlcera péptica gastroduodenal y exposición a la nocturnidad.	14

1.3.4 Factores Asociados a Accidentes, Enfermedades y Ausentismo Laboral: Análisis de Cohorte de Trabajadores Formales en Chile.	15
---	----

## **Capítulo II: Obesidad**

2.1 Definiciones: Sobrepeso y Obesidad	19
2.2 Tipos de Obesidad	20
2.2.1 Tipos de obesidad según el riesgo	20
2.2.2 Tipos de obesidad según la distribución de grasa	21
2.2.3 Tipos de obesidad según la causa	22
2.2.4 Otros tipos de obesidad	23
2.3 Causas del sobrepeso y la obesidad	24
2.4 Consecuencias comunes del sobrepeso y la obesidad para la salud	
Física	26
2.4.1 Físicas	26
2.4.2 Económicas y Sociales: Afrontar una doble carga de morbilidad	27
2.5 Reducción del sobrepeso y la obesidad	28
2.6 Estadísticas Mundiales de la Obesidad	31
2.7 Obesidad en México	33
2.7.1 Sobrepeso y Obesidad	33
2.7.2 Consecuencias	34
2.7.3 Programa de Prevención y Control de la Obesidad.	34
2.7.4 Normatividad	35
2.7.5 Acciones	35

## **Capítulo III: Análisis de Datos categóricos**

3.1 Distribuciones e inferencias para datos categóricos	38
3.1.1 Escalas de Medición	38
3.1.2 Distribuciones de Probabilidad	40
3.1.3 Inferencias Estadísticas	44

3.2 Tablas de Contingencia e Inferencias	47
3.2.1 Definición Tablas de Contingencia	47
3.2.2 Pruebas de Independencia	49
3.2.3 Riesgo Relativo y Cociente de Momios	50
3.3 Modelos Lineales Generalizados	53
3.3.1 Varianza no constante	56
3.3.2 Normalidad	58
3.3.3 Función vínculo/ligadura	60
3.3.4 Especificación del modelo	60
3.3.5 Especificación de la estimación	64
3.3.6 Construcción y evaluación de un GLM	65
3.4 Regresión Logística	69
3.5 Interpretación de los modelos de regresión logística	71

### **Capítulo III: Aplicación del modelo**

4.1 Encuesta Nacional de Salud y Nutrición 2012	75
4.1.1 Antecedentes ENSANUT 2012	75
4.1.2 Objetivos	77
4.1.3 Metodología	77
4.1.4 Alcances de la ENSANUT	82
4.2 Análisis de las variables objeto de estudio	84

4.3 Análisis de Datos	86
4.3.1 Resultados	87
4.3.1.1 Inferencias tabla de frecuencias	88
4.3.1.2 Cociente de momios de obesidad según algunos factores socioeconómicos.	90
4.3.1.3 Regresión Logística	94
4.3.1.4 Tabla de clasificación	98
Conclusiones	101
Bibliografía	103

## ÍNDICE DE GRÁFICOS TABLAS Y CUADROS

Gráfico 1. La definición de la Obesidad según la OMS.....	20
Gráfico 2. Tipos de Obesidad Abdominal.....	22
Gráfico 3. Consecuencias sobrepeso y obesidad.....	27
Gráfico 4. Sobrepeso y Obesidad Mundial (Hombres).....	32
Gráfico 5. Sobrepeso y Obesidad Mundial (Mujeres).....	33
Tabla 1. Estadísticas Obesidad en México (por grupos).....	34
Tabla 2. Test de Independencia en Tablas de Contingencia.....	49
Gráfico 6. Varianza no Constante (Supuesto I).....	56
Gráfico 7. Varianza no Constante (Supuesto II).....	57
Gráfico 8. Varianza no Constante (Supuesto III).....	57
Gráfico 9. Varianza no Constante (Supuesto IV).....	58
Cuadro I. Funciones de ligadura/vínculo más utilizadas.....	62
Cuadro II. Funciones de vínculo para la distribución de los errores.....	63
Cuadro III. Combinaciones de variables de respuesta y explicativas.....	64
Cuadro IV. Criterios de estratificación de la ENSANUT 2012.....	80
Cuadro V. Parámetros usados para la estimación de tamaño de muestra.....	81
Cuadro VI. Variables Objeto de Estudio.....	84
Gráfico 10. Prevalencia de Obesidad en México por sexo y Grupos de Edad...87	
Tabla 3. Tablas de Frecuencia Obesidad.....	89
Gráfico 11. Delta Chi Cuadrada.....	96
Tabla 4. Tablas de Clasificación.....	98

## ***PRÓLOGO***

Al día de hoy la obesidad ha pasado de ser una condición de salud a un problema médico complejo, que según la organización mundial de la salud se ha convertido en epidemia mundial.

Es importante resaltar que esta condición genera problemas no sólo físicos sino también psicológicos desencadenando en una serie de enfermedades, que llevan al desmejoramiento de las condiciones de vida de quienes lo padecen y en algunos casos a la muerte.

En particular para el caso de México tanto la obesidad como el sobrepeso son el principal problema de Salud Pública, pues nuestro país es el primer lugar mundial en niños con obesidad y sobrepeso, y segundo en adultos, derivado de esto México gasta alrededor del 7% del presupuesto destinado a salud para atender la obesidad. Como se puede ver para el caso de nuestro país, el incremento de la incidencia de las patologías asociadas a la obesidad, constituyen también un serio problema económico ya que genera un claro incremento de la demanda de atención sanitaria, poniendo en una situación aún más complicada a los ya saturados sistemas sanitarios.

Por tal motivo, el objetivo de esta investigación es conocer las principales variables que se relacionan con dicho padecimiento, a través de un Análisis de Regresión Logística (se calculará la importancia de cada variable y el riesgo relativo para cada una de ellas). Con el objetivo de promover medidas y/o campañas de salud que sean dirigidas a aquellos grupos en los cuales la probabilidad de padecer dicho padecimiento sea mayor.

## ***INTRODUCCIÓN***

Al día de hoy y como consecuencia del creciente impacto de la globalización, muchos procesos sociales, económicos, políticos y culturales se han modificado. Esto ha generado cambios importantes los cuales han transformado los hábitos y costumbres de la población, todo esto ha sido un determinante para el cambio en los estilos de vida.

Como consecuencia de lo anterior la sociedad ha optado por un decremento en la calidad de los alimentos que se consume y un incremento en la cantidad de alimentos consumidos, adicional también, a la falta de actividad física (como consecuencia de las múltiples opciones que existen en los medios de transporte y las nuevas formas de trabajo y entretenimiento), lo anterior ha contribuido a un aumento considerable de las enfermedades ligadas al consumo excesivo de alimentos (conocido como obesidad).

La obesidad, por su parte, al día de hoy constituye un problema de salud pública, la cual ha sido calificada como la “epidemia del siglo”, esto ha dado como resultado que una gran cantidad de recursos tanto humanos como económicos sean empleados para su prevención y control.

De acuerdo con datos de la OMS, hace diez años había en el mundo un aproximado de 330 millones de adultos obesos; en 2005 alcanzó los 400 millones de personas, y en 2008 la cifra asciende a 1,400 millones con sobrepeso y más de 500 millones con obesidad.

Para el caso de México, dicho padecimiento es al día de hoy un problema de Salud Pública el cual ha generado un incremento importante en el gasto público asignado a combatir problemas de salud a consecuencia de dicho padecimiento.

Por todo lo anterior, a lo largo de este trabajo de investigación que se ha realizado, se podrá estudiar más acerca de la Obesidad en México, a través de diversas variables como lo son: Ingreso, Edad, Región (rural o urbana), Género, Escolaridad, Afiliación, Región del país e Ingresos; después de esto se procederá con un análisis de regresión logística para medir los impactos de cada variable en nuestra variable de respuesta (medida con 1 y 0 en caso de existir o no obesidad).

El modelo, el cual será un análisis de regresión logística binaria, tiene como objetivo principal analizar el problema de la obesidad a través de las variables descritas para poder conocer cuáles de ellas son las que afectan en mayor proporción a la aparición del padecimiento, todo esto para poder identificar y mitigar el problema a través de programas sociales que se enfoquen en las variables con mayor dependencia.

Para tal efecto, la hipótesis propuesta es: “Es posible realizar un modelo logístico binario que permita predecir la obesidad en México y conocer las variables principales que afectan la aparición de este padecimiento”.

Es importante mencionar que la metodología utilizada es un análisis multivariado con el cual se obtendrá la estimación de los parámetros, la relación de las variables, las inferencias a través de las razones de momios (que es la posibilidad de que una condición de salud o enfermedad se encuentre en determinados grupos) y las tablas de frecuencias. Para tal efecto se utilizará el paquete estadístico Minitab.

Cabe recordar que la regresión logística, intenta modelar la relación entre la probabilidad (proporción) de una respuesta binaria y las variables explicativas, los valores numéricos de 0 y 1 se asignan a las dos categorías de una variable binaria, el 0 representa una respuesta negativa y el 1 representa una afirmación, la media de esta variable será la proporción de respuestas positivas (presencia de obesidad).

De igual manera, cabe mencionar que esta técnica modela cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores (variables utilizadas en el modelo) y el valor o nivel de los mismos, ella

precisa el efecto de un grupo de variables, consideradas potencialmente influyentes, sobre la ocurrencia de un determinado proceso.

Como punto medular es importante hacer mención de las preguntas de investigación “¿Por qué existen problemas de Obesidad en México?, ¿Se puede estimar la presencia o no del padecimiento dadas las variables objeto de estudio?, ¿Se pueden generar programas sociales que se enfoquen únicamente en las variables que más influyen a la presencia de la obesidad?, todas estos cuestionamientos se pretenderán responder durante todo el contenido del presente trabajo.

El trabajo de Investigación, consta de cuatro capítulos, los cuales se resumen a continuación.

A lo largo del primer capítulo se conocerá la historia de la obesidad en el mundo y cómo esta pasó de ser un padecimiento a un problema de salud mundial, catalogado por muchos como una epidemia. Adicional, se revisará algunos artículos que utilizan modelos multivariados y de regresión lineal en los cuales se abordan problemas de salud; esto con el objetivo de tener referencias sobre modelos aplicados a problemas de salud.

En el segundo capítulo se revisarán los conceptos básicos de Obesidad, desde la definición misma, hasta tipos obesidad, consecuencias de la misma y medidas de prevención. Por último se revisarán algunas estadísticas locales y globales que nos permitirán evaluar de manera general el problema de salud que se presenta hoy en día.

Como tercer capítulo, se hondará sobre el análisis de regresión y la diferencia que este guarda con el análisis de regresión logística, se revisará para poder analizar el porqué del uso de este tipo de análisis. En este mismo capítulo se revisará el Riesgo Relativo y el cociente de momios, se conocerá de manera general su definición y cómo el cálculo de éstos ayuda a nuestro tema de investigación.

Por otro lado, se revisarán también las tablas de frecuencia y contingencia con la finalidad de poder presentar una introducción y lectura de las mismas, ya que éstas serán utilizadas de igual manera en nuestro modelo. Finalmente, se revisarán a detalle los modelos lineales generalizados, y la aplicación de los mismos, para con esto poder

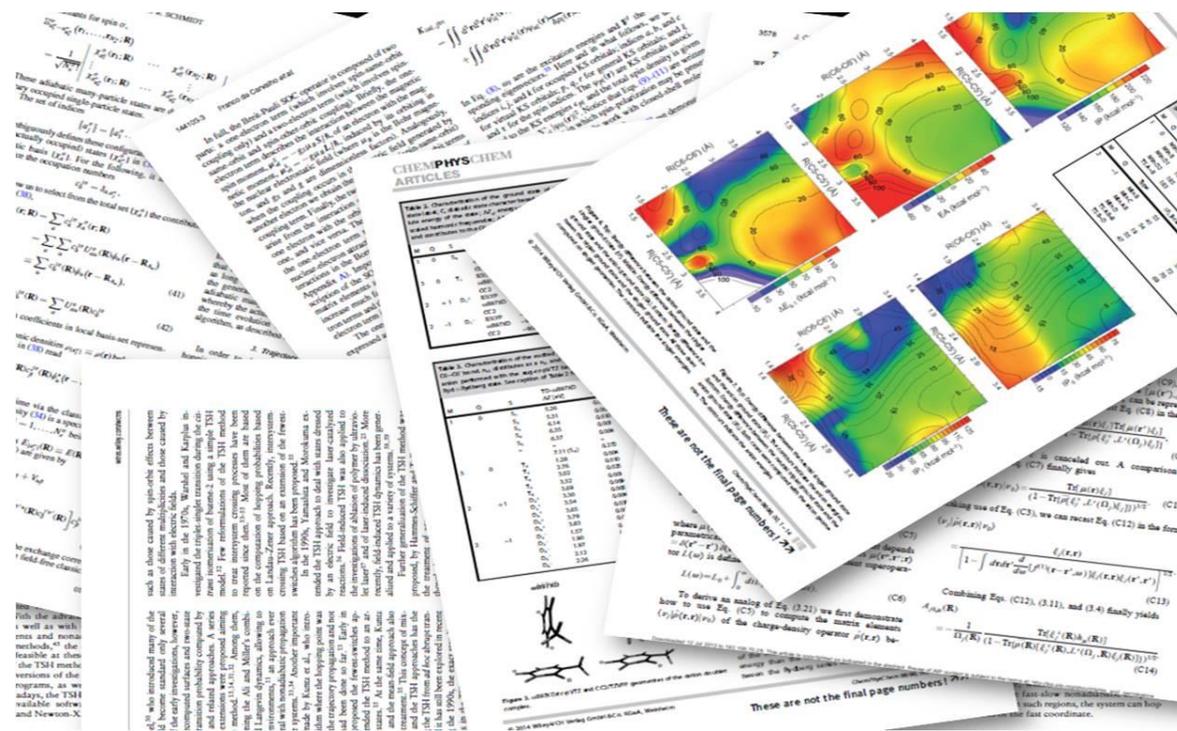
aterrizarlos dentro del análisis de regresión logística binaria (que será nuestro objeto de estudio en el capítulo 3)

Por último en el cuarto y último capítulo se revisará la metodología de muestreo utilizada en la Encuesta Nacional de Salud y Nutrición 2012 para conocer más acerca de nuestras variables objeto de estudio, adicional se analizará la forma en que se codificaron y el porqué de la inclusión de las mismas en nuestro modelo. Como segundo paso se analizarán los resultados obtenidos al realizar los cálculos del cociente de momios y el riesgo relativo, con esto y las tablas de contingencia se podrán realizar inferencias por grupos y de acuerdo a nuestras variables.

Como tercer punto se revisarán los resultados del modelo de regresión logística y se procederá con el análisis de los mismos.

Finalmente, se presentarán las conclusiones finales sobre los resultados obtenidos.

# CAPÍTULO I. MARCO TEÓRICO Y REFERENCIAL



Durante este capítulo se hará una revisión de la historia de la obesidad con el objetivo de poder analizar los aspectos sociales, políticos y culturales de este padecimiento a través del tiempo. Además se conocerán las distintas técnicas empleadas para la disminución de la obesidad.

Para tal efecto, el marco teórico se dividirá en 4 rubros principales: La obesidad en la antigüedad, Edad Media, Edad Moderna y Edad Contemporánea.

Cabe destacar que lo citado aquí tiene base en el artículo académico de Mario Fox (Catedrático de Medicina y Profesor Emérito de la Universidad Autónoma de Barcelona, España) "Historia de la Obesidad".

Como segundo punto dentro de este capítulo se analizarán 3 artículos relacionados con la salud, los cuales tienen como base el uso de técnicas estadísticas multivariadas; esto con el objetivo de conocer las metodologías y su aplicación en las ciencias de la salud. Esto nos servirá como una referencia para el modelo que se revisará dentro de este trabajo de investigación.

## **1. Marco Teórico**

### **1.1 Obesidad en la Antigüedad**

#### **1.1.1 Prehistoria**

La prehistoria se divide en 2 grandes periodos, la edad de piedra 2.5m-3000 a.C (Paleolítico, Mesolítico y Neolítico) y la edad de los metales 5.000-800 siglos d.C. (Cobre, Bronce y Hierro)

Como primera instancia, cabe mencionar que durante el paleolítico, la economía se basaba en la caza y la recolección, por lo tanto el ser humano debía de resistir a los frecuentes periodos de carencia de alimentos.

Dicho hecho forzó al cambio en el genoma humano de aquellos <<genes ahorradores>> que favorecían el depósito de energía y permitía que estos individuos tuvieran una mayor supervivencia y alcanzaran la edad de reproducción.

A partir del Mesolítico, se hizo presente la economía, la caza, la recolección y la pesca y más adelante la aparición de la agricultura y la ganadería. Esta transformación económica, conocida como la Revolución neolítica, puede ser considerada como el suceso más importante en la historia humana y el lejano antecedente de las sociedades modernas que favorecen a la obesidad, ya que permitió el crecimiento de la población y la evolución hacia sociedades complejas y civilizadas.

La única constatación que se tiene de la existencia de la obesidad en tiempos prehistóricos proviene de estatuas de la edad de piedra representando la figura femenina con exceso de volumen en sus formas. La más conocida es la Venus de Willendorf, una pequeña estatua de la edad de piedra que tiene una antigüedad aproximada de 25.000 años y que está expuesta en el museo de Historia.

### 1.1.2 Egipto

En el antiguo Egipto se tienen registros de restos humanos que presentaban obesidad, esto es a consecuencia de que la dieta en dicha época era rica principalmente para las personas de rangos elevados.

Adicional, se creía que la corpulencia u obesidad era un signo de ahorro de energía y de poder, y por lo tanto, un estado deseable.

De hecho en el antiguo testamento el Faraón agradecido promete a José <<toda la grasa de la Tierra>> o se señala que <<el virtuoso florecerá como el árbol de la palma...ellos traerán abundante fruto en la edad avanza: ellos serán gordos y florecientes>><sup>1</sup>

Sin embargo, también se tienen registros de que se tenía el conocimiento del beneficio de llevar dietas equilibradas. De hecho en el Imperio medio egipcio (siglos XXI-XVII a.C), en las enseñanzas del Kagemni, puede leerse por primera vez en la historia una asociación de la glotonería con la obesidad y una condena y estigmatización del comer con exceso. Así, en este texto señala: “La glotonería es grosera y censurable”.

---

<sup>1</sup> (Salmos XCII,13)

### 1.1.3 Grecia

Tanto los griegos como los romanos, son considerados los iniciadores de la dietoterapia<sup>2</sup> de hecho Hipócrates, la gran figura de la medicina griega (siglo V a.C), realiza un minucioso estudio de la enfermedad a través de la observación de los síntomas que presenta el paciente. En relación a la obesidad, Hipócrates señala que “la muerte súbita es más frecuente en los obesos que en los delgados”, y que la obesidad es una causa de infertilidad en las mujeres y que la frecuencia de menstruación está reducida.

Por otro lado el gran filósofo Platón (siglo V-IV a.C) proclama una certera observación sobre la alimentación y la obesidad al señalar que la dieta equilibrada es la que contiene todos los nutrientes en cantidades moderadas y que la obesidad se asocia con la disminución en la esperanza de vida.

### 1.1.4 Roma: Galeno

Galeno (siglo II a.C), que nació en Grecia pero residió en Roma, es una de las personalidades más importantes en la historia de la medicina. Escribió más de 125 libros médicos y en relación a la obesidad elaboro algunos conceptos juiciosos de valor clínico, descritos a continuación:

Se identificaron 2 tipos de obesidad: moderada e inmoderada. La primera la considera como natural y la segunda como mórbida.

En su libro “De Sanitate Tuenda”, Galeno expone que: “El arte higiénico promete mantener en buena salud a aquellos que lo obedecen pero no así a aquellos que no lo hacen”. Galeno veía, por tanto, la obesidad en relación a un estilo de vida inadecuado. Algunos de sus consejos van dirigidos a intentar corregir esta desviación: “Yo he conseguido adelgazar a un paciente obeso en un tiempo breve aconsejándole que corriera velozmente. Tras el ejercicio, “... le di abundante comida poco nutritiva a fin de saciarle y de que aquélla se distribuyera poco en el cuerpo entero”.

---

<sup>2</sup> La Dietoterapia, según González Cruz, Francisco, es la terapia que se basa en la prevención y curación de las enfermedades o la eliminación de sus síntomas, a través de los alimentos que ingerimos o los que eliminamos de la dieta.

### 1.1.5 Cultura cristiana: ideas sobre la glotonería

San Pablo (siglo I d.C), en su Epístola a los Filipenses, afirmó: <<Los enemigos de la cruz de Cristo cuyo final es la destrucción, cuyo dios es su barriga>>.

Tanto San Agustín en el siglo V, como Gregorio I en el siglo VII, incorporaron la Glotonería como uno de los siete pecados capitales.

La obesidad era infrecuente en esta época y, por tanto, no se solía relacionar con la glotonería durante este periodo

## **1.2 Obesidad en la modernidad.**

### 1.2.1 Siglo XV- Obesidad en la Cultura Cristiana

Como se revisó en la obesidad en la época antigua, los cristianos condenaban claramente la glotonería, claramente relacionada con la obesidad, de hecho el pintor Hieronymus Bosch (1450-1516) en su célebre obra “Los siete pecados capitales” represento el “Ojo de Dios”, que todo lo ve, en un pequeño círculo central, mientras que los pecados eran representados en siete paneles formando en conjunto un gran círculo. El pintor represento así de forma gráfica que una transgresión fácilmente puede conducir a otra.

### 1.2.2 Siglos XVI-XVII

Durante estos siglos, aparecen las primeras monografías cuyo tema principal era la obesidad, todos estos textos se publicaron en latín y hablaron principalmente de aspectos clínicos de la obesidad. Los autores principales de estas teorías fueron el italiano Giovanni Alfonso Borelli y el médico flamenco Jan Baptiste van Helmont.

Cabe destacar que durante esta época aún se consideraba a los obesos como símbolos de fecundidad y de atractivo sexual.

### 1.2.3 Siglo XVIII-primera mitad

Durante la primera mitad del siglo, tuvieron auge los estudios elaborados por Flemmyng en los cuales proporcionaba consejos terapéuticos para combatir la “corpulencia” de acuerdo con las causas que él creía responsables de dicho padecimiento.

Cabe destacar que dichos consejos se resumen a continuación: la dieta deberá de ser moderada en cantidad y los individuos deberán levantarse de la mesa sin haber saciado completamente su apetito, Flemmyng también aconsejó la ingesta pobre en grasa, el ejercicio y tomar baños fríos y por último recomendaba realizar gran cantidad de ejercicio para aumentar la evacuación.

Sin embargo es hasta la segunda mitad del siglo XVIII que Joannes Baptista publica la obra en la que se inicia la teoría anatomopatológica de la obesidad como enfermedad, básicamente lo que aquí se describe son 2 descripciones de obesidad visceral con comorbilidades asociadas las cuales son consideradas un antecedente histórico de la obesidad androide.

Finalmente a finales de dicho siglo, diversos textos médicos refieren agudas y certeras intuiciones acerca del equilibrio calórico, la importancia de la producción de calor, y de la consideración de la obesidad como una enfermedad.

### 1.2.4 Siglo XIX

Durante este siglo existen grandes nombres de autores ingleses que vale la pena resaltar Graves, Stokes, Corrigan, Hodgkin, Bright, Addison y Gull de la escuela de Londres.

Las contribuciones en la obesidad en esta época se limitan a aspectos descriptivos en distintas obras.

### 1.2.5 Siglo XX

Finalizada la segunda guerra mundial, Estados Unidos afianza su liderazgo en la investigación médica y por tanto también en el campo de la obesidad. Se inicia el estudio de la obesidad experimental y la profundización en los estudios metabólicos para mejorar la comprensión de los mecanismos de la acumulación adiposa.

Finalmente en el año de 1994, se produce un descubrimiento de gran importancia en la investigación básica sobre la obesidad. La revista Nature publica el artículo “Positional cloning of the mouse obese gene and its human homologue”, publicado por el grupo de Friedman, de la Rockefeller University. En esta publicación se describe el gen ob en un ratón, la proteína que codifica, la leptina y el gen homólogo en el ser humano. El descubrimiento de la leptina produjo una extraordinaria avalancha de publicaciones que permitió avanzar no solo en el conocimiento de los efectos reguladores de la leptina sobre el apetito del ser humano, si no también y especialmente en el conocimiento de su función de señal (de déficit nutricional y del nivel suficiente de reservas energéticas para poner en marcha la activación del eje hipotálamo en la pubertad).

## 1.3 Marco Referencial

1.3.1 Artículos académicos que abordan problemas de salud a través de un análisis multivariado.

1.3.2 Aplicación de análisis multivariado para la diferenciación de individuos sanos según su contenido sérico de minerales.

Como primer artículo analizaremos el uso de un análisis multivariado para poder diferenciar a un individuo sano de acuerdo a su contenido de minerales, la fuente de dicho artículo proviene del Departamento de Química Analítica, Nutrición y Bromatología. Universidad de La Laguna. Santa Cruz de Tenerife y del Departamento de Ciencias Clínicas. Universidad de Las Palmas de Gran Canaria. España, los autores

de dicho artículo se citan a continuación: E. Rodríguez Rodríguez, P. Henríquez Sánchez, F. López Blanco, C. Díaz Romero y L. Serra Majem.

A manera de resumen el objetivo primordial de la investigación fue lograr diferenciar a los individuos a través de los niveles de concentración de Na, K, Ca, Mg, Fe, Cu, Zn, Se, Mn y P, y utilizando técnicas multivariantes considerando las variables como residencia, edad, sexo, ejercicio, consumo de agua y tabaquismo.

La metodología aplicada fue a través de una muestra de suero de individuos de entre 20 y 38 años. Con dicha muestra se obtuvieron estadísticas de las variables objeto de estudio como: residencia, edad, sexo, ejercicio, consumo de agua y tabaquismo. La determinación de los minerales se realizó por fotometría de llama<sup>3</sup> (Na y K) y espectrofotometría de absorción atómica<sup>4</sup> con llama aire/acetileno (Ca, Mg, Fe, Cu y Zn), con generación de hidruros<sup>5</sup> (Se) o por cámara de grafito (Mn). El P se determinó por colorimetría<sup>6</sup>

<sup>7</sup>Los resultados obtenidos muestran que el sexo y edad de los individuos afectaron las concentraciones séricas de algunos minerales, Cu y Fe y P y Se respectivamente. La residencia afectó las concentraciones medias de la mayoría de los minerales analizados. Los hábitos de fumar y beber no parecen afectar de forma importante los contenidos medios de los minerales. La práctica de ejercicio físico influyó sobre el contenido sérico de P, Cu y Mn. El tipo de agua consumida influyó sobre el contenido sérico de los electrolitos, y elementos plásticos analizados, pero no en el de los elementos traza.

---

<sup>3</sup> La fotometría de llama se emplea para determinar el sodio y el calcio en una muestra biológica.

Fuente: <http://www.elergonomista.com/tecnicas/fotometria.htm>

<sup>4</sup> En química analítica, la espectrometría de absorción atómica es una técnica para determinar la concentración de un elemento metálico determinado en una muestra

Fuente: <http://www.espectrometria.com>

<sup>5</sup> La generación de hidruros sigue el mismo principio que las otras técnicas de absorción atómica, variando la forma de atomizar los metales a medir. En este caso, en lugar de aplicar calor se añade un reactivo a la muestra a temperatura ambiente que reacciona con los elementos y genera hidruros del elemento gaseosos

Fuente: <http://www.microlabindustrial.com>

<sup>6</sup> Se conoce como colorimetría a la ciencia encargada de medir los colores para obtener la cuantificación de los mismos, favoreciendo así su estandarización.

Fuente: <http://www.quiminet.com>

<sup>7</sup> E. Rodríguez Rodríguez, P. Henríquez Sánchez, F. López Blanco, C. Díaz Romero y L. Serra Majem.

Después de aplicar análisis discriminante, se observa que los individuos menores de 18 años se diferencian razonablemente bien (89% de casos correctamente clasificados) del resto. Existe una cierta tendencia a diferenciarse los individuos en base a la residencia.

Por lo anterior, la diferenciación de los individuos mediante análisis discriminante aplicado sobre los contenidos séricos de los minerales analizados en función del sexo, provincia, residencia y hábitos o estilo de vida fue baja. Sin embargo, los adultos se diferencian razonablemente de los adolescentes y niños, y los individuos residentes en Lanzarote y La Palma, tienden a separarse del resto de los individuos de su provincia.

### 1.3.3 Análisis Multivariado Discriminante en trabajadores con úlcera péptica gastroduodenal y exposición a la nocturnidad.<sup>8</sup>

Como segundo artículo se utilizó el análisis discriminante para medir la relación de los trabajadores con úlcera péptica gastroduodenal y su exposición a la nocturnidad, el autor de dicho artículo es Jesús Romero Madero (Médico especialista de I grado en Gastroenterología, Master en Salud de los Trabajadores. Departamento de Gastroenterología, Hospital Universitario 'Julio Trigo López'. La Habana, Cuba).

El objetivo de la investigación es calcular un modelo matemático-estadístico que evalúe las relaciones entre los factores de riesgo para la aparición de la UPGD en trabajadores expuestos y no expuestos a nocturnidad, y como segundo punto determinar los factores de riesgo, expresados por las variables de los cuestionarios (Inventario para factores de personalidad, Escala sintomática de estrés, Cuestionario de síntomas para exploración de fatiga laboral, Cuestionario de la situación laboral, Escala analógico-visual de adaptación al turno de trabajo) que más contribuyen a las diferencias encontradas entre los trabajadores expuestos y los no expuestos a nocturnidad.

Para tal efecto, la metodología aplicada fue un modelo matemático-estadístico por medio de un análisis discriminante a una muestra de 76 trabajadores, para diferenciar

---

<sup>8</sup>Jesús Romero Madero, Médico especialista de I grado en Gastroenterología, Master en Salud de los Trabajadores. Departamento de Gastroenterología, Hospital Universitario 'Julio Trigo López'. La Habana, Cuba

los trabajadores expuestos y no expuestos a nocturnidad, basándonos en la expresión de los factores de riesgo relacionados con la úlcera péptica gastroduodenal ( $p \leq 0,05$ ). El diseño de estudio fue analítico de corte transversal, con un muestreo no probabilístico. Los instrumentos aplicados para explorar los factores de riesgo fueron: factores de individualidad, personalidad, escala sintomática de estrés, exploración de fatiga laboral, situación laboral y escala analógico-visual de adaptación al turno de trabajo. La función discriminante calculada tuvo una capacidad de pronóstico del 86,8%, y las variables con mayor peso en las diferencias fueron: esfuerzo extrínseco, extroversión, psicotismo y razón esfuerzo extrínseco/baja recompensa. Por último, los resultados obtenidos a manera de resumen, son enlistados a continuación.

1. La función discriminante calculada permitió diferenciar o separar los trabajadores de los grupos de expuestos y de no expuestos a nocturnidad que padecían UPGD, con una efectividad o capacidad de pronóstico del 86,8% ( $p \leq 0,005$ ) en nuestra muestra de estudio.
2. Los factores de riesgo (expresados por las variables contenidas en los cuestionarios) con mayor peso o contribución a las diferencias entre los grupos fueron: esfuerzo extrínseco, extroversión, psicotismo y razón esfuerzo extrínseco/baja recompensa.
3. La inclusión de instrumentos como la función discriminante, calculada en este estudio en los chequeos pre-empleos y sistemáticos de los trabajadores, y tomando en cuenta principios ergonómicos, podría favorecer la prevención de la UPGD en los mismos, y puede constituir un primer paso para su erradicación, dado el carácter multifactorial de su etiología.

#### 1.3.4 Factores Asociados a Accidentes, Enfermedades y Ausentismo Laboral: Análisis de Cohorte de Trabajadores Formales en Chile.

Como tercer artículo, se revisará un artículo sobre los factores que afectan en mayor medida a un grupo de trabajadores en cuanto a los accidentes que pudieran sufrir, las enfermedades y el ausentismo laboral. Los autores de dicho artículo se enlistan a continuación: Lorena Hoffmeister, Carolina Vidal y Clelia Vallebuona<sup>9</sup>, adición a Nelly Ferrer, Pamela Vásquez, MsC y Gabriela Núñez<sup>10</sup>

El objetivo principal propuesto por los autores se describe a continuación: Estimar la asociación entre factores de riesgo y la ocurrencia de enfermedades, accidentes y ausentismo laboral en trabajadores formales.

La metodología asociada es un estudio de cohorte retrospectivo, usando dos fuentes de información: registro de evaluaciones ocupacionales preventivas y de episodios de accidentes y enfermedades laborales (2009-2012). Se realizó un análisis descriptivo de variables de salud y de factores de riesgo. Regresiones logísticas para la ocurrencia de enfermedades y accidentes. Se usó una regresión Poisson-cero-inflado<sup>11</sup> para el ausentismo laboral.

Los resultados obtenidos, muestran que los trabajadores > 60 años tienen un exceso de riesgo de enfermedad laboral (OR-ajustada 19.18 respecto a < 30 años) y los sedentarios una OR-ajustada 1.75. Los < 30 años tienen una OR-ajustada de 1.38, los que tienen circunferencia de cintura (CC) superior a la recomendada una OR-ajustada de 1.31 y los sedentarios 1.23 para la probabilidad de accidentes de trabajo. Las

---

<sup>9</sup> Escuela de Salud Pública, Facultad de Medicina, Universidad Mayor. Santiago, Chile.

<sup>10</sup> Asociación Chilena de Seguridad, Santiago, Chile.

<sup>11</sup> Los modelos de mezclas de distribuciones son combinaciones lineales convexas de distribuciones de probabilidad. Los ceros de estos modelos pueden ser una mezcla de ceros estructurales y muestrales. De esta clase de modelos, el llamado modelo de regresión Poisson Inflado con Ceros (PIC) es el más usado para datos de conteo con exceso de ceros. Este modelo fue propuesto por Lambert (1992). En estos modelos los ceros se dividen en dos grupos, uno tiene los ceros provenientes de la distribución que genera a la variable respuesta, el otro grupo tiene a los ceros "extra". Los ceros del primer grupo se modelan con la distribución Poisson. Un cero en este grupo ocurre con probabilidad  $1-p$ . Los ceros extra ocurren con probabilidad  $p$ . Lambert (1992), Welsh et al. (1996), y Böhning et al. (1999) presentan aplicaciones del modelo PIC.

Fuente: <http://www.uv.mx>

mujeres tienen una OR-ajustada 1.99 y CC-superior 1.29 para la probabilidad de accidente de trayecto. Ser mujer (IRR 1.45), tener > 60 años (IRR 2.69), trabajar en actividades inmobiliarias (IRR 2.37) y en explotación de minas (IRR 2.38), tienen una probabilidad mayor de más días de ausentismo. Por lo tanto, los factores no modificables (sexo y edad), junto con factores modificables, como el sedentarismo y el estado nutricional, contribuyen a eventos laborales no deseados.

A lo largo de estos artículos se observó el uso de técnicas estadísticas multivariadas para diversos análisis relacionados con la salud, ya que el uso de estas técnicas permite analizar de forma simultánea dos o más variables, lo cual ofrece una visión de un conjunto de fenómenos objeto de estudio y puede revelar las complejas interacciones o correlaciones que a menudo existen entre ellos.

## CAPÍTULO II. OBESIDAD



Durante este segundo capítulo se abordarán de manera general algunos conceptos básicos sobre obesidad y sobrepeso. Se conocerá la definición de la obesidad, los tipos de obesidad, las principales causas y consecuencias y las medidas para la reducción de la misma.

Por último se revisarán estadísticas tanto mundiales como en México y se abordará con mayor detalle la situación actual en México.

## **2.1 Definiciones: Sobrepeso y Obesidad**

El sobrepeso y la obesidad<sup>12</sup> se han convertido al día de hoy en un problema de salud mundial que ha cobrado muchas vidas y es el detonante de muchas enfermedades. Ésta es medida a través del indicador IMC<sup>13</sup>, el cual mide la relación entre el peso y la talla, este indicador es el que se utiliza frecuentemente para identificar el sobrepeso y la obesidad en los adultos.

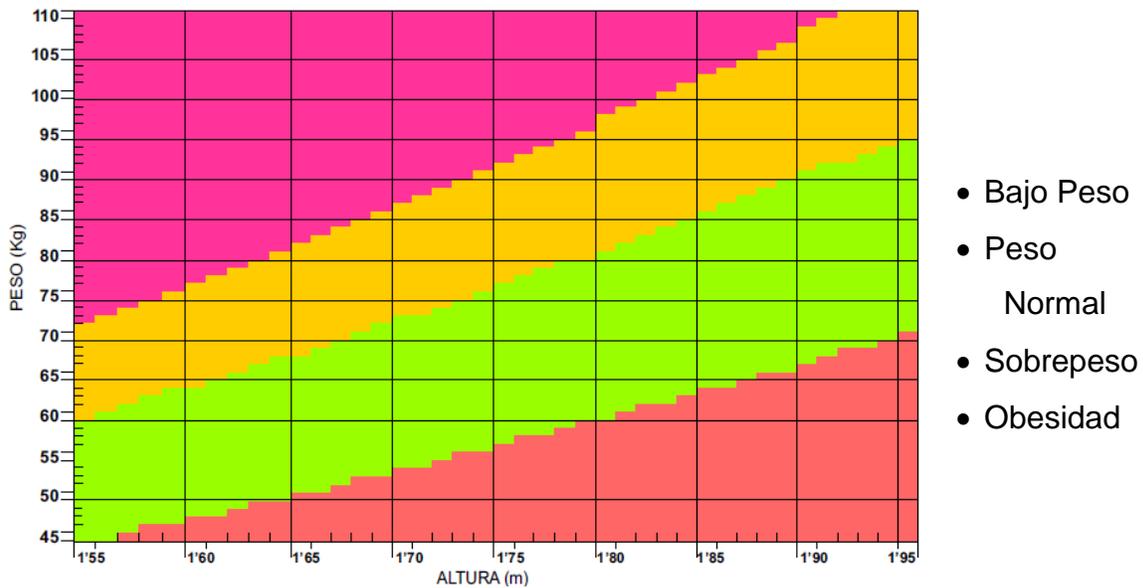
El IMC proporciona la medida más útil del sobrepeso y la obesidad en la población, puesto que es la misma para ambos sexos y para los adultos de todas las edades. Sin embargo, hay que considerarla a título indicativo porque es posible que no se corresponda con el mismo nivel de grosor en diferentes personas.

---

<sup>12</sup> Definida por la Organización Mundial de la Salud como una acumulación anormal o excesiva de grasa que puede ser perjudicial para la salud.

<sup>13</sup> El índice de masa corporal (IMC) es un indicador simple que se calcula dividiendo el peso de una persona en kilos por el cuadrado de su talla en metros (kg/m<sup>2</sup>).

**Gráfico 1 La definición de la Obesidad según la OMS<sup>14</sup>**



Fuente: Elaboración Propia, según SENC: Guías alimentarias para la población española, recomendaciones para una dieta saludable, 2001

## 2.2 Tipos de Obesidad

De acuerdo a la Organización Española de la Obesidad<sup>15</sup>, la obesidad puede clasificarse de acuerdo a los siguientes factores: riesgo, la distribución de grasa, la causa u otros factores.

### 2.2.1 Tipos de obesidad según el riesgo

Según la Organización Mundial de la Salud (OMS) la obesidad puede dividirse en:

- Obesidad grado 1: IMC 30-34 Kg/m<sup>2</sup>
- Obesidad grado 2: IMC 35-39,9 kg/m<sup>2</sup>
- Obesidad grado 3: IMC 40-49,9 kg/m<sup>2</sup>
- Obesidad grado 4 u obesidad extrema: IMC >50 kg/m<sup>2</sup>

<sup>14</sup> Organización Mundial de Salud, 1948

<sup>15</sup> Fuente: Leal Lombardo, María (2002) La Obesidad, *Web Oficial de la obesidad.org*. Fecha de consulta: 20 de Mayo 2015. URL: <http://laobesidad.org.es/tipos-obesidad>

### 2.2.2 Tipos de obesidad según la distribución de grasa

- Obesidad abdominal o androide (forma de manzana): la localización del exceso de grasa se encuentra en el abdomen, el tórax y la cara. Está muy asociada con la diabetes y enfermedades del corazón.
- Obesidad periférica o ginoide (forma de pera): la grasa acumulada se encuentra en muslos y caderas. Se da especialmente en mujeres y está relacionada con problemas como las várices y la artrosis<sup>16</sup> en las rodillas.
- Obesidad homogénea: no existen una predominancia de exceso de grasa en una zona localizada sino que la grasa se reparte por el cuerpo en las mismas proporciones.

---

<sup>16</sup> La **artrosis** u **osteoartritis** (nombre derivado del término anglosajón *osteoarthritis*, menos usado en la práctica clínica) es una enfermedad producida por el desgaste del cartílago, tejido que hace de amortiguador al proteger los extremos de los huesos y que favorece el movimiento de la articulación. Es la enfermedad reumática más frecuente, especialmente entre personas de edad avanzada.

## Gráfico 2. Tipos de Obesidad Abdominal.



Fuente: Fuente: Leal Lombardo, María (2002) La Obesidad, Web Oficial de la obesidad.org. Fecha de consulta: 20 de Mayo 2015. URL: <http://laobesidad.org.es/tipos-obesidad>

### 2.2.3 Tipos de obesidad según la causa

- **Genética:** es una de las más comunes. Es aquella en la que el paciente ha recibido una herencia genética o predisposición para tener obesidad.
- **Dietética:** Es otra de las más comunes. Se da por llevar un estilo de vida sedentario unido a la ingesta de alimentos de alto valor calórico.
- **Obesidad por desajuste:** aparece debido a un fallo (desajuste) en la saciedad. La persona nunca se encuentra satisfecha al comer y siempre siente la necesidad de seguir ingiriendo más alimentos.
- **Defecto termogénico:** no es un tipo de obesidad frecuente. La causa el organismo al no quemar las calorías eficientemente.
- **Obesidad Nerviosa:** la sufren aquellas personas que sufren otras enfermedades como la hipo actividad u otros tipos de problemas psicológicos. La obesidad la produce el sistema nervioso central cuando altera los mecanismos de saciedad.

- Enfermedades endocrinas<sup>17</sup>: Tampoco es muy frecuente. Es generada por enfermedades endocrinas como el hipercortisolismo o el hipotiroidismo.
- Obesidad por medicamentos: algunos medicamentos producen acumulación de grasa. Ocurre con algunos tipos de antidepresivos o corticoides por ejemplo.
- Obesidad cromosómica: se asocia a defectos cromosómicos como aquellas personas que sufren síndrome de Down o de Turner.

#### 2.2.4 Otros tipos de obesidad

- Secundaria: Es consecuencia de algunas enfermedades que dan lugar al aumento de la grasa corporal.
- Primaria: Cuando existe un desequilibrio entre gasto energético y la ingesta de alimentos.
- Hipertrófica: Cuando aumenta el volumen de adipocitos.
- Hiperplásica: Cuando aumenta el número de células adiposas.

---

<sup>17</sup> Las enfermedades endocrinas son trastornos del sistema endocrino, un complejo sistema de órganos y glándulas que ayudan a controlar las funciones vitales del organismo a través de unas sustancias llamadas hormonas. Los desequilibrios de las hormonas pueden afectar a la salud de muchas formas, y algunas enfermedades endocrinas son más comunes que otras. Las enfermedades endocrinas incluyen la diabetes, una enfermedad endocrina muy común, pero también dentro de estas enfermedades hay otras más raras como la acromegalia y la enfermedad de Cushing.

## 2.3 Causas del sobrepeso y la obesidad

A manera de resumen, las causas fundamentales del sobrepeso y la obesidad están relacionadas principalmente con: desequilibrio energético entre calorías consumidas y gastadas, descenso en la actividad física, medio ambiente, genética, medicamentos, factores emocionales y otros; a continuación una breve descripción de cada uno de ellos:

- a) **Falta de balance energético:** En la mayoría de las personas, el sobrepeso y la obesidad se producen por falta de balance energético. Para que haya balance energético, la energía que se ingiere en los alimentos debe ser igual a la que se gasta. La energía que se ingiere es la cantidad de energía o de calorías que se obtiene de los alimentos y bebidas. La energía que se gasta es la cantidad de energía que el cuerpo usa en funciones como respirar, digerir los alimentos y mantenerse activo.
- b) **Un estilo de vida poco activo:** Las personas poco activas tienen más probabilidades de subir de peso porque no queman las calorías que consumen en los alimentos y bebidas. Un estilo de vida poco activo también eleva el riesgo de sufrir enfermedades coronarias, presión alta, diabetes, cáncer del colon y otros problemas de salud.
- c) **Medio ambiente:** Nuestro medio ambiente no siempre contribuye a que se tengan hábitos saludables; de hecho, estimula la obesidad. Esto se debe a razones como las siguientes: Los horarios de trabajo, pasan mucho tiempo yendo entre la casa y el trabajo, Las raciones extra grandes de comida, La falta de acceso a alimentos saludables, la publicidad de la industria de alimentos que anuncian bocadillos con alto contenido de calorías y grasas, y bebidas llenas de azúcar.
- d) **Los genes y los antecedentes familiares:** En estudios realizados con gemelos idénticos que se criaron en hogares distintos se ha demostrado que los genes

tienen mucha influencia en el peso de una persona. El sobrepeso y la obesidad tienden a ser hereditarios. Sus probabilidades de tener sobrepeso son mayores si uno de sus padres o ambos tienen sobrepeso u obesidad.

- e) **Medicina:** Ciertas medicinas pueden provocar aumento de peso. Entre ellas se cuentan algunos corticosteroides<sup>18</sup>, antidepresivos y anticonvulsivantes. Estas medicinas pueden disminuir la velocidad en la que el cuerpo quema calorías, aumenta el apetito o hace que el cuerpo retenga más agua, todo lo cual puede conducir a un aumento de peso.
- f) **Factores emocionales:** Algunas personas comen más de lo acostumbrado cuando están aburridas, enojadas o estresadas. Con el tiempo, comer en exceso conducirá a un aumento de peso y puede causar sobrepeso u obesidad.
- g) **Otros factores:** Hormonales, hábito de fumar, edad, embarazo y desequilibrios en el sueño.

---

<sup>18</sup> Los corticosteroides (del lat. cortex, —ícis, corteza, y esteroide) o corticoides son una variedad de hormonas del grupo de los esteroides (producida por la corteza de las glándulas suprarrenales) y sus derivados.

Los corticosteroides están implicados en una variedad de mecanismos fisiológicos, incluyendo aquellos que regulan la inflamación, el sistema inmunitario, el metabolismo de hidratos de carbono, el catabolismo de proteínas, los niveles electrolíticos en plasma y, por último, los que caracterizan la respuesta frente al estrés.

Estas sustancias pueden sintetizarse artificialmente y tienen aplicaciones terapéuticas, utilizándose principalmente debido a sus propiedades antiinflamatorias e inmunosupresoras y a sus efectos sobre el metabolismo.

## 2.4 Consecuencias comunes del sobrepeso y la obesidad para la salud

### 2.4.1 Físicas:

Un IMC elevado es un importante factor de riesgo de enfermedades no transmisibles, como:

- Las enfermedades cardiovasculares<sup>19</sup> (principalmente cardiopatía y accidente cerebrovascular).
- La diabetes<sup>20</sup>
- Trastornos del aparato locomotor
- Cánceres<sup>21</sup> (del endometrio, la mama y el colon).

---

<sup>19</sup> Según la OMS, las enfermedades cardiovasculares (ECV), es decir, del corazón y de los vasos sanguíneos, son:

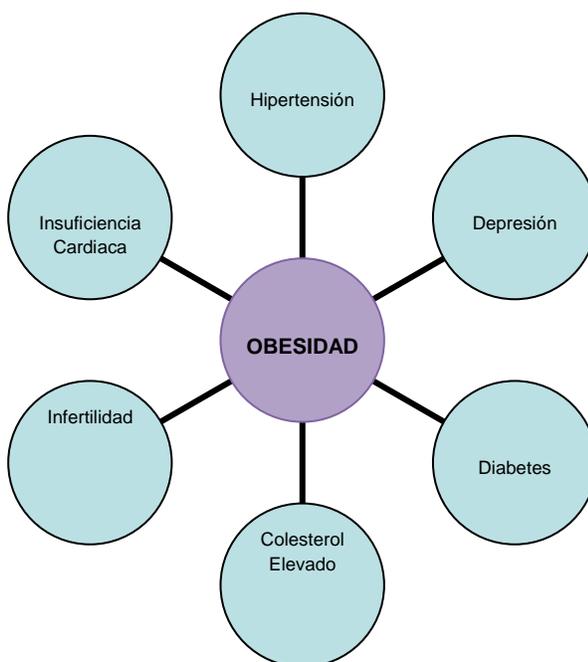
- La cardiopatía coronaria – enfermedad de los vasos sanguíneos que irrigan el músculo cardíaco (miocardio);
- Las enfermedades cerebrovasculares – enfermedades de los vasos sanguíneos que irrigan el cerebro;
- Las arteriopatías periféricas – enfermedades de los vasos sanguíneos que irrigan los miembros superiores e inferiores;
- La cardiopatía reumática – lesiones del miocardio y de las válvulas cardíacas debidas a la fiebre reumática, una enfermedad causada por bacterias denominadas estreptococos;
- Las cardiopatías congénitas – malformaciones del corazón presentes desde el nacimiento; y

Las trombosis venosas profundas y embolias pulmonares – coágulos de sangre (trombos) en las venas de las piernas, que pueden desprenderse (émbolos) y alojarse en los vasos del corazón y los pulmones.

<sup>20</sup> Según la International Diabetes Federation la diabetes es una afección crónica que se desencadena cuando el organismo pierde su capacidad de producir suficiente insulina o de utilizarla con eficacia. La insulina es una hormona que se fabrica en el páncreas y que permite que la glucosa de los alimentos pase a las células del organismo, en donde se convierte en energía para que funcionen los músculos y los tejidos. Como resultado, una persona con diabetes no absorbe la glucosa adecuadamente, de modo que ésta queda circulando en la sangre (hiperglucemia) y dañando los tejidos con el paso del tiempo.

<sup>21</sup> Según el Instituto Nacional de Cáncer éste es definido como un conjunto de enfermedades relacionadas en las cuales algunas de las células del cuerpo empiezan a dividirse sin detenerse y se diseminan a los tejidos del derredor.

### Gráfico 3. Consecuencias sobrepeso y obesidad



Fuente: Elaboración Propia en base a Organización Mundial de la Salud, 2012

#### 2.4.2 Económicas y Sociales: Afrontar una doble carga de morbilidad

Muchos países de ingresos bajos y medianos actualmente están afrontando una "doble carga" de morbilidad.

Mientras continúan lidiando con los problemas de las enfermedades infecciosas y la desnutrición, estos países están experimentando un aumento brusco en los factores de riesgo de contraer enfermedades no transmisibles como la obesidad y el sobrepeso, en particular en los entornos urbanos.

No es raro encontrar la desnutrición y la obesidad coexistiendo en un mismo país, una misma comunidad y un mismo hogar.

En los países de ingresos bajos y medianos, los niños son más propensos a recibir una nutrición prenatal, del lactante y del niño pequeño insuficiente. Al mismo tiempo, están expuestos a alimentos hipercalóricos ricos en grasa, azúcar y sal y pobres en micronutrientes, que suelen ser poco costosos. Estos hábitos alimentarios, juntamente con una escasa actividad física, tienen como resultado un crecimiento brusco de la obesidad infantil, al tiempo que los problemas de la desnutrición continúan sin resolver.

## **2.5 Reducción del sobrepeso y la obesidad**

### a) Acciones Individuales

El sobrepeso y la obesidad, así como sus enfermedades no transmisibles asociadas, son en gran parte prevenibles. En el plano individual, las personas pueden: limitar la ingesta energética procedente de la cantidad de grasa total y de azúcares; aumentar el consumo de frutas y verduras, así como de legumbres, cereales integrales y frutos secos; realizar una actividad física periódica (60 minutos diarios para los jóvenes y 150 minutos semanales para los adultos).

La responsabilidad individual solamente puede tener pleno efecto cuando las personas tienen acceso a un modo de vida saludable. Por consiguiente, en el plano social es importante: dar apoyo a las personas en el cumplimiento de las recomendaciones mencionadas más arriba, mediante un compromiso político sostenido y la colaboración de las múltiples partes interesadas públicas y privadas, y lograr que la actividad física periódica y los hábitos alimentarios más saludables sean económicamente asequibles y fácilmente accesibles para todos, en particular las personas más pobres.

La industria alimentaria puede desempeñar una función importante en la promoción de una alimentación saludable: reduciendo el contenido de grasa, azúcar y sal de los alimentos elaborados; asegurando que todos los consumidores puedan acceder física y económicamente a unos alimentos sanos y nutritivos; poner en práctica una

comercialización responsable, y asegurar la disponibilidad de alimentos sanos y apoyar la práctica de una actividad física periódica en el lugar de trabajo.

## b) Acciones Globales

La respuesta de la OMS Adoptada por la Asamblea Mundial de la Salud<sup>22</sup> en 2004, sobre régimen alimentario, actividad física y salud expone las medidas necesarias para apoyar una alimentación saludable y una actividad física periódica. La Estrategia exhorta a todas las partes interesadas a adoptar medidas en los planos mundial, regional y local para mejorar los regímenes de alimentación y actividad física entre la población.

La Declaración Política de la Reunión de Alto Nivel de la Asamblea General de las Naciones Unidas sobre la Prevención y el Control de las Enfermedades No Transmisibles adoptada en septiembre de 2011, reconoce la importancia crucial de reducir el nivel de exposición de las personas y las poblaciones a dietas poco sanas y al sedentarismo.

Esa Declaración manifiesta el compromiso de promover la aplicación de la Estrategia Mundial OMS sobre Régimen Alimentario, Actividad Física y Salud, incluida, según proceda, la introducción de políticas y medidas orientadas a promover dietas sanas e incrementar la actividad física en toda la población.

Dicho Plan de acción trata de partir de la labor del Convenio Marco de la OMS para el Control del Tabaco y de la Estrategia Mundial OMS sobre Régimen Alimentario, Actividad Física y Salud. El Plan de acción contribuirá a avanzar en las 9 metas mundiales relacionadas con las enfermedades no transmisibles que deben haberse cumplido en 2025, entre las que cabe mencionar una reducción relativa de un 25%

---

<sup>22</sup> La Asamblea Mundial de la Salud es el órgano decisorio supremo de la Organización Mundial de la Salud. Su reune en Ginebra en mayo de cada año con la asistencia de delegaciones de los Estados Miembros. La función principal es determinar las políticas de la Organización, designar al Director General, supervisar las políticas financieras, y revisar y adoptar el programa de presupuesto propuesto.

en la mortalidad prematura debida a enfermedades no transmisibles y la contención de los índices mundiales de obesidad en los correspondientes a 2010.

Para informar mejor y preparar una respuesta integral a la obesidad en la niñez en particular, la Directora General de la OMS estableció una comisión de alto nivel para acabar con la obesidad infantil, integrada por 15 eminentes personalidades de diversas procedencias profesionales y geográficas. La Comisión examinará los mandatos y estrategias existentes y los ampliará para corregir sus deficiencias, fomentará la sensibilidad sobre la obesidad infantil y generará una dinámica favorable para la acción.

## **2.6 Estadísticas Mundiales de la Obesidad**

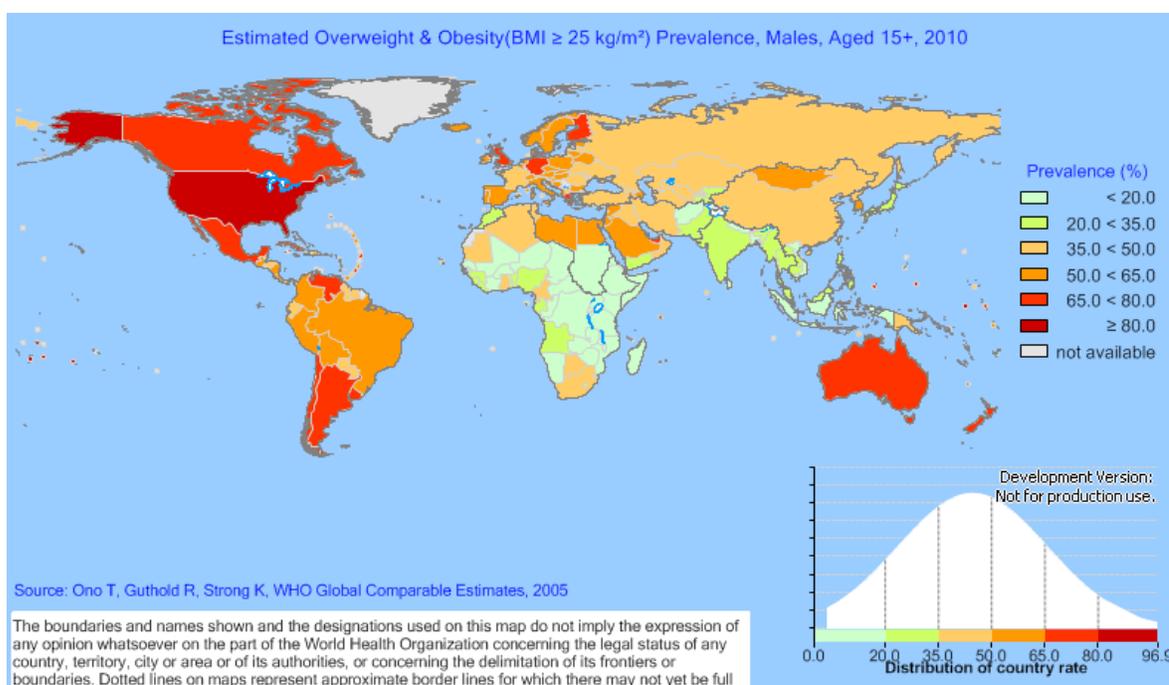
A continuación se presentan algunas estimaciones mundiales recientes de la OMS:

- Desde 1980, la obesidad se ha más que doblado en todo el mundo.
- En 2014, más de 1900 millones de adultos de 18 o más años tenían sobrepeso, de los cuales, más de 600 millones eran obesos.
- En 2014, el 39% de las personas adultas de 18 o más años tenían sobrepeso, y el 13% eran obesas.
- La mayoría de la población mundial vive en países donde el sobrepeso y la obesidad se cobran más vidas de personas que la insuficiencia renal.
- En 2013, más de 42 millones de niños menores de cinco años tenían sobrepeso.
- En general, en 2014 alrededor del 13% de la población adulta mundial (un 11% de los hombres y un 15% de las mujeres) eran obesos.
- La prevalencia mundial de la obesidad se ha multiplicado por más de dos entre 1980 y 2014.
- En 2013, más de 42 millones de niños menores de cinco años de edad tenían sobrepeso. Si bien el sobrepeso y la obesidad tiempo atrás eran considerados un problema propio de los países de ingresos altos, actualmente ambos trastornos están aumentando en los países de ingresos

bajos y medianos, en particular en los entornos urbanos. En los países en desarrollo con economías emergentes (clasificados por el Banco Mundial en países de ingresos bajos y medianos) el incremento porcentual del sobrepeso y la obesidad en los niños ha sido un 30% superior al de los países desarrollados.

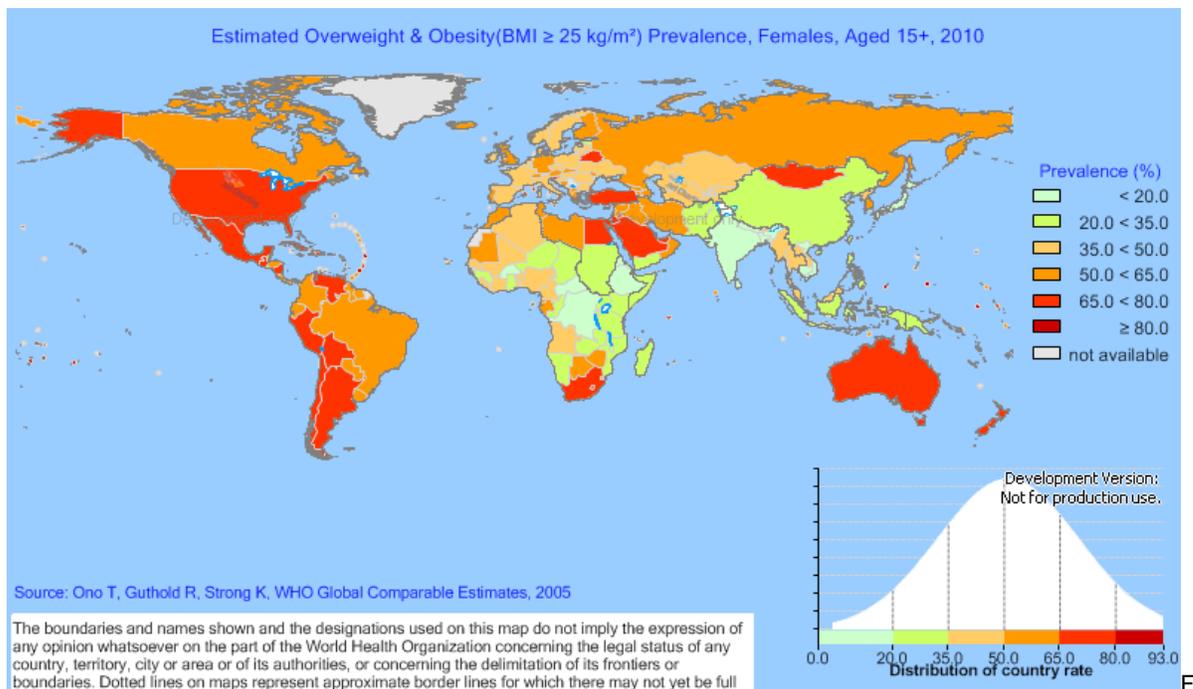
- En el plano mundial, el sobrepeso y la obesidad están relacionados con un mayor número de defunciones que la insuficiencia renal. La mayoría de la población mundial vive en países donde el sobrepeso y la obesidad se cobran más vidas que la insuficiencia renal (estos países incluyen a todos los de ingresos altos y la mayoría de los de ingresos medianos).

**Gráfico 4. Sobrepeso y Obesidad Mundial (Hombres)**



Fuente: Organización Mundial de la salud, 2005

## Gráfico 5. Sobrepeso y Obesidad Mundial (Mujeres)



Fuente: Organización Mundial de la salud, 2005.

## 2.7 Obesidad en México

### 2.7.1 Sobrepeso y Obesidad

De acuerdo a la Secretaría de Salud de México, la obesidad y el sobrepeso son el principal problema de Salud Pública en México, pues nuestro país es el primer lugar mundial en niños con obesidad y sobrepeso<sup>23</sup>, y segundo en adultos.

México gasta 7% del presupuesto destinado a salud para atender la obesidad, sólo debajo de Estados Unidos que invierte el 9%.

<sup>23</sup> Fuente: OMS (Organización Mundial de la Salud), 2012

La mala alimentación, el sedentarismo, la falta de acceso a alimentos nutritivos, son factores determinantes del sobrepeso y la obesidad.

**Tabla 1. Estadísticas Obesidad en México (por grupos)**

Grupos	México		Distrito Federal	
	%	No. Personas	%	No. Personas
Mujeres mayores de 20 años	72	20.52 millones	75.4	2.3 millones
Hombres mayores de 20 años	66	16.96 millones	69.8	69.8 millones
Niños en edad escolar	26	5.54 millones	35	481,785

Fuente: Olaiz-Fernández, Rivera-Dommarco , Shamah-Levy , Rojas R, Villalpando-Hernández S, Hernández-Avila M, Sepúlveda-Amor J. ENSANUT 2006. Cuernavaca, México: INSP.

### 2.7.2 Consecuencias

- Mortalidad 12 veces mayor en jóvenes de 25 a 35 años
- 25% de las incapacidades laborales son por padecimientos relacionados con la obesidad
- Gastos de entre 22% y 34% superiores en el ingreso familiar
- Tres de cada cuatro camas de hospital las ocupan pacientes con enfermedades relacionadas con la obesidad

### 2.7.3 Programa de Prevención y Control de la Obesidad.

- \*Áreas de Intervención
- \*Promoción de la Salud
- \*Atención integral al paciente
- \*Atención especializada
- \*Vigilancia epidemiológica

#### 2.7.4 Normatividad

Ley para la Prevención y Tratamiento de la Obesidad, el Sobrepeso y los Trastornos Alimenticios en el D.F.

La Secretaría de Salud del Distrito Federal debe:

1. Diseñar, realizar y coordinar, campañas de prevención sobre nutrición y alimentación sana, difundiendo en los centros de salud, hospitales, planteles escolares y espacios públicos, las causas que provocan el sobrepeso, la obesidad y los trastornos alimenticios, así como las formas de prevenir y atender estos problemas;
2. Aplicar un programa masivo para incentivar una alimentación saludable entre la población del D.F., y
3. Generar y difundir bases de datos, desagregadas por grupo de edad, sexo y ubicación geográfica que registren la incidencia de trastornos alimenticios en la población indicando peso, talla y masa corporal, poniendo especial énfasis en los planteles de educación básica.

2.7.5 Acciones: Programa Muévete y Métete en cintura, cuyos principales objetivos son:

1. Generar un amplio movimiento social informado y organizado para promover estilos de vida saludables, a través acciones de orientación alimentaria y fomento de la actividad física.
2. Incidir en la modificación de los determinantes del sobrepeso, obesidad y sedentarismo.  
Fomentar la incorporación de la actividad física en la vida cotidiana de la comunidad capitalina, enfatizando en niños y adolescentes.

3. Mediciones de Índice de Masa Corporal en espacios públicos y eventos masivos  
Pláticas, cursos, talleres y seminarios en unidades médicas
4. Distribución de material informativo e interactivo a población abierta
5. Clínica de Bariatría en el Hospital Rubén Leñero
6. 7 clínicas de atención al sobrepeso, la obesidad y los trastornos alimenticios (en proceso)
7. Detección y tratamiento oportuno de pacientes en hospitales y centros de salud
8. Educación en salud en las escuelas (clases y libros de texto)
9. Regulación de compra y venta de alimentos, dentro y fuera de escuelas (lineamientos determinados por las autoridades federales y locales)
10. Hábitos alimenticios saludables (orientación a padres de familia)
11. Ejercicio y deporte (ampliar los horarios de ejercicio en escuelas e incentivar la actividad física en el hogar)
12. Mejorar la calidad nutritiva de los productos (acuerdo con empresas y cámaras de comercio)
13. Reforzar el Programa “Ángel” en materia de detección y tratamiento del sobrepeso y la obesidad.

A lo largo de este capítulo se revisaron algunos conceptos básicos sobre la obesidad, desde su definición, los tipos de obesidad, las principales causas que provocan un incremento en el IMC, las consecuencias de la obesidad y sobrepeso tanto en la salud como el impacto económico; todo esto nos servirá para poner en contexto el tema tan grave que representa y cómo se puede mitigar.

Adicional a esto se revisaron algunas estadísticas que nos permiten conocer la situación actual en el caso particular de México.

## CAPÍTULO III. ANÁLISIS DE DATOS CATEGÓRICOS



A lo largo de este capítulo se revisarán algunos conceptos básicos para el análisis de datos categóricos, como lo son las distribuciones de probabilidad, escalas de medición, inferencias estadísticas (pruebas de hipótesis), tablas de contingencia, pruebas de independencia, riesgo relativo y cociente de momios.

Como segundo punto se revisarán los modelos lineales generalizados y sus características más sobresalientes.

Se procederá con la revisión de la regresión logística y el por qué del uso de ella en nuestro modelo.

Cabe destacar que para este segundo capítulo, la información está basada en los libros y artículos siguientes: Salud Pública en México (2000), Hamza K (1995), Guerriero V. (2000), Agresti (2013) y John A (1983-1989).

La información aquí presentada considera un resumen de contenido de los trabajos presentados por estos autores.

### **3.1 Distribuciones e inferencias para datos categóricos**

#### **3.1.1 Escalas de Medición.**

La medición<sup>24</sup>, consiste en asignar un número o una calificación a alguna propiedad específica de un individuo, una población o un evento usando ciertas reglas, para tal efecto existen distintas escalas de medición, las cuales serán revisadas a detalle a continuación:

Las variables nominal<sup>25</sup> y ordinal<sup>26</sup> se denominan también categóricas, por otra parte las variables de escala de intervalo<sup>27</sup> o de razón<sup>28</sup> se denominan variables numéricas.

---

<sup>24</sup> Fuente: Salud Pública Méx 2000; Vol. 42(4):337-348

<sup>25</sup> La escala nominal sólo permite asignar un nombre al elemento medido. Esto la convierte en la menos informativa de las escalas de medición.

<sup>26</sup> La escala ordinal, además de las propiedades de la escala nominal, permite establecer un orden entre los elementos medidos.

Para el caso de los datos categóricos, las escalas más usadas son: Binarias, Nominales, Ordinales, Discretos de Intervalo con pocos valores, continuos agrupados en categorías.

Ejemplos de variables con tipo de escala nominal:

- Preferencias Electorales
- Uso de anticonceptivos
- Número de CURP

. Ejemplos de variables con tipo de escala ordinal:

- Preferencia a productos de consumo.
- Etapa de desarrollo de un ser vivo.
- Clasificación de películas por una comisión especializada.

Los siguientes son ejemplos de variables de escala de intervalo:

- Temperatura de una persona.
- Sobrepeso respecto de un patrón de comparación.

Finalmente algunos ejemplos de variables con la escala de razón:

- Altura de personas.
- Cantidad de litros de agua consumido por una persona en un día.

---

<sup>27</sup> La escala de intervalo, además de todas las propiedades de la escala ordinal, hace que tenga sentido calcular diferencias entre las mediciones.

<sup>28</sup> La escala de razón permite, además de lo de las otras escalas, comparar mediciones mediante un cociente.

- Velocidad de un auto en la carretera.
- Número de goles marcados por un jugador de básquetbol en un partido.

### 3.1.2 Distribuciones de Probabilidad

A) **Distribución Binomial:** El experimento binomial, consiste en una secuencia de  $n$  ensayos idénticos, y sólo hay 2 posibles resultados éxito o fracaso (en cada ensayo), adicional es importante mencionar que los ensayos son independientes.

La función de distribución binomial es<sup>29</sup>

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Donde  $x = \{0, 1, 2, \dots, n\}$ ,

$$\text{Siendo } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Las combinaciones de  $n$  en  $x$  ( $n$  elementos tomados de  $x$  en  $x$ )

B) **Distribución Multinomial:** El experimento multinomial, consiste en una secuencia de  $n$  ensayos idénticos, con  $m$  posibles resultados, adicional es importante mencionar que los ensayos son independientes.

Por lo tanto, sea la variable aleatoria  $x_i$ , que indica el número de veces que se ha dado el resultado  $i$  sobre los  $n$  sucesos. El vector  $x = (x_1, \dots, x_k)$  sigue una distribución multinomial con parámetros  $n$  y  $p$ , donde  $P = (p_1, \dots, p_k)$ .

Nótese que en algunos campos las distribuciones categóricas y multinomial se encuentran unidas, y es común hablar de una distribución multinomial cuando el término más preciso sería una distribución categórica.

---

<sup>29</sup> Fuente: Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statist. Probab. Lett.* 23 21–25.

La función de probabilidad de la distribución multinomial es como sigue<sup>30</sup>:

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \Pr(X_1 = x_1 \text{ y } \dots \text{ y } X_k = x_k)$$
$$= \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}, & \text{cuando } \sum_{i=1}^k x_i = n \\ 0 & \end{cases}$$

C) **Distribución Poisson:** Dicha distribución posee las siguientes características:

Una variable aleatoria con distribución Poisson es útil para estimar el número de ocurrencias de cierto evento sobre un intervalo de tiempo o espacio.

Es una variable aleatoria discreta que puede asumir una secuencia infinita de valores ( $x=0, 1, 2, 3, \dots$ ). Un claro ejemplo sería:

- El número de vehículos que llegan a una caseta de cobro en un cierto período de tiempo
- El número de clientes que llegan a un cajero automático durante un cierto período de tiempo

Las propiedades de un experimento Poisson son:

- La probabilidad de ocurrencia es la misma para cualquiera dos intervalos de igual longitud
- La ocurrencia o no ocurrencia en cualquier intervalo es independiente de la ocurrencia o no ocurrencia en cualquier otro intervalo

---

<sup>30</sup> Fuente: Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statist. Probab. Lett.* 23 21–25.

La función de probabilidad de la distribución Poisson es como sigue<sup>31</sup>:

$$f(k, \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Donde

$k$  es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente  $k$  veces).

$\lambda$  es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y se está interesado en la probabilidad de que ocurra  $k$  veces dentro de un intervalo de 10 minutos, se usará un modelo de distribución de Poisson con  $\lambda = 10 \times 4 = 40$ .

$e$  es la base de los logaritmos naturales ( $e = 2,71828\dots$ )

**D) Distribución Hipergeométrica:** La distribución hipergeométrica está muy relacionada con la Binomial ya que una variable aleatoria hipergeométrica estima el número de éxitos en  $n$  ensayos. Sin embargo tiene las siguientes características:

- Los ensayos no son independientes
- La probabilidad de éxito cambia de ensayo a ensayo

La distribución de probabilidad hipergeométrica se usa para calcular la probabilidad de que en una muestra aleatoria de  $n$  elementos, seleccionados sin reemplazo, se obtengan  $x$  éxitos y  $n-x$  fracasos.

$$f(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}, \text{ para } 0 \leq x \leq r$$

---

<sup>31</sup> Fuente: Guerriero V. «Power Law Distribution: Method of Multi-scale Inferential Statistics». J. Mod. Math. Fr.

Dónde:  $f(x)$ =probabilidad de  $x$  éxitos en  $n$  ensayos

$n$  = número de ensayos

$N$  = número de elementos en la población

$r$  = número de éxitos en la población

E) Distribución Chi-Cuadrada<sup>32</sup>, ésta es muy importante en el análisis de datos categóricos, no como una distribución para los datos, sino como una distribución muestral para varios estadísticos

**Importante:**

Una distribución Chi-cuadrada con grados de libertad  $gl$ , tiene media  $gl$  y varianza  $2(gl)$

Si  $Z$  es  $N(0,1)$ , entonces  $Z^2$  tiene una distribución Chi-cuadrada con 1  $gl$

Si  $Z_1, \dots, Z_v$  son Normales independientes con media 0 y desviación estándar 1, entonces  $Z_i^2 \sim \chi_v^2$

### 3.1.3 Inferencia Estadística

La inferencia estadística<sup>33</sup> nos servirá de apoyo para conocer el comportamiento de la población, para tal efecto se hará uso de las siguientes técnicas:

---

<sup>32</sup> En realidad la distribución ji-cuadrada es la distribución muestral de  $s^2$ . O sea que si se extraen todas las muestras posibles de una población normal y a cada muestra se le calcula su varianza, se obtendrá la distribución muestral de varianzas.

<sup>33</sup> La inferencia estadística es el conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una muestra, cual es el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad.

Los métodos paramétricos de la inferencia estadística se pueden dividir, básicamente, en dos: métodos de estimación de parámetros y métodos de contraste de hipótesis. Ambos métodos se basan en el conocimiento teórico de la distribución de probabilidad del estadístico muestral que se utiliza como estimador de un parámetro.

### a) Estimación por máxima verosimilitud<sup>34</sup>

Una vez que se tienen los datos, la función de verosimilitud es la probabilidad de obtener esos datos como función de los parámetros desconocidos

- El estimador de máxima verosimilitud es el valor del parámetro que maximiza dicha función
- Es común maximizar Log(función de verosimilitud)

Los estimadores de máxima verosimilitud:

- Con muestras grandes tienden a tener distribución Normal
- Son asintóticamente consistentes, convergiendo al parámetro cuando n crece
- Son asintóticamente eficientes, produciendo errores estándar no mayores que los de otros métodos de estimación

### b) Inferencias para la distribución Binomial

- El estimador de MV:  $\hat{\pi} = y/n$

$$E\hat{\pi} = \pi, \text{Var}\hat{\pi} = \pi(1-\pi)/n$$

- El estadístico de Wald para probar

$$H_0: \pi = \pi_0$$

$$Z_W = \frac{\hat{\pi} - \pi_0}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} \sim N(0,1)$$

---

<sup>34</sup> Sea  $X_1, \dots, X_n$  una muestra aleatoria de una población  $X$  con función de verosimilitud  $L(\theta)$ . Para cada muestra particular  $(x_1, \dots, x_n)$ , la estimación de máxima verosimilitud de  $\theta$  es el valor  $\hat{\theta}^{MV}$  que maximiza la verosimilitud. Es decir:  $L(x_1, \dots, x_n; \hat{\theta}^{MV}) = \max_{\theta} L(x_1, \dots, x_n; \theta)$  El estimador de máxima verosimilitud,  $\hat{\theta}^{MV}(X_1, \dots, X_n)$ , es aquél que evaluado en cada muestra particular nos da la estimación de máxima verosimilitud ( $\hat{\theta}^{MV}(x_1, \dots, x_n)$ ).

La forma normal del estadístico de score es preferible ya que utiliza  $\pi_0$  en lugar de  $\pi$  en el denominador

- Intervalo de confianza de Wald:  $\pi \pm z\alpha/2\pi(1-\pi)n$

c) Inferencias para la distribución Multinomial

- El estimador de MV:

$$\hat{\pi}(j) = n_j/n$$

- Prueba  $\chi^2$  de Pearson evalúa si los parámetros multinomiales son iguales a ciertos valores:

$$H_0: \pi_j = \pi_{j0}, j=1, \dots, c \text{ donde } \sum \pi_{j0} = 1$$

Cuando la hipótesis nula es cierta,  $\mu_j = n\pi_{j0}, j=1, \dots, c$

Pearson propuso:

$$X^2 = \sum (n_j - \mu_j)^2 / \mu_j \sim \chi^2_{c-1}$$

d) Prueba de Máxima Verosimilitud: Una prueba alternativa a la de Pearson es la prueba de máxima verosimilitud, basada en el estadístico:

$$G^2 = 2 \sum_j n_j \log (n_j / n\pi_{j0}) \sim \chi^2_{c-1}$$

Este estadístico se conoce como el estadístico Chi-cuadrado de MV. Tanto el estadístico de Pearson como el de MV tienen en el límite distribuciones  $\chi^2$

## 3.2 Tablas de Contingencia e Inferencia

### 3.2.1 Definición Tablas de Contingencia

La tabla ji- cuadrada o de contingencia<sup>35</sup> ( $\chi^2$ ) se utiliza principalmente:

- Para probar si una serie de datos observada, concuerda con el modelo (serie esperada) de la información.
- Para probar las diferencias entre las proporciones de varios grupos (tabla de contingencia).

Para todos los casos,

H<sub>0</sub>: No hay diferencia o no hay dependencia entre variables

H<sub>1</sub>: Hay diferencia o si hay dependencia entre variables

#### **Pasos para realizar la tabla de contingencias $\chi^2$**

1) Plantear las hipótesis:

$$H_0 = p_1 = p_2 = p_3 \dots = p_k$$

H<sub>1</sub>: al menos dos proporciones son diferentes.

- 2) Construir una tabla que contenga los valores observados.
- 3) Sumar los totales de los renglones y columnas de los valores observados.
- 4) Debajo de cada valor observado poner el valor esperado utilizando la fórmula:

---

<sup>35</sup> En estadística las tablas de contingencia se emplean para registrar y analizar la relación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales)

$$E_{ij} = \frac{(\text{total de } i - \text{ésimo renglón} \times \text{total de } j - \text{ésima columna})}{n}$$

- 4) Calcular el valor del estadístico de prueba  $\chi^2$  usando la fórmula:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

dónde:

$O_{ij}$  = Valor observado de la celda i,j.

$E_{ij}$  = Valor esperado de la celda i,j

- 5) Determinar los grados de libertad mediante:

$$gl = (r-1)(c-1)$$

dónde

r = número de renglones

c = número de columnas

- 6) Calcular el valor crítico en la tabla  $\chi^2$

- 7) Criterio de decisión: si el valor crítico < valor del estadístico de prueba entonces se rechaza  $H_0$

### 3.2.2 Pruebas de Independencia

Adicional a la prueba  $\chi^2$  existen otras pruebas que de igual manera nos ayudan a evaluar la presencia de independencia. Estas son descritas a continuación.

**Tabla 2 Test de Independencia en Tablas de Contingencia**

Tipos de Pruebas	¿Qué es?	¿En qué se basa?	Prueba de Hipótesis	Observaciones	ESTADÍSTICO
Prueba X cuadrada	Prueba de independencia	Compara los valores observados contra los estimados.	H0: independencia	-Tratan a las clasificaciones como nominales. -No contestan todos los cuestionamientos por lo que se debe analizar los residuales. -Requiere muestras grandes. -No distingue entre variables respuesta y explicatorias.	El estadístico $X^2$ de Pearson se define como: $X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \sim \chi_{(I-1)(J-1)}^2 \text{ asintóticamente}$ Para el caso de tablas 2x2: $X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1+}n_{2+}n_{+1}n_{+2}}$
Prueba de Máxima Verosimilitud G cuadrada	Prueba de independencia	Compara los valores observados contra los estimados.	H0: $\pi_{ij} = (\pi_{i+})(\pi_{j+})$ Independencia	-Tratan a las clasificaciones como nominales. -Requiere muestras grandes. -No distingue entre variables respuesta y explicatorias.	El estadístico $G^2$ de MV se define como: $G^2 = 2 \sum_i \sum_j n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right) \sim \chi_{(I-1)(J-1)}^2 \text{ asintóticamente}$
Prueba M cuadrada	Prueba de independencia VS la alternativa de correlación diferente de 0.	Entre más grande es r en valor absoluto, más se rechaza H0. Por lo que grandes valores de $M^2$ , contradicen la independencia.	H0: independencia Ha: $\rho \neq 0$	-Se calcula usando el r de Pearson. Para grandes n, se aproxima a una Chi-cuadrada con $gdl=1$ . -Se asignan scores. -Es más potente para datos ORDINALES. -No distingue entre variables respuesta y explicatorias.	$M^2 = (n-1)r^2 \sim \chi_1^2 \text{ asintóticamente}$
Prueba exacta de Fisher	Es una pruebas de independencia para tablas 2x2.	Usa el odds ratio y la cola derecha de la distribución hipergeométrica para la prueba.	H0: $\theta=1$ Independencia en las dos variables. Ha: $\theta > 1$	El valor p es para 2 colas, debes de dividir entre dos por la Ha: $\theta > 1$ -Por la distribución hipergeométrica, no es posible lograr un nivel de significancia específico, como 0.05	Si $t_0 = \text{valor observado de } n_{11}$ $\text{valor} - p = P(n_{11} \geq t_0)$

\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

### 3.2.3 Riesgo Relativo y Cociente de Momios

El Riesgo relativo<sup>36</sup>, está definido como un cociente de probabilidades

Como se sabe la determinación de la presencia o ausencia de un suceso en dos grupos diferentes es una situación habitual en los estudios clínicos. Los resultados se muestran en una tabla de contingencia 2 x 2.

<sup>36</sup> También conocido como el cociente entre el riesgo en el grupo con el factor de exposición o factor de riesgo y el riesgo en el grupo de referencia (que no tiene el factor de exposición) como índice de asociación.

El procedimiento estadístico habitual para contrastar la presencia de asociación es utilizar la **prueba del  $\chi^2$**  o la **prueba exacta de Fisher**, cuando las frecuencias de las celdas son pequeñas.

Existen diferentes procedimientos para cuantificar la importancia de la asociación. Uno de ellos utiliza el cociente de esos dos riesgos, y nos indica cuánto más probable es que ocurra el suceso en el primer grupo frente al segundo. Es lo que se conoce como **Riesgo Relativo (RR)**.

Actualmente otra medida muy utilizada es la que se denomina "**odds ratio**" (**OR**), o bien cociente de momios. La ventaja del odds es que es otra forma de representar un riesgo, mediante el cociente entre el número de veces que ocurre el suceso frente a cuántas veces no ocurre.

El cociente de los *odds* de los dos grupos es lo que se denomina *odds ratio* y constituye otra forma de cuantificar la asociación entre dos variables dicotómicas. El cálculo del *odds ratio* a partir de los datos de frecuencia de una tabla 2 x 2 es muy sencillo:

$$OR = \frac{a \times d}{b \times c}$$

Razones para utilizar el *odds ratio*

Está claro que la interpretación del riesgo relativo es más intuitiva que la del odds ratio ¿por qué entonces se usa tanto éste último?. Hay varias razones. En primer lugar puede ser calculado en cualquier tabla 2 x 2, mientras que el riesgo relativo no se puede calcular en los estudios caso-control.<sup>37</sup>

---

<sup>37</sup> La primera investigación de este tipo apareció en el año 1962, los estudios de casos y controles han hecho aportaciones sobresalientes a la medicina. Fueron de particular importancia los estudios realizados durante la década de los cincuenta, ya que estos originaron los procedimientos metodológicos generales que hoy en día son utilizados.

Por otro lado permite examinar el efecto que otras variables pueden causar en esa asociación (como podrían ser por ejemplo la edad, el sexo, etc), ya que  $exp(\text{coeficiente de la ecuación de regresión logística})$  pueden ser interpretados como el *odds ratio* de la variable correspondiente.

Otra propiedad interesante del odds ratio es que no varía aunque se cambie el orden de las categorías en cualquiera de las variables, lo que no ocurre con el riesgo relativo.

Por otro lado, cuando los riesgos (o los odds) en ambos grupos son pequeños (inferiores al 20 %) el odds ratio se aproxima bastante al riesgo relativo pudiendo considerarse como una buena aproximación de éste.

Sin embargo, cuando se trata de sucesos frecuentes la discrepancia entre ambos parámetros se acentúa. Así por ejemplo si la frecuencia del suceso fuese 10 veces mayor en ambos grupos que la del ejemplo original, la tabla que se tendría sería:

a	b
c	D

$$\frac{a}{(a+c)}$$

$$\frac{b}{(b+d)}$$

### Error estándar del *Odds Ratio* y del Riesgo Relativo

---

El objetivo principal de los estudios retrospectivos es probar alguna hipótesis planteada sobre la etiología de una enfermedad, es decir, que estos se dedican al análisis de una presunta relación entre algún factor o característica sospechosa y el desarrollo de cierto padecimiento.

Los estudios retrospectivos se realizan basándose en observaciones clínicas, o a través de análisis especiales, estos revisan situaciones de exposición a factores sospechosos, comparando grupos de individuos enfermos (casos), con grupos de individuos sanos (controles). A partir de la frecuencia observada en cada uno de los grupos expuestos al factor en estudio se realiza un análisis estadístico.

Si luego de haber realizado un análisis estadístico de los datos, se logra determinar la presencia del factor en estudio, entonces se puede realizar una asociación entre dicho factor y la enfermedad.

Básicamente se puede decir que este tipo de estudios busca las causas a partir de un efecto que ya se presentó. Los estudios retrospectivos parten de un efecto y regresan a buscar la causa.

*Fuente: Polanco, Francisco (2002) Estudios Prospectivo y Retrospectivo. Fecha de consulta: 10 de Mayo 2015. URL: <http://www.monografias.com/trabajos5/retropros/retropros.shtml>*

El valor del *odds ratio* está limitado en un extremo, no pudiendo ser negativo, y vale 0 cuando o bien no hay sucesos en el primer grupo ( $a = 0$ ) o en el segundo grupo todos los casos presentan el suceso ( $d = 0$ ). Sin embargo no está limitado su valor superior, pudiendo tomar cualquier valor positivo, por lo que tiene una distribución asimétrica. El **logaritmo del *odds ratio***, en cambio puede tener cualquier valor (positivo o negativo), y tiene una distribución simétrica que es aproximadamente normal; de ahí que se utilice esta transformación tanto para su representación gráfica como para el cálculo de los intervalos de confianza. El **error estándar del logaritmo del *odds ratio*** se estima como:

$$Err. est. (\ln O) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Otro tanto se puede decir del riesgo relativo, para el que, con la configuración de la tabla anterior se tiene.

$$Err. est. (\ln RR) = \sqrt{\frac{1}{a} - \frac{1}{a+c} + \frac{1}{b} - \frac{1}{b+d}}$$

### 3.3 Modelos lineales generalizados

Los modelos lineales (regresión, ANOVA<sup>38</sup>, ANCOVA<sup>39</sup>), se basan en los siguientes supuestos:

1. Los errores se distribuyen normalmente.
2. La varianza es constante.
3. La variable dependiente se relaciona linealmente con la(s) variable(s)

Independiente(s); de manera analítica se tendría:

$$Y_i = B_1 + B_2X_{2i} + \dots + B_pX_{Pi} + U_i$$

$$E(U_i) = 0 \quad i=1, \dots, n$$

Tomando la esperanza

$$E(Y_i) = B_1 + B_2X_{2i} + \dots + B_pX_{Pi}$$

Además se supone que existe un comportamiento normal

---

<sup>38</sup> El análisis de varianza, conocido como ANOVA sus siglas en inglés, puede ser utilizado en los casos en los que existen más de dos grupos.

Cuando sólo tienes dos muestras puedes utilizar la prueba t para comparar las medias de las muestras, pero esto podría ser poco fiable en caso de más de dos muestras. Si sólo comparas dos medias, entonces la prueba t (muestras independientes) arrojará los mismos resultados que el ANOVA.

Se utiliza para comparar las medias de más de dos muestras. Esto puede quedar más claro si se brinda un ejemplo.

<sup>39</sup> El análisis de la covarianza o ANCOVA, acrónimo del inglés analysis of covariance, es un modelo lineal general con una variable cuantitativa y uno o más factores. El ANCOVA es una fusión del ANOVA y de la regresión lineal múltiple. Es un procedimiento estadístico que permite eliminar la heterogeneidad causada en la variable de interés (variable dependiente) por la influencia de una o más variables cuantitativas (covariables). Básicamente, el fundamento del ANCOVA es un ANOVA al que a la variable dependiente se le ha eliminado el efecto predicho por una o más covariables por regresión lineal múltiple. La inclusión de covariables puede aumentar la potencia estadística porque a menudo reduce la variabilidad.

$$U_i \sim N(0, \sigma^2)$$

Siendo,

$$\mu = E(Y_i) = B_1 + B_2 X_{2i} + \dots + B_p X_{Pi}$$

Por otro lado en el modelo de regresión lineal múltiple se tiene un predictor lineal  $B_1 + B_2 X_{2i} + \dots + B_p X_{Pi}$ , la relación entre la esperanza y el predictor lineal es la identidad, es decir

$$E(Y_i) = B_1 + B_2 X_{2i} + \dots + B_p X_{Pi}$$

La distribución de probabilidad para  $Y_i$  es la normal

En muchas ocasiones, sin embargo, se encuentra con que uno o varios de estos supuestos no se cumplen por la naturaleza de la información.

Estos problemas se pueden llegar a solucionar mediante la transformación de la variable respuesta (por ejemplo tomando logaritmos).

Sin embargo estas transformaciones no siempre consiguen corregir la falta de normalidad, la heterocedasticidad<sup>40</sup> (varianza no constante) o la no linealidad de nuestros datos.

Además resulta muchas veces interpretar los resultados obtenidos, si se utilizan las transformaciones de la variable.

Una alternativa a la transformación de la variable dependiente/respuesta y a la falta de normalidad es el uso de <sup>41</sup>los modelos lineales generalizados (GLM de las siglas en inglés de *Generalized Linear Models*).

---

<sup>40</sup> La heterocedasticidad se presenta cuando la varianza de las perturbaciones no es constante a lo largo de las observaciones.

<sup>41</sup> McCullagh, Peter & Nelder, John A. (1983, 1989) *Generalized Linear Models*, Chapman & Hal

Los GLM son, por tanto, una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (binomiales, Poisson, gamma, etc) y varianzas no constantes.

Ciertos tipos de variables dependientes sufren invariablemente la violación de estos dos supuestos de los modelos normales y los GLM ofrecen una buena alternativa para tratarlos.

*Se estaría en el supuesto GLM* cuando la variable dependiente/respuesta/endógena es:

- Un variable de conteo, en concreto, casos (ejemplo: número de colisiones, accidentes, viviendas destruidas...)
- Un variable de conteo de casos expresados éstos como proporciones (ejemplo; porcentaje de heridos graves en accidentes, porcentaje de no carnet...)
- Una variable establecida como binaria (como en este caso lo es la obesidad)

### 3.3.1 Varianza no Constante

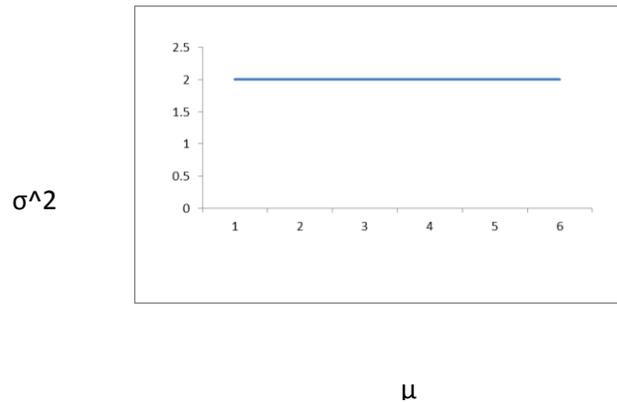
Al analizar submuestras<sup>42</sup> de datos de los que se dispone, se puede encontrar el investigador con diversas realidades:

- El supuesto central que se hace en los modelos lineales es que la varianza es constante así se tendría que al variar la media la varianza se mantiene constante:

---

<sup>42</sup> Es una parte más pequeña de una muestra original, creado por el recorte, la subdivisión, la división o la recolección diferenciada de la muestra original. (Ciencias) Una porción de la muestra original que es representativa en la naturaleza a la de la muestra original, asegurando de este modo la equivalencia en los resultados de las pruebas y análisis ya sea sobre la submuestra o el material original, independiente de su tamaño.

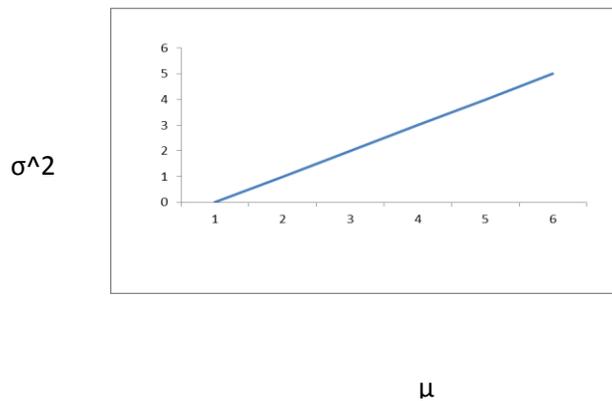
**Gráfico 6. Varianza Constante (Supuesto I)**



\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

- En el caso de variables de conteo como variable dependiente, sin embargo, donde ésta se expresa en números enteros y en dónde puede haber muchos ceros en los datos, la varianza se suele incrementar linealmente con la media:

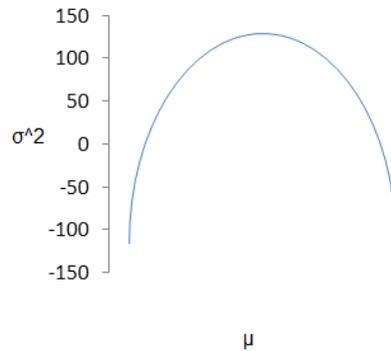
**Gráfico 7. Varianza no Constante (Supuesto II)**



\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

- Con proporciones de eventos como variable explicada es muy posible que la varianza se comporte en forma de U invertida :

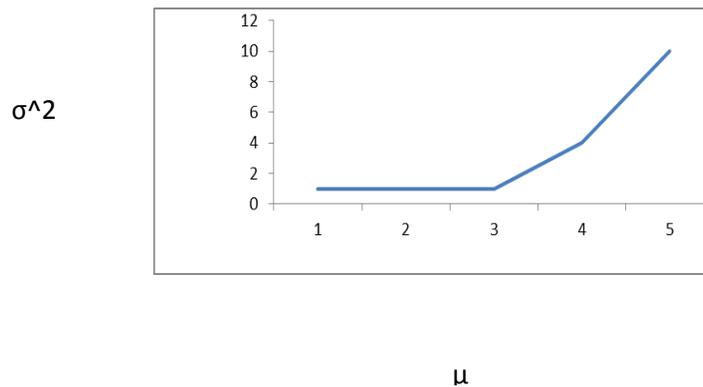
### Gráfico 8. Varianza no Constante (Supuesto III)



\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

- Cuando la variable respuesta/dependiente se aproxime a una distribución Gamma, entonces la varianza se incrementa de una manera no lineal con respecto a la media:

### Gráfico 9. Varianza no Constante (Supuesto IV)



\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

Son en estos casos de varianza no constante donde será necesario aplicar los GLM.

### 3.3.2 Normalidad

Muchos datos tienen una estructura no normal. Las herramientas habituales para tratar la ausencia de normalidad eran la transformación de la variable respuesta o la adopción de métodos no paramétricos<sup>43</sup>. Otra alternativa, son los modelos lineales generalizados o GLM. Los GLM permiten especificar distintos tipos de distribución de errores así:

- Poisson, muy útiles para conteos de acontecimientos. Ejemplo: número de heridos por accidentes de tráfico; número de hogares asegurados que dan parte de siniestro al día.
- Binomiales, de gran utilidad para proporciones y datos de presencia/ausencia. Ejemplo: tasas de mortalidad; tasas de infección; porcentaje de siniestros mortales.
- Gamma, muy útiles con datos que muestran un coeficiente de variación constante, esto es, en donde la varianza aumenta según aumenta la media de la muestra de manera constante. Ejemplo: número de heridos en función del número de siniestros.
- Exponenciales, muy útiles para los análisis de supervivencia.

Además, los modelos lineales, habituales, asumen que tanto la variable respuesta como los errores del modelo siguen una distribución normal<sup>44</sup>. Una distribución normal que es, como es sabido continua. En ocasiones, sin embargo, la variable dependiente sigue una distribución que no es continua y, por tanto, los valores estimados por el modelo han de seguir el mismo tipo de distribución que los datos de partida. Cualquier otro tipo de valor estimado por el modelo no deberá ser válido desde un punto de vista lógico, aunque en la práctica no se presta mucha atención a esto. Por ejemplo, un investigador está interesado en predecir cuántos accidentes se producen al día en un determinado

---

<sup>43</sup> Prueba estadística que no genera premisas respecto a la distribución de las observaciones.

<sup>44</sup> En estadística y probabilidad se llama distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece aproximada en fenómenos reales.

*Fuente: Weisstein, Eric W. «Normal Distribution Function». En Weisstein, Eric W. MathWorld. Wolfram Research.*

municipio en base a datos de días con determinado número de accidentes. En este caso, es razonable asumir que la variable dependiente seguirá una distribución de tipo Poisson y no una normal como en muchas ocasiones se utiliza por “comodidad”.

### 3.3.3 Función vínculo/ligadura

Otra razón por la que un modelo lineal puede no ser adecuado para describir un fenómeno determinado es que la relación entre la variable respuesta y la(s) variable(s) independiente(s) no es siempre lineal.

La función de vínculo se encarga de linealizar la relación entre la variable dependiente y la(s) variable(s) independiente(s) mediante la transformación de la variable respuesta. Por ejemplo la relación entre el número de incendios forestales y la temperatura exterior.

Esta relación no es del todo lineal, pero se puede linealizar tomando logaritmos en la variable respuesta.

Otra de las utilidades de la función de vínculo, es la de conseguir que las predicciones de nuestro modelo queden acotadas. Por ejemplo, si se tiene datos de conteo, no tiene sentido que nuestras predicciones arrojen resultados negativos, como en el caso del número de incendios. En este caso, una función de vínculo de tipo logarítmica resolverá el problema de la acotación. En otras situaciones la variable respuesta/dependiente es una proporción, entonces los valores estimados tienen que estar entre 0 y 1 o 0 y 100 (valores por debajo de 0 o por encima de 1 o 100 no tienen ningún sentido). En este otro caso, una función de vínculo de tipo 'logit' será más apropiada.

### 3.3.4 Especificación del modelo

La especificación de un modelo lineal generalizado, se realiza en 3 partes:

- a) La componente aleatoria corresponde a la variable  $Y_i$ , que sigue una distribución de la familia exponencial (normal, log normal, poisson, gamma). Además se denotará por  $\mu$  a su esperanza matemática.

- b) La componente sistemática, también llamada predictor lineal<sup>45</sup>, se denota por  $N$  y corresponde al vector de  $n$  componentes. Siendo cada una de ellas igual a:

$$\eta_i = \sum_{j=1}^p \beta_j x_{ji} = \mathbf{x}'_i \boldsymbol{\beta}$$

- c) La función de ligadura relaciona la esperanza matemática de la variable dependiente con el predictor lineal  $N_i = g(\mu_i)$   $i=1, \dots, n$ . La función de ligadura debe ser monótona y diferenciable.

Una variable aleatoria  $Y$  sigue una distribución de la familia exponencial, si su densidad puede escribirse como:

$$f_Y(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\}$$

Siendo  $a(\cdot)$ ,  $b(\cdot)$  y  $c(\cdot)$  funciones conocidas.

Si  $\theta$  es conocido, se cumplirá que  $E(Y) = \mu = b'(\theta)$  donde ' indica la diferenciación respecto a  $\theta$ . Además  $\text{var}(Y) = b''(\theta)a(\phi)$

Cuando se escriba la varianza como  $\text{var}(\mu)$  es porque estará en función del parámetro  $\mu$

---

<sup>45</sup> Predicción lineal es una operación matemática donde los valores futuros de una señal discreta en el tiempo se calculan como una función lineal de muestras anteriores

**Cuadro I Funciones de ligadura/vínculo más utilizadas**

<b>Función de Vínculo</b>	<b>Fórmula</b>	<b>Uso</b>
<b>Identidad</b>	$\mu$	Datos continuos con errores normales (regresión y ANOVA)
<b>Logarítmica</b>	$\text{Log}(\mu)$	Conteos con errores de tipo Poisson
<b>Logit</b>	$\text{Log}(\mu / n-\mu)$	Proporciones (datos entre 0 y 1) con errores binomiales
<b>Recíproca</b>	$1/\mu$	Datos continuos con errores gamma
<b>Raíz Cuadrada</b>	$(\mu)^{(1/2)}$	Conteos
<b>Exponencial</b>	$\mu^n$	Funciones de potencia

\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

Se denominan funciones de ligadura/vínculo canónicas a las funciones que se aplican por defecto a cada una de las distribuciones de errores. Esto no significa que siempre se deba usar una única función de vínculo para una determinada distribución. De hecho, puede ser recomendable comparar diferentes funciones de vínculo para un mismo modelo y ver con cuál se obtiene un mejor ajuste del modelo a los datos. En la siguiente tabla se plasma las funciones de vínculo canónicas para cada una de las distribuciones de errores, así como otras posibles funciones de vínculo habitualmente usadas.

**Cuadro II Funciones de vínculo para la distribución de los errores**

<b>Distribución de errores</b>	<b>Función de vínculo canónica</b>	<b>Otras funciones de vínculo posible</b>
<b>Normal</b>	Identidad	Logarítmica
<b>Poisson</b>	Logarítmica	Identidad, raíz cuadrada
<b>Binomial</b>	Logit	Logarítmica
<b>Gamma</b>	Recíproca	Identidad, Logarítmica

\*Fuente: Agresti, A., (2013), *Categorical Data Analysis*, 3rd Edition, Wiley Series in Probability and Statistics.

En la siguiente tabla se muestran algunas de las combinaciones más comunes de variables respuestas y variables explicativas con distintos tipos de funciones de vínculo y distribuciones de errores.

**Cuadro III Combinaciones de variables de respuesta y explicativas**

<b>Tipo de Análisis</b>	<b>Variable de Respuesta</b>	<b>Variable Explicativa</b>	<b>Función de Vínculo</b>	<b>Distribución de Errores</b>
<b>Regresión</b>	Continua	Continua	Identidad	Normal
<b>ANOVA</b>	Continua	Factor	Identidad	Normal
<b>Regresión</b>	Continua	Continua	Recíproca	Gamma
<b>Regresión</b>	Conteo	Continua	Logarítmica	Poisson
<b>Tabla de Contingencia</b>	Conteo	Factor	Logarítmica	Poisson
<b>Proporciones</b>	Proporción	Continua	Logit	Binomial
<b>Regresión Logística</b>	Binaria	Continua	Logarítmica	Binomial
<b>Análisis de Supervivencia</b>	Tiempo	Continua	Reciproca	Exponencial

\*Fuente: Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

### 3.3.5 Especificación de la estimación

La estimación de los parámetros  $\beta_1 \dots \beta_p$  se realiza por el método de máxima verosimilitud.

Para valorar el ajuste de los modelos lineales generalizados se podrá utilizar el estadístico de Chi Cuadrado. Se define como el doble de la diferencia entre el máximo del logaritmo de la verosimilitud que se podría conseguir con la mínima o máxima parametrización y el valor del máximo del logaritmo de la verosimilitud que se consigue con el modelo que se quiere evaluar.

Los ajustes de  $\mu^i$  que se calculan como  $G^{-1} (\sum_{j=1} B_j X_{ji})$  una vez estimados los parámetros del vector  $\beta$ .

Los residuos de Pearson son los más utilizados y se definen como:

$$r_P = \frac{(y_i - \hat{\mu}_i)}{\sqrt{V(\hat{\mu}_i)}}.$$

### 3.3.6 Construcción y evaluación de un GLM

En la construcción de modelos lineales generalizados es importante tener en cuenta una cosa: no existe un único modelo que sea válido. En la mayoría de los casos, habrá un número variable de modelos plausibles que puedan ajustarse a un conjunto determinado de datos. Parte del trabajo de construcción y evaluación del modelo es determinar cuál de todos estos modelos son adecuados, y entre todos los modelos adecuados, cuál es el que explica la mayor proporción de la varianza sujeto a la restricción de que todos los parámetros del modelo deberán ser estadísticamente significativos. Esto es lo que se conoce como el modelo adecuado mínimo. En algunos casos habrá más de un modelo que describan los datos igual de bien. En estos casos queda a nuestro criterio elegir uno u otro, aunque puede ser recomendable utilizarlos todos y discutir las limitaciones que esto presenta desde el punto de vista inferencial. Los pasos que hay que seguir en la construcción y evaluación de un GLM son muy similares a los de cualquier modelo estadístico. No obstante se detallan:

- Exploración de los datos. (E.D.A. exploratory data analysis ,J.Tukey) Conviene conocer nuestros datos. Puede resultar interesante obtener gráficos que nos muestren la relación entre la variable explicada y cada una de las variables explicativas, gráficos de caja (box-plot) para variables categóricas, o matrices de correlación<sup>46</sup> entre las variables explicativas. El objetivo de este análisis

---

<sup>46</sup> La matriz de correlación R es una matriz cuadrada n x n constituida por los coeficientes de correlación de cada pareja de variables; de manera que tendrá unos en su diagonal principal, y en los elementos no diagonales (i,j) los correspondientes coeficientes de correlación  $r_{ij}$  . La matriz de correlación será, obviamente, simétrica, y conservará las propiedades de ser definida-positiva y tener un determinante no

exploratorio<sup>47</sup> es: a) Buscar posibles relaciones de la variable respuesta/dependiente con la(s) variable(s) explicativa(s); b) Considerar la necesidad de aplicar transformaciones de las variables; c) Eliminar variables explicativas que estén altamente correlacionadas.

- Elección de la estructura de errores y función de vínculo. A veces resultará fácil elegir estas propiedades del modelo. Otras resultará tremendamente difícil, y será a posteriori cuando se compruebe, analizando los residuos, la idoneidad de la distribución de errores elegida. Por otro lado, puede ser una práctica recomendable el comparar modelos con distintas funciones de vínculo para ver cuál se ajusta mejor a nuestros datos.

Ajuste del modelo a los datos. Se debe prestar particular atención a:

- a) Los tests de significación para los estimadores del modelo;
- b) La cantidad de varianza explicada por el modelo. Esto en GLM se conoce como devianza  $D^2$ . La devianza nos da una idea de la variabilidad de los datos. Por ello, para obtener una medida de la variabilidad explicada por el modelo, se deberá de comparar la devianza del modelo nulo (Null deviance) con la devianza residual (Residual

---

negativo, (además el determinante será siempre menor o igual que 1). Puede considerarse como la matriz de varianzas entre las variables tipificadas.

<sup>47</sup> El análisis exploratorio tiene como objetivo identificar el modelo teórico más adecuado para representar la población de la cual proceden los datos muestrales. Dicho análisis se basa en gráficos y estadísticos que permiten explorar la distribución identificando características tales como: valores atípicos o outliers, saltos o discontinuidades, concentraciones de valores, forma de la distribución, etc. Por otra parte, este análisis se puede realizar sobre todos los casos conjuntamente o de forma separada por grupos. En este último caso los gráficos y estadísticos permiten identificar si los datos proceden de una o varias poblaciones, considerando la variable que determina los grupos como factor diferenciador de las poblaciones. También permite comprobar, mediante técnicas gráficas y contrastes no paramétricos, si los datos han sido extraídos de una población con distribución aproximadamente normal.

*Fuente: Universidad de Barcelona (2000) Análisis Exploratorio de Datos, Web Oficial de la Universidad de Barcelona. Fecha de consulta: 10 de Abril 2015.*

*URL: [http://www.ub.edu/aplica\\_infor/spss/cap2-3.htm](http://www.ub.edu/aplica_infor/spss/cap2-3.htm)*

deviance), esto es, una medida de cuánto de la variabilidad de la variable respuesta no es explicado por el modelo, o lo que es lo mismo:

$$D^2 = \frac{\text{Devianza.modelo.nulo} - \text{Devianza.residual}}{\text{Devianza.modelo.nulo}} \cdot 100$$

- Criterios de evaluación de modelos. Se podrá utilizar la reducción de la devianza como una medida del ajuste del modelo a los datos. Los tests de significación para los parámetros del modelo son también útiles para ayudarnos a simplificar el modelo. Sin embargo, un criterio comúnmente utilizado es el llamado Criterio de Información de Akaike (AIC del inglés Akaike Information Criterion). Aunque no se va a explicar aquí los fundamentos matemáticos de este índice, sí se dirá que es un índice que evalúa tanto el ajuste del modelo a los datos como la complejidad del modelo. Cuanto más pequeño es el AIC mejor es el ajuste. El AIC es muy útil para comparar modelos similares con distintos grados de complejidad o modelos iguales (mismas variables) pero con funciones de vínculo distintas.
- Análisis de los residuos. Los residuos son las diferencias entre los valores estimados por el modelo y los valores observados. Sin embargo, muchas veces se utilizan los residuos estandarizados, que tienen que seguir una distribución normal. Conviene analizar los siguientes gráficos:

1.-Histograma de los residuos.

2.-Gráfico de residuos frente a valores estimados. Estos gráficos pueden indicar falta de linealidad, heterocedasticidad (varianza no constante) y valores atípicos.

3.-El gráfico de normalidad (q-q plot), que permite contrastar la normalidad (simetría) de la distribución de los residuos; opcionalmente, pueden ser también de gran utilidad los siguientes gráficos:

En caso necesario, sería preciso volver a plantear el modelo (paso 2), tal vez utilizando una estructura de errores más adecuada, otra función de vínculo o incluso eliminando ciertos datos que pueden estar desviando nuestro análisis.

- Simplificación del modelo. El principio de parsimonia<sup>48</sup> requiere que el modelo sea tan simple como sea posible. Esto significa que no debe contener parámetros o niveles de un factor que sean redundantes. La simplificación del modelo implica por tanto:

1.-La eliminación de las variables explicativas que no sean significativas.

2.-La agrupación de los niveles de factores (variables categóricas) que no difieran entre sí. Esto significa que cada vez que se simplifica el modelo se deben de repetir los pasos 3 y 4. La simplificación del modelo tiene que tener, además, una cierta lógica para el analista y no debe incrementar de manera significativa la devianza residual. Por ello y para llegar a entender bien los datos y las relaciones existentes entre las variables conviene evitar, en la medida de lo posible, los procedimientos automatizados (*p.e. backward/forward stepwise regression procedures*).

### **3.4 Regresión logística**

La regresión logística es una herramienta estadística con una gran capacidad para el análisis de datos de la investigación clínica, de la investigación tecnológica, de la ingeniería y de la epidemiología.

La regresión logística se usa para analizar los problemas en los cuales hay una o más variables independientes que determinan un resultado que se mide con una variable dicótoma que toma solamente dos posibles resultados.

---

<sup>48</sup> El principio de parsimonia (también conocido como navaja de *Occam u Ockham*), el cual establece que la solución más simple suele ser la mejor.

También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico).

En la regresión logística, la variable dependiente es binaria o dicótoma, es decir contiene solamente los datos cifrados como 1 (VERDADERO) o 0 (FALSO), con ella se utiliza un modelo matemático, la transformación logit, para predecir el comportamiento de la variable dependiente.

La regresión logística intenta modelar la relación entre la probabilidad (proporción) de una respuesta positiva y las variables explicativas, los valores numéricos de 0 y 1 se asignan a las dos categorías de una variable binaria, el 0 representa una respuesta negativa y el 1 representa una respuesta positiva, la media de esta variable será la proporción de respuestas positivas.

Esta técnica modela cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos, ella precisa el efecto de un grupo de variables, consideradas potencialmente influyentes, sobre la ocurrencia de un determinado proceso.

La regresión logística compite con el análisis discriminante como un método para analizar las variables de respuesta discreta, la regresión logística es más versátil y más adecuada que el análisis discriminante porque no asumen que las variables independientes se distribuyen normalmente.

Esta herramienta resulta de gran utilidad para establecer el efecto de un grupo de variables, consideradas potencialmente influyentes, sobre la ocurrencia de un determinado proceso.

Se clasifica el valor de la variable respuesta como 0 cuando no se presenta el suceso y con el valor 1 cuando sí está presente y si se busca cuantificar la posible relación entre la presencia del suceso y, un posible factor de riesgo y se toma cómo variable dependiente la probabilidad  $p$  de que ocurra el suceso y se construye la siguiente función:

$$\ln(p/1-p)$$

La relación  $p/(1-p)$  es llamada las probabilidades y el logit es el logaritmo de las probabilidades, o sólo probabilidades de registro.

Si se tiene una variable  $p$  que puede tomar cualquier valor, por lo se puede elaborar para ella una ecuación de regresión tradicional:

$$\ln(p/1-p) = a + bx$$

Este es el tipo de ecuación que se conoce como modelo logístico, donde puede haber varios factores:

$$\ln(p/1-p) = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

### **3.5 Interpretación de los resultados de los modelos de regresión logística.**

En el modelo de regresión lineal el resultado que se obtiene se puede interpretar como la magnitud del cambio de la variable dependiente si se incrementa en una unidad el valor de la variable independiente (en el caso de que la variable independiente sea de tipo continuo) y la magnitud del cambio en la variable dependiente si una característica determinada está o no presente (en el caso de tratar variables de tipo categórico).

Para que la aplicación de un modelo de regresión lineal resulte procedente debe cumplirse que los valores de respuesta ( $y$ ) sean independientes entre sí y la relación entre las variables sea lineal de la forma:

$$Y = f(X_1, X_2, \dots) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$$

Para poder interpretar el resultado del modelo de regresión logística se debe de recurrir al concepto de 'odds', una de las medidas de las que se dispone para poder cuantificar el riesgo. De esta forma el o la 'odds' se define como el cociente de la probabilidad de presentar una característica y la probabilidad de no presentarla, o lo que es lo mismo el

cociente del número de casos que presentan la característica entre el número de casos que no la presentan.

$$\text{Odds} = p / (1-p)$$

Se pueden comparar dos odds, por ejemplo entre los pacientes que padecen una cierta enfermedad si presentan cierta característica o no la presentan, en forma de cociente de ambas (denominada odds ratio), de manera que se pueda concluir si por ejemplo la enfermedad es más frecuente entre los pacientes que presentan dicha característica o no la presentan. Se puede demostrar que los coeficientes obtenidos en la regresión logística son medidas que cuantifican el riesgo de presentar cierta característica respecto a no presentarla en base a la variable de estudio, de manera que:

$$\text{Exp}(\beta) = \text{OR}$$

Donde  $\beta$  es el coeficiente resultado de la regresión logística asociado a una cierta variable participante en el modelo. Cuando la variable independiente tratada es numérica, este valor se interpreta como el cambio en el riesgo cuando se incrementa en uno el valor de la variable, mientras que el resto de variables permanecen constantes.

Siempre que se construye un modelo de regresión es fundamental, antes de pasar a extraer conclusiones, el corroborar que el modelo calculado se ajusta efectivamente a los datos usados para estimarlo. En el caso de la regresión logística una idea bastante intuitiva es calcular la probabilidad de aparición del suceso. Si el ajuste es bueno, es de esperar que un valor alto de probabilidad se asocie con presencia real de la enfermedad y viceversa, si el valor de esa probabilidad calculada es bajo, cabe esperar también ausencia de enfermedad. Esta idea intuitiva se lleva a cabo formalmente mediante la prueba conocida como de Hosmer-Lemeshow<sup>49</sup> la cual, se contrasta mediante una prueba de Ji-Cuadrado<sup>50</sup>.

---

<sup>49</sup> Es otra prueba para evaluar la bondad del ajuste de un modelo de regresión logística, aunque su uso está más discutido que la anterior. La idea es si el ajuste es bueno, un valor alto de la  $p$  predicha se asociará (con una frecuencia parecida a la  $p$ ) con el resultado 1 de la variable binomial. Se trata de calcular para cada observación del conjunto de datos las probabilidades de la variable dependiente que

Finalmente, debe evitarse que en el modelo de regresión planteado pueda producirse el fenómeno de la colinealidad<sup>51</sup>, que daría lugar a soluciones inestables.

A lo largo de este capítulo, se realizó la revisión de los conceptos básicos del análisis multivariado, así como sus aplicaciones.

Se revisó la información de manera inductiva, comenzando de lo particular a lo general, es decir se estudiaron las distintas distribuciones de probabilidad, la inferencia estadística (esto con el objetivo de conocer las pruebas necesarias para la validación correcta de nuestro modelo), los test de independencia y las tablas de contingencia (la cual sirvió para conocer la dependencia entre cada una de las variables), también se revisó la literatura sobre los Modelos Lineales Generalizados con el objetivo de determinar el modelo a elegir para el caso de estudio analizado.

Como último punto se revisó la interpretación de resultados, para poder aplicarlo a nuestro estudio durante el próximo capítulo.

---

predice el modelo, agruparlas y calcular, a partir de ellas, las frecuencias esperadas y compararlas con las observadas mediante la prueba  $\chi^2$ .

*Fuente: Comunidad de Madrid (2005) Prueba Hosmer-Lemeshow, Web Oficial del Hospital Universitario Ramón y Carbajal. Fecha de consulta: 10 de Abril 2015. URL: [http://www.hrc.es/bioest/Reglog\\_5.html](http://www.hrc.es/bioest/Reglog_5.html)*

<sup>50</sup> El estadístico ji-cuadrado sirve para probar la asociación entre dos variables, así como para evaluar en qué medida se ajusta la distribución de frecuencias obtenida con los datos de una muestra, a una distribución teórica o esperada. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula.

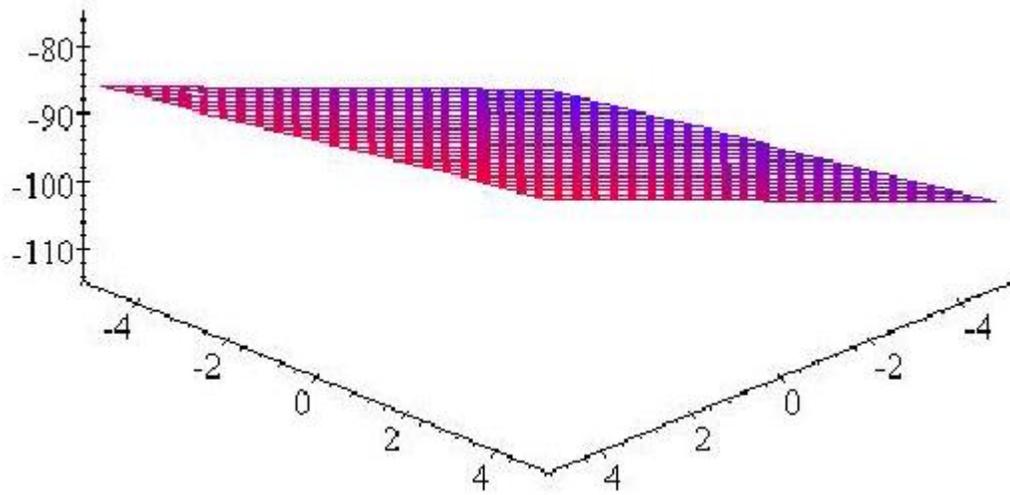
*Fuente: Quevedo, Fernando (2011) Prueba ji-cuadrada, Web Oficial de Revista Biomédica. Fecha de consulta: 1 de Abril 2015. URL:*

*<http://www.medwave.cl/link/Medwave/Series/MBE04/5266>.*

<sup>51</sup> Se habla de colinealidad cuando dos o más variables independientes que se introducen en el modelo de regresión están altamente correlacionadas entre sí.

*Fuente: Moral Peláez, Irene (2014) Modelos de Regresión: lineal simple y regresión logística, Web Oficial Revistas Edén. Fecha de consulta: 20 de Marzo 2015. URL: <http://www.revistaseden.org/files/14-CAP%2014.pdf>*

## CAPÍTULO IV. APLICACIÓN DEL MODELO



A lo largo de este capítulo se conocerá acerca de nuestras variables objeto de estudio y como se obtuvieron los datos (ENSANUT 2012), para tal efecto se analizarán algunos puntos importantes de la Encuesta Nacional de Salud y Nutrición como lo son conocer sus objetivos, su metodología y sus alcances.

Después se procederá con la revisión de la codificación de las variables y como éstas fueron incluidas en el modelo; una vez realizado esto se analizarán las tablas de contingencia para conocer algunos puntos importantes sobre nuestras variables objeto de estudio.

Como paso adicional se calcularán los cocientes de momios para medir la relación entre nuestras variables y si éstas tienen correlación positiva o negativa.

Finalmente, para la construcción del modelo de regresión logística se utilizó el método “introducción” o “enter”, en el cual el proceso de selección de las variables es manual. Se parte de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se va evaluando qué variable es la que menos participa en él y se elimina, volviendo a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que lo componen.

## **4.1 Encuesta Nacional de Salud y Nutrición 2012**

### **4.1.1 Antecedentes ENSANUT 2012<sup>52</sup>**

Desde hace más de 25 años la Secretaría de Salud inició un esfuerzo sistemático por generar un conjunto de datos fidedignos que permitieran conocer las condiciones de salud de la población y sus tendencias, así como la utilización y percepción sobre los servicios de salud. El Instituto Nacional de Salud Pública (INSP) se sumó a esta labor desde una óptica académica y científica con el propósito de desarrollar el abordaje metodológico y las herramientas apropiadas para el desarrollo de encuestas en el ámbito de la salud en el país.

---

<sup>52</sup> Fuente: ENSANUT 2012.

Como parte de la responsabilidad social asumida por el INSP para el estudio científico de los problemas de salud pública, la Encuesta Nacional de Salud y Nutrición 2012 (ENSANUT 2012), observación más reciente del eje del Sistema Nacional de Encuestas de Salud, permitirá contar con información actualizada y 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.

La ENSANUT 2012 permite contar con información de base poblacional y probabilística sobre la cobertura de programas de salud en áreas básicas como las inmunizaciones, la atención a los niños, la salud reproductiva, la atención a padecimientos crónicos, así como sobre los retos en salud como el control de la hipertensión arterial, la diabetes, el sobrepeso y la obesidad, los accidentes y la violencia, entre otros. Permite asimismo analizar la protección en salud en México, y el desempeño de los proveedores de servicios de salud.

Además sienta las bases para afrontar los rezagos en salud y nutrición de los mexicanos y los retos que plantean el acceso a la salud y la protección social en salud. La solidez de la información que provee permite considerarla como base para la conformación del Plan Nacional de Desarrollo y el Programa Sectorial de Salud del gobierno entrante.

#### 4.1.2 Objetivos

Los objetivos de la ENSANUT 2012 son:

- Cuantificar la frecuencia, distribución y tendencia de las condiciones y determinantes de salud y nutrición.
- Examinar la respuesta social organizada frente a los problemas de salud y nutrición de la población, en particular, examinar la cobertura, utilización y calidad percibida de los servicios de salud. Generar evidencia sobre la cobertura de protección en salud de las familias mexicanas.

- Informar sobre el desempeño del sistema de salud y sobre los retos para los planes futuros.

#### 4.1.3 Metodología

Para lograr este objetivo, la ENSANUT 2012 obtuvo información de 50 528 hogares en los cuales se aplicaron 96 031 cuestionarios individuales en los diferentes grupos de edad, así como 14 104 cuestionarios de utilizadores de servicios de salud. La ENSANUT 2012 incluyó la toma de mediciones antropométricas<sup>53</sup> y de sangre capilar en individuos seleccionados, la medición de tensión arterial en una submuestra<sup>54</sup> de adolescentes y adultos y la toma de sangre venosa en una submuestra de individuos de cada grupo de edad, así como información dietética y sobre actividad física.

Los grupos de interés fueron los siguientes:

- Niños de 0 a 4 años
- Niños de 5 a 9 años

---

<sup>53</sup> Las medidas antropométricas se refieren a las mediciones que se realizan en diferentes partes del organismo. Incluyen las destinadas a determinar de forma indirecta la cantidad de grasa total, que están encaminadas a clasificar los individuos según su peso (índice de masa corporal y medición de pliegues entre otros), así como las medidas que conducen a la determinación del reparto de grasa en los diferentes compartimentos del organismo (como el pliegue tricipital, circunferencia del brazo, muñeca, cintura y cadera

*Fuente: Agencia Española de Seguridad Alimentaria y Nutrición (1996) Mediciones Antropométricas, Web Oficial Programa Perseo. Fecha de consulta: 20 de Marzo 2015. URL: <http://www.perseo.aesan.msssi.gob.es/es/profesionales/secciones/mediciones.shtml>*

<sup>54</sup> Submuestra (submuestras plural) Una parte más pequeña de una muestra original, creado por el recorte, la subdivisión, la división o la recolección diferenciada de la muestra original. (Ciencias) Una porción de la muestra original que es representativa en la naturaleza a la de la muestra original, asegurando de este modo la equivalencia en los resultados de las pruebas y análisis ya sea sobre la submuestra o el material original, independiente de su tamaño.

*Fuente: Simple Dictionary (2005) Submuestra, Web Oficial Simple Dictionary. Fecha de consulta: 10 de Marzo 2015. URL: <http://spanish-s.simple-dictionary.com/definitions/?word=subsample>*

- Adolescentes de 10 a 19 años
- Adultos de 20 años o más
- Utilizadores de servicios de salud ambulatorios en los últimos 15 días

El diseño muestral se trata de una encuesta probabilística nacional con representatividad estatal, por estratos nacionales urbano y rural, y una sobremuestra de los hogares con mayores carencias del país<sup>55</sup>.

Para la distribución de las viviendas por entidad se incluyó la sobrerrepresentación de los hogares del país en condiciones de mayor vulnerabilidad, en el supuesto de que en estos hogares se focalizan los apoyos de los programas de salud y sociales.

Para implementar esta sobrerrepresentación, se construyó un índice de rezago social para las áreas geoestadísticas básicas (AGEB) que son las unidades primarias de muestreo. El índice se construyó de forma análoga al índice de rezago social construido por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval)<sup>56</sup> para las localidades en 2005. Luego, las AGEB se ordenaron de acuerdo con el índice en cada estado y se definió para cada uno el estrato de mayor rezago social que agrupaba a 20% de la población. Una vez construido el estrato de mayor rezago social por estado, las AGEB se clasificaron de acuerdo con dos criterios.

---

<sup>55</sup> Los hogares con mayores carencias se identificaron mediante la construcción de un índice de rezago social definido para las áreas geoestadísticas básicas; el índice que se construyó es similar al índice de rezago social construido por el Consejo Nacional de Evaluación de la Política de Desarrollo Social para las localidades en 2005.

*Fuente: Consejo Nacional de Evaluación de la Política de Desarrollo Social (2014) , Índice de Rezago Social, Web Oficial CONEVAL. Fecha de consulta: 10 de Marzo 2015. URL: <http://www.coneval.gob.mx/Paginas/principal.aspx>*

<sup>56</sup> El índice de rezago social considera variables que miden carencias en educación, vivienda, seguridad social y bienes.

Así, en la ENSANUT 2012 se formaron siete estratos<sup>57</sup>, seis que resultan de considerar las combinaciones de urbano, rural y metropolitano, con mayor rezago y menor rezago, y un estrato para las localidades de nueva creación. En el caso del Distrito Federal se formaron tres estratos geográficos dentro del estrato metropolitano dado que todas las viviendas del D.F. fueron clasificadas como metropolitanas.

Se seleccionaron 1 440 viviendas en cada entidad para replicar la distribución por condición de urbanidad y una sobremuestra de 288 viviendas seleccionadas del estrato con mayor rezago social.

Para el marco de muestreo se utilizó la información del Censo de Población y Vivienda 2005 desagregada por AGEB y el listado de localidades de nueva aparición en el Censo 2010.

---

<sup>57</sup> Estrato tiene su origen en el vocablo latino stratus y permite hacer referencia al conjunto de elementos que comparten ciertos caracteres comunes y que se integra con otros conjuntos para la formación de una entidad.

*Fuente: Definición. De (2008) , Definición Estrato, Web Oficial Definición. De . Fecha de consulta: 8 de Marzo 2015. URL:<http://definicion.de/estrato/#ixzz3bD7tQcw0>*

## Cuadro IV: Criterios de estratificación de la ENSANUT 2012

---

a. Mayor (AGEB de mayor rezago social que agrupan a 20% de la población)	1. Rural. Localidades con menos de 2 500 habitantes
b. Menor (AGEB con el restante 80% de la población)	2. Urbano. Localidades con 2 500 o más habitantes y menos de 100 mil habitantes no incluidas en el estrato metropolitano
	3. Metropolitano. Localidades con 100 mil habitantes o más, capitales de los estados o áreas metropolitanas.
	4. Localidades de nueva creación. Localidades que aparecen en el Censo de Población 2010 y no aparecen en el Censo de Población 2005

---

\*Fuente: ENSANUT 2012

Las unidades de análisis definidas para la encuesta son las siguientes:

- 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.
- Utilizadores de servicios de salud son las personas del hogar que buscaron o recibieron atención ambulatoria dentro de las dos semanas anteriores a la fecha de la entrevista por enfermedad, lesión, accidente, rehabilitación, servicio dental, programa de control de diabetes o hipertensión arterial.
- Niños preescolares son las personas del hogar con edad entre 0 y 4 años cumplidos.
- Niños escolares son las personas del hogar con edad entre 5 y 9 años cumplidos.
- Adolescentes son las personas del hogar en el grupo de edad de 10 a 19 años cumplidos.
- Adultos son las personas del hogar con edad mayor a 20 años (cuadro V).

**Cuadro V Parámetros usados para la estimación de tamaño de muestra**

Grupo	Tasa de Respuesta R*	K
Preescolar	0.81	0.2750
Escolar	0.81	0.3077
Adolescente	0.783	0.4837
Adulto	0.783	0.9963
Utilizador	0.81	0.2742

Fuente: ENSANUT 2012

\* Considera la no respuesta de viviendas e individuos

El efecto del procedimiento de muestreo se puede cuantificar por el índice  $deff^{58}$ , que puede interpretarse como el número de unidades colectadas por un procedimiento de muestreo específico que son equivalentes, para propósitos de estimación, con una unidad colectada mediante muestreo aleatorio simple. Para la planeación de la ENSANUT 2012 se utilizó un  $deff=1.836$ , estimado a partir de los resultados de la ENSANUT 2006 y de considerar el sobremuestreo del estrato de mayor rezago social. El tamaño de muestra fue calculado mediante la fórmula siguiente:

$$n = Z^2 \frac{p(1-p)}{\delta^2 (RK)} deff$$

Dónde:

$n$  = Tamaño de muestra en viviendas

$p$  = Proporción a estimar

$Z$  = Cuantil 97.5 de una distribución normal unitaria ( $Z=1.96$ )

$\delta$  = Semiapertura del intervalo de confianza,  $\Pr(|P - P^l| \leq \delta) = 0.95$

$R$  = Tasa de respuesta esperada

$K$  = Porcentaje de viviendas donde se espera obtener alguna persona del grupo de interés

$deff$  = Efecto de diseño de muestra (1.83)

<sup>58</sup>  $deff = Var(\theta^{\wedge}; \text{muestreo M}) / Var(\theta^{\wedge}; \text{muestreo aleatorio simple})$  El  $deff$  compara a la varianza de un procedimiento de muestreo M con la varianza de un muestreo aleatorio simple.

Con base en los parámetros descritos se propuso un tamaño de muestra de 1 719 hogares por entidad. Tamaño de muestra que es suficiente para estimar prevalencias de 10% con las siguientes semiamplitudes: de 2% en adultos, de 3% en adolescentes y de 4% en preescolares y escolares.

#### 4.1.4 Alcances de la ENSANUT

La ENSANUT 2012 es una encuesta que fue diseñada para estimar prevalencias y proporciones de condiciones de salud y nutrición, acceso a servicios y determinantes de la salud. En consecuencia, todas las estimaciones presentadas tienen un error de muestreo debido a que se obtienen conclusiones a partir una muestra y no de la población. El error de muestreo<sup>59</sup> puede ser cuantificado por un intervalo de confianza<sup>60</sup> que puede ser construido para cualquier estimador. En este reporte no se incluyeron intervalos de confianza para todas las estimaciones con el ánimo de facilitar la redacción; sin embargo, éstos pueden ser calculados en las bases de datos de la misma.

Adicionalmente, es importante señalar que dada la cantidad de intervalos de confianza que se generan existe una probabilidad positiva de que algunos de ellos sean equivocados en el sentido de no incluir al parámetro. Por ejemplo, si se calcula un intervalo de confianza al 95% para todos los estados, es de esperarse que en  $(32 \times 0.05 = 1.6)$  el promedio de intervalos equivocados sea 1.6.

---

<sup>59</sup> El error de muestreo es la desviación de la muestra seleccionada de las verdaderas características, rasgos, comportamientos, cualidades o figuras de toda la población.

*Fuente: Explorable AS (2015) Error de Muestreo, Web Oficial Explorable. Fecha de consulta: 20 de Marzo 2015. URL: <https://explorable.com/es/error-de-muestreo>*

<sup>60</sup> En estadística, se llama a un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. La probabilidad de éxito en la estimación se representa con  $1 - \alpha$  y se denomina nivel de confianza. En estas circunstancias,  $\alpha$  es el llamado error aleatorio o nivel de significación, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo.

*Fuente: Sotomayor Velasco, Gabriel; Wisniewski, Piotr Marian (2001). «10.2. Intervalos de confianza para medias». Probabilidad y estadística para ingeniería y ciencias. Cengage Learning Editores. p. 230.*

En consecuencia, no se espera que la encuesta reproduzca con exactitud el tamaño poblacional para cualquier grupo de edad, en particular para grupos de edad donde el tamaño de muestra puede ser insuficiente.

## 4.2 Análisis de las variables objeto de estudio

**Cuadro VI: Variables Objeto de Estudio**

<b>Variable</b>	<b>Codificación</b>	<b>Descripción</b>
<b>Obesidad</b>	SI (1)	Se presenta obesidad cuando se tiene un IMC > 30
	NO (0)	
<b>Sexo</b>	Hombre (0)	
	Mujer (1)	
<b>Edad</b>	20 a 109 años	
<b>Área</b>	Rural (1)	Rural. Localidades con menos de 2 500 habitantes
	Urbana (2)	Urbano. Localidades con 2 500 o más habitantes y menos de 100 mil habitantes no incluidas en el estrato metropolitano
<b>Región</b>	Norte (1)	
	Centro (2)	
	Área Metropolitana (3)	

	Sur (4)	
	Ninguna (1)	
<b>Escolaridad</b>	Primaria y Secundaria (2)	
	Más de Secundaria (3)	
	Alto (4 y 5)	
<b>Nivel Socioeconómico</b>	Medio (2,3,4)	Medido en Percentiles
	Bajo (1,2)	
	Ninguno (0)	
	IMSS (1)	
	ISSSTE (2)	
<b>Afiliación</b>	PEMEX (4)	
	DEFENSA / MARINA (5)	
	SEGURO POPULAR (6)	
	PRIVADO (7)	
	OTRO (8)	
	NA (99)	

---

\*Elaboración Propia, basado en la ENSANUT 2012

### 4.3 Análisis de los datos

Para poder realizar este análisis se utilizaron diversas técnicas estadísticas, para conocer las variables que más impactan la presencia o ausencia de la obesidad.

Para tal efecto, y como se describe a continuación, se utilizó el modelo de regresión logística binaria, el cual nos permitirá evaluar la variable dicotómica (presencia o ausencia de la obesidad) y el impacto de cada una de las variables independientes (objeto de estudio).

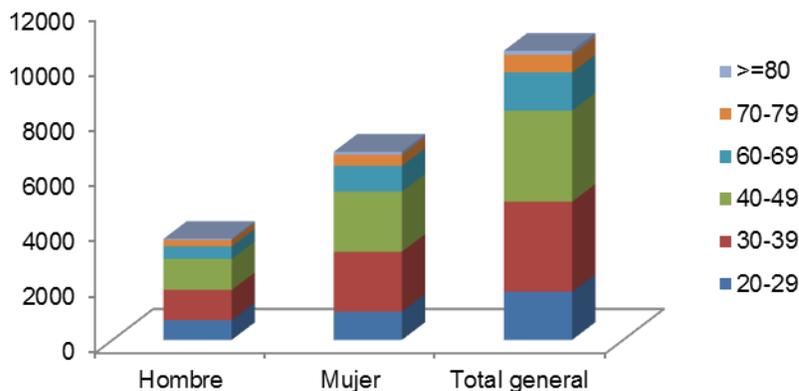
El análisis estadístico se realizó usando el programa Minitab v16 para Windows. Se determinaron prevalencias específicas para cada una de las variables explicativas. Inicialmente, se estableció la influencia en la obesidad de cada una de las variables considerada de manera aislada. Para tal fin se establecieron los ORs crudos y ajustados por edad y sexo, con su respectivo intervalo de confianza. Para establecer la asociación entre la obesidad y las variables explicativas se utilizó el estadístico chi cuadrado, con una significancia del 5%. En cada una de las variables se estableció como grupo de referencia aquel que tuviera una menor probabilidad de padecer obesidad. Para establecer los determinantes sociales y económicos de la obesidad se construyó un modelo de regresión logística incluyendo las variables explicativas: nivel educativo, nivel socioeconómico, área, región, afiliación a servicios de salud, edad y sexo, con la obesidad como variable dependiente.

Para la construcción del modelo de regresión logística se utilizó el método “introducción” o “*enter*”, en el cual el proceso de selección de las variables es manual. Se parte de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se va evaluando qué variable es la que menos participa en él y se elimina, volviendo a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que lo componen.

### 4.3.1 Resultados

Se evaluaron 50,756 personas entre los 20 y los <80 años, se tuvo información de completa para el análisis de 37,580 que representan el 74% del total de observaciones.

**Gráfico 10. Prevalencia de Obesidad en México por sexo y Grupos de Edad.**



\* Elaboración propia / ENSANUT 2012

Como se muestra en el gráfico 1 se podrá concluir que la obesidad en las mujeres es mayor en un 30% y los grupos de edad con mayor índices de obesidad son entre los 30 y los 50 años de edad (alrededor de un 50% de la obesidad se concentra en estos grupos de edad).

#### 4.3.1.1 Inferencias tabla de frecuencias:

Los resultados obtenidos, muestran lo siguiente:

- Se encontró una prevalencia de obesidad de 34% para el conjunto de la población.
- No se encontraron diferencias significativas entre las áreas urbana y rural.
- Las mujeres presentan mayores prevalencias de obesidad (38%) que los hombres (28%).
- Por grupos de edad, la obesidad es mayor en edades de los 40 a los 60 años.

- No se encontraron diferencias significativas entre las regiones del país.
- Se presenta un mayor índice de obesidad en las personas con nivel educativo de primaria y secundaria (36%)
- A mayor nivel socioeconómico se presentan mayores índices de obesidad.
- Los afiliados a PEMEX, Seguros Privados y Otros presentan mayores niveles de obesidad con un porcentaje mayor a un 40%.

Lo anterior nos permitirá conocer los grupos en los que la probabilidad de padecer obesidad es mayor. Adicional a que podremos conocer las condiciones más favorables para la presencia de la obesidad.

Todo lo anterior, nos permitirá establecer las políticas de Salud que podrán tener más impacto para mitigar el problema de la obesidad en nuestro país.

**Tabla 3. Tablas de Frecuencia Obesidad**

		SI		NO		TOTAL		Valor p Chi Cuadrado
		N	%	NO	%	N	%	
SEXO	Hombre	4407	28%	11555	72%	15962	100%	Pearson Chi-Square = 484.652, DF = 1, <b>P-Value = 0.000</b> Likelihood Ratio Chi-Square = 490.225, DF = 1
	Mujer	8319	38%	13299	62%	21618	100%	
	<b>Total</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
EDAD	20-29	1736	23%	5888	77%	7624	100%	Pearson Chi-Square = 948.125, DF = 6 Likelihood Ratio Chi-Square = 992.677, DF = 6, <b>P-Value = 0.000</b>
	30-39	3254	36%	5884	64%	9138	100%	
	40-49	3295	41%	4685	59%	7980	100%	
	50-59	2277	40%	3384	60%	5661	100%	
	60-69	1375	36%	2419	64%	3794	100%	
	70-79	636	27%	1748	73%	2384	100%	
	>=80	153	15%	846	85%	999	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
AREA	Rural	3818	29%	9254	71%	13072	100%	Pearson Chi-Square = 194.042, DF = 1 Likelihood Ratio Chi-Square = 196.433, DF = 1, <b>P-Value = 0.000</b>
	Urbano	8908	36%	15600	64%	24508	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
REGIÓN	Norte	3281	38%	5253	62%	8534	100%	Pearson Chi-Square = 117.003, DF = 3, <b>P-Value = 0.000</b> Likelihood Ratio Chi-Square = 115.848, DF = 3
	Área Metropolitana	371	34%	706	66%	1077	100%	
	Centro	4352	31%	9478	69%	13830	100%	
	Sur	4722	33%	9417	67%	14139	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
ESCOLARIDAD	Ninguna	1057	28%	2741	72%	3798	100%	Pearson Chi-Square = 110.159, DF = 2, <b>P-Value = 0.000</b> Likelihood Ratio Chi-Square = 112.111, DF = 2
	Primaria y Secundaria	8528	36%	15424	64%	23952	100%	
	Más de Secundaria	3141	32%	6689	68%	9830	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
NIVEL SOCIOECONÓMICO	BAJO	3828	29%	9263	71%	13091	100%	Pearson Chi-Square = 197.333, DF = 2, <b>P-Value = 0.000</b> Likelihood Ratio Chi-Square = 199.488, DF = 2
	MEDIO	4455	36%	8049	64%	12504	100%	
	ALTO	4443	37%	7542	63%	11985	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	
AFILIACIÓN	Ninguno	2316	31%	5070	69%	7386	100%	Pearson Chi-Square = 111.882, DF = 8, <b>P-Value = 0.000</b> Likelihood Ratio Chi-Square = 111.120, DF = 8
	DEFENSA/MARINA	44	37%	76	63%	120	100%	
	IMSS	3676	36%	6490	64%	10166	100%	
	ISSSTE	1281	38%	2064	62%	3345	100%	
	PEMEX	53	45%	64	55%	117	100%	
	PRIVADO	49	45%	61	55%	110	100%	
	SEGURO POPULAR	5229	32%	10900	68%	16129	100%	
	OTRO	69	43%	92	57%	161	100%	
	NA	9	20%	37	80%	46	100%	
	<b>Total general</b>	<b>12726</b>	<b>34%</b>	<b>24854</b>	<b>66%</b>	<b>37580</b>	<b>100%</b>	

La tabla de frecuencia anterior, nos muestra los grupos y probabilidades en los que la obesidad tiene mayor probabilidad de presentarse

Como se puede observar, se utiliza la prueba  $\chi^2$  que es una prueba de independencia a través del estadístico. Dado que el valor P es pequeño, indica asociación entre variables

El estadístico  $X^2$  de Pearson se define como:

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)} \text{ asintóticamente}$$

Para el caso de tablas 2x2:

$$X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1+n_2+n_1n_2}$$

#### 4.3.1.2 Análisis Estadístico de los casos de estudio de acuerdo factores socioeconómicos y físicos.

##### 1. Odds ratios de obesidad según la Edad

Hombre			
EDAD	SI	NO	TOTAL
20-29	714	2,635	3,349
30-39	1,091	2,456	3,547
40-49	1,119	2,157	3,276
50-59	754	1,709	2,463
60-69	450	1,271	1,721
70-79	230	891	1,121
>=80	49	436	485
<b>Todas</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>

Mujer			
EDAD	SI	NO	TOTAL
20-29	1,022	3,253	4,275
30-39	2,163	3,428	5,591
40-49	2,176	2,528	4,704
50-59	1,523	1,675	3,198
60-69	925	1,148	2,073
70-79	406	857	1,263
>=80	104	410	514
<b>Todas</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

- **Odds Hombre<sub>20:50</sub> vs 50 89:**  $(2924 \cdot 4307) / (1483 \cdot 7248) = 1.2$ , es decir los odds muestrales de que exista obesidad de los 20 a los 50 años es 1.2 veces más probable que de los 50 en adelante.
- **Odds Mujeres<sub>20:50</sub> vs 50 89:**  $(5361 \cdot 4090) / (2958 \cdot 9209) = 0.8$ , es decir los odds muestrales de que exista obesidad de los 20 a los 50 años es 80% más probable que de los 50 en adelante.

## 2. Odds ratios de obesidad según la Área

Hombre				Mujer			
ÁREA	SI	NO	TOTAL	ÁREA	SI	NO	TOTAL
Rural	1,227	4,453	5,680	Rural	2,591	4,801	7,392
Urbano	3,180	7,102	10,282	Urbano	5,728	8,498	14,226
<b>Todas</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>	<b>Todas</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

- **Odds Hombre** rural vs urbano:  $(1227 \cdot 7102) / (4453 \cdot 3180) = 0.61$ , es decir los odds muestrales de que exista obesidad en el área rural son 0.61 veces los odds que en área urbana.
- **Odds Mujeres** rural vs urbano:  $(2591 \cdot 8498) / (4801 \cdot 5728) = 0.8$ , es decir los odds muestrales de que exista obesidad en el área rural son 0.8 veces los odds que en área urbana.

## 3. Odds ratios de obesidad según la Región

Hombre				Mujer			
REGIÓN	SI	NO	Todos	REGIÓN	SI	NO	Todos
AM	135	327	462	AM	236	379	615
Centro	1,403	4,415	5,818	Centro	2,949	5,063	8,012
Norte	1,232	2,626	3,858	Norte	2,049	2,627	4,676
Sur	1,637	4,187	5,824	Sur	3,085	5,230	8,315
<b>Todas</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>	<b>Todas</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

- El porcentaje de obesidad en el caso de las mujeres se acentúa en el norte con un 44%
- El porcentaje de obesidad en el caso de los hombres se acentúa en el norte con un 32%

#### 4. Odds ratios de obesidad según la Escolaridad.

Hombre				Mujer			
ESCOLARIDAD	SI	NO	Todas	ESCOLARIDAD	SI	NO	Todas
+ Secundaria	1,387	3,020	4,407	+ Secundaria	1,754	3,669	5,423
Ninguna	263	1,178	1,441	Ninguna	794	1,563	2,357
Primaria / Secundaria	2,757	7,357	10,114	Primaria / Secundaria	5,771	8,067	13,838
<b>Todos</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>	<b>Todos</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

- **Odds Hombre** +Secundaria vs -Secundaria : Son 0.77, es decir los odds muestrales de que exista obesidad en personas con escolaridad mayor son 0.77 veces los odds que con escolaridad de primaria, secundaria o sin escolaridad.
- **Odds Mujeres** +Secundaria vs -Secundaria: Son 1.42, es decir los odds muestrales de que exista obesidad en personas con escolaridad mayor son 1.42 veces los odds que con escolaridad de primaria, secundaria o sin escolaridad.

#### 5. Odds ratios de obesidad según el Nivel Socioeconómico.

Hombre				Mujer			
Nivel Socioeconómico	SI	NO	Total	Nivel Socioeconómico	SI	NO	Total
Alto	1,731	3,379	5,110	Alto	2,712	4,163	6,875
Bajo	1,204	4,447	5,651	Bajo	2,624	4,816	7,440
Medio	1,472	3,729	5,201	Medio	2,983	4,320	7,303
<b>Total</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>	<b>Total</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

- El porcentaje de obesidad incrementa a mayor nivel socioeconómico en el caso de los hombres de un 21% en niveles bajos a un 33% en nivel alto.
- El porcentaje de obesidad en el caso de las mujeres se acentúa en un nivel socioeconómico medio con un 41%.

6. Odds ratios de obesidad según la Afiliación.

<b>Hombre</b>				<b>Mujer</b>			
<b>AFILIACIÓN</b>	<b>SI</b>	<b>NO</b>	<b>TODOS</b>	<b>AFILIACIÓN</b>	<b>SI</b>	<b>NO</b>	<b>Todas</b>
DEFENSA/MARINA	15	32	47	DEFENSA/MARINA	29	44	73
IMSS	1,363	3,051	4,414	IMSS	2,313	3,439	5,752
ISSSTE	455	906	1,361	ISSSTE	826	1,158	1,984
NA	5	23	28	NA	4	14	18
Ninguno	938	2,735	3,673	Ninguno	1,378	2,335	3,713
OTRO	25	46	71	OTRO	44	46	90
PEMEX	24	26	50	PEMEX	29	38	67
PRIVADO	25	26	51	PRIVADO	24	35	59
SEGURO POPULAR	1,557	4,710	6,267	SEGURO POPULAR	3,672	6,190	9,862
<b>Todos</b>	<b>4,407</b>	<b>11,555</b>	<b>15,962</b>	<b>Todos</b>	<b>8,319</b>	<b>13,299</b>	<b>21,618</b>

\*Los afiliados a PEMEX, Seguros Privados y Otros presentan mayores niveles de obesidad con un porcentaje mayor a un 40%.

#### 4.3.1.3 Regresión Logística

Para establecer la prevalencia de la obesidad medida a través de factores como sexo, edad, área (rural o urbana), región geográfica, escolaridad, nivel socioeconómico y afiliación a servicios de salud, se procederá a las regresiones logísticas incluyendo y omitiendo variables para observar cual es el modelo que mejor explica a la obesidad.

Como primer paso se corre la regresión considerando todas las variables anteriores, sin embargo dentro del apartado del análisis de la bondad de ajuste, se muestra que los p-valores asociados a los contrastes de Pearson, Hosmer-Lemeshow son 0.00, ambos menores a 0.05 lo que nos dice que nuestro modelo no se encuentra bien explicado con las variables incluidas.

Por lo tanto se procederá con sacar del modelo las variables escolaridad y afiliación que parecen no ser significativas.

### Regresión Logística Binaria

**Variable Dependiente:** Obesidad

**Variable Independiente:** Sexo, Edad Actual, Área y Nivel Socioeconómico.

**Función Link:** LOGIT

Variable	Valor	Cuenta
Obesidad	1	12,726
No		
Obesidad	0	24,854
<b>Total</b>		<b>37,580</b>

### Tabla de Regresión Logística

Predictor	Coef	SE	Coef	Z	P-Ratio
Constant	-		1.315	0.053	- 24.770 -
Sexo	-		0.499	0.023	- 22.030 - 0.610
Edad			0.005	0.001	- 7.660 - 1.010
Area (Rural/Urbano)			0.233	0.026	- 9.040 - 1.260
Nivel socioeconómico (quintiles)			0.076	0.009	- 8.940 - 1.080

#### 95% CI

Predictor	Lower	Upper
Constant		
Sexo	0.58	0.63
Edad	1	1.01
Area (Rural/Urbano)	1.2	1.33
Nivel socioeconómico (quintiles)	1	1.1

Log-Likelihood = -23,647.454

Test that all slopes are zero:  $G = 817.086$ ,  $DF = 4$ ,  $P\text{-Value} = 0.000$

#### Prueba de Bondad de Ajuste

Method	Chi-Square	DF	P
Pearson	2,559.65	1,437	0.000
Deviance	2,827.01	1,437	0.000
Hosmer Lemeshow	59.67	8	0.000

#### Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

VALUE Group	1		0		Total
	Observed	Expected	Observed	Expected	
1	746	843	3,035	2,938	3,781
2	970	979	2,813	2,804	3,783
3	1,058	1,090	2,728	2,696	3,786
4	1,225	1,159	2,548	2,614	3,773
5	1,385	1,235	2,416	2,566	3,801
6	1,313	1,310	2,462	2,465	3,775
7	1,457	1,412	2,326	2,371	3,783
8	1,496	1,511	2,289	2,274	3,785
9	1,560	1,586	2,200	2,174	3,760
10	1,516	1,601	2,037	1,952	3,553

### Measures of Association:

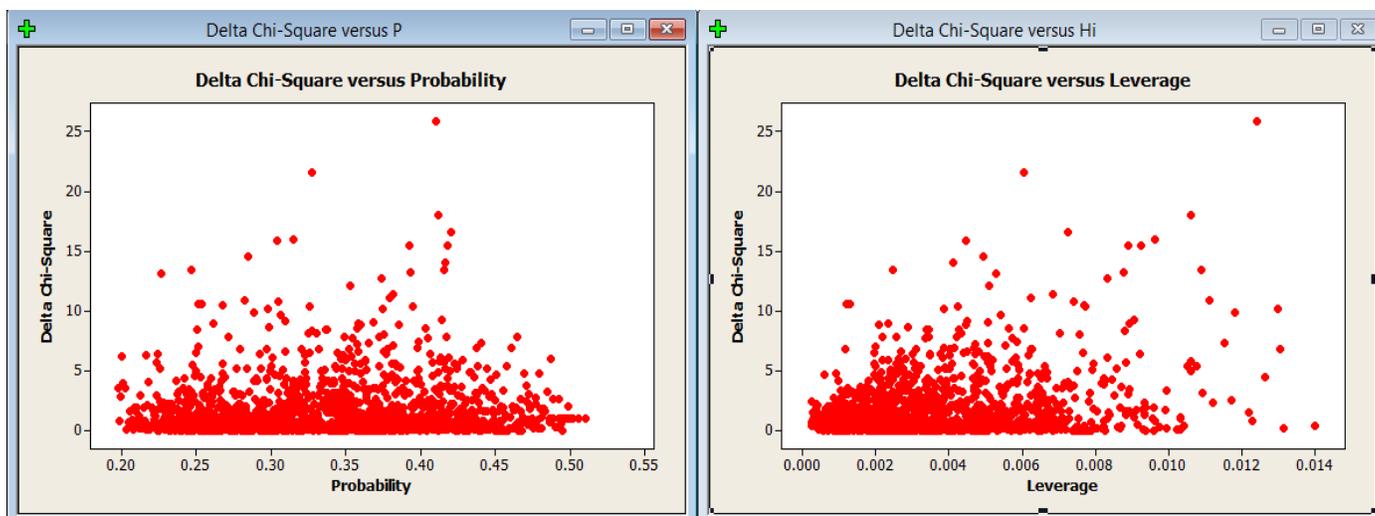
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent
Concordant	185,296,012	58.6%
Discordant	129,512,141	40.9%
Ties	1,483,851	0.5%
<b>Total</b>	<b>316,292,004</b>	<b>100.0%</b>

### Summary Measures

<b>Somers' D</b>	<b>0.18</b>
<b>Goodman-Kruskal Gamma</b>	<b>0.18</b>
<b>Kendall's Tau-a</b>	<b>0.08</b>

Gráfico 12. Delta Chi Square (Modelo II)



\*Fuente: Elaboración Propia / Minitab

### Interpretación:

La tabla de regresión logística muestra los valores estimados para los coeficientes del modelo ( $\beta_0 = -1.3$ ,  $\beta_1 = -0.49$ ,  $\beta_2 = 0.5$ ,  $\beta_3 = 0.005$ ,  $\beta_4 = 0.23$ ,  $\beta_5 = 0.07$  junto con sus p-valores asociados (0.00 respectivamente). Por lo tanto se pueden interpretar a los coeficientes como el cambio que se produce en el término Logit al incrementarse en una unidad la variable explicativa asociada. Cuando se usó la regresión logística, también se obtuvo un odds-ratio mayores a 1 (a excepción de sexo) lo que nos indica

que un incremento de 1 unidad en nuestras variables, incrementa los odds de obesidad en cuando menos un 1% hasta un 26%.

Se observa que los p-valores asociados a los coeficientes  $\beta$ s son inferiores a 0,05. Por tanto, para un nivel de significación  $\alpha = 0,05$ , se rechaza la hipótesis nula de que dichos coeficientes son nulos (i.e.: que la variable asociada a los mismos no es relevante en el modelo).

El estadístico G sirve para contrastar la hipótesis nula de que todos los coeficientes asociados con variables explicativas son nulos. Dado que el p-valor obtenido es de 0.00, se puede rechazar dicha hipótesis nula y concluir que, como mínimo, uno de los coeficientes será distinto de cero.

El apartado Goodness-of-Fit Tests muestra los p-valores asociados a los contrastes de Pearson, Hosmer-Lemeshow son 0.00, ambos menores a 0.05 lo que nos dice que nuestro modelo se encuentra bien explicado con las variables incluidas.

#### 4.3.1.4 Tabla de clasificación

Si bien los coeficientes de bondad de ajuste nos indican falta de ajuste, se analizará la tabla de clasificación, que es normalmente el criterio que se debe de seguir para indicar la bondad de ajuste del modelo. En esta tabla se muestran los casos bien clasificados en la diagonal principal, y los casos mal clasificados en la segunda diagonal.

**Tabla 4. Tablas de Clasificación**

Observado			Pronosticado		
			obesidad		Porcentaj e correcto
			.00	1.00	
Paso 1	Obesidad	.00	24,851	3	100.0
		1.00	12,724	2	.0
Porcentaje global					66.1

a El valor de corte es .500

\*Fuente: Elaboración Propia / Minitab

El porcentaje global de aciertos nos muestra un 66%, lo que nos indica que de los valores pronosticados el 66% de ellos fue correcto.

Finalmente, en los gráficos se observa la existencia de una observación que no es bien explicada por el modelo.

Para determinar si el modelo puede ser simplificado, se observa que el P-valor más alto para las pruebas de probabilidad es 0.000, perteneciendo a todas las variables incluidas, este P-valor es menor de 0.05 y es un estadístico significativo a un nivel del 95.0%. Por lo tanto, no se debe remover ninguna variable del modelo.

Por ser el P-valor del modelo menor de 0.05, hay una relación estadística significativa entre las variables a un nivel de confianza del 95.0%.

Por lo tanto, la ecuación del modelo final es:

$$\text{OBESIDAD} = \exp(\eta) / (1 + \exp(\eta))$$

Dónde:

$\eta = -1.3 + 0.005 * (\text{Edad}) - 0.49 * (\text{Sexo}) + 0.23 * (\text{Área Rural o Urbana}) + 0.07 * (\text{Nivel Socioeconómico})$

# CONCLUSIONES



En conclusión, analizando de manera general los datos, se observó que la obesidad en las mujeres es mayor en un 30% y los grupos de edad con mayor índices de obesidad son entre los 30 y los 50 años de edad (alrededor de un 50% de la obesidad se concentra en estos grupos de edad).

La tabla de frecuencias y el cociente de momios, nos muestra lo siguiente:

- Se encontró una prevalencia de obesidad de 34% para el conjunto de la población.
- No se encontraron diferencias significativas entre las áreas urbana y rural.
- Las mujeres presentan mayores prevalencias de obesidad (38%) que los hombres (28%).
- Por grupos de edad, la obesidad es mayor en edades de los 40 a los 60 años.
- No se encontraron diferencias significativas entre las regiones del país. El porcentaje de obesidad en el caso de las mujeres se acentúa en el norte con un 44% y el % de obesidad en el caso de los hombres se acentúa en el norte con un 32%
- Se presenta un mayor índice de obesidad en las personas con nivel educativo de primaria y secundaria (36%)
- A mayor nivel socioeconómico se presentan mayores índices de obesidad (sólo para el caso de las mujeres)
- Los afiliados a PEMEX, Seguros Privados y Otros presentan mayores niveles de obesidad con un porcentaje mayor a un 40%.

En conclusión y observando el modelo, se puede ver que la edad es un factor que afecta de manera positiva a la obesidad, a mayor edad mayor incremento en la probabilidad de padecer obesidad.

Para el caso del sexo se observa que los hombres (marcados como 1) disminuyen la probabilidad de padecer obesidad. En cuanto al área rural o Urbana, se puede observar que se guarda una relación positiva lo que nos indica que incrementa también la probabilidad de padecer obesidad cuando el área es urbana y finalmente el nivel socioeconómico también guarda una relación positiva (a mayor nivel socioeconómica la afectación en la variable obesidad es mayor).

## BIBLIOGRAFÍA

Academy of Management (2016), Regression Models. Recuperado de: <[http://www.aom.pace.edu/rmd/1997\\_forum\\_regression\\_models.html](http://www.aom.pace.edu/rmd/1997_forum_regression_models.html)>

Agresti, A., (2010), Analysis of Ordinal Categorical Data, 2nd Edition. John Wiley and Sons, Inc.

Agresti, A., (2013), Categorical Data Analysis, 3rd Edition, Wiley Series in Probability and Statistics.

Alonso A, Álvaro-Gracia JM, Andreu JL et al. Manual S.E.R. de las enfermedades reumáticas. Sociedad Española de Reumatología. 3ª ed. 2000. p. 464

Artís, M.; Suriñach, J.; et al (2002): "Econometría". Ed. Fundación para la Universidad Oberta de Catalunya, Barcelona.

Charles F. Manski and Daniel L. McFadden (1981): Structural Analysis of Discrete Data and Econometric Applications. Ed. Cambridge: The MIT Press.

Cuernavaca, Mexico: Instituto Nacional de Salud Pública (MX), 2012.

Daniel A. Powers and Yu Xie, Statistical Methods for Categorical Data Analysis, 2nd Ed.

Departamento de Ciencias Clínicas. Universidad de Las Palmas de Gran Canaria. España.

Departamento de Química Analítica, Nutrición y Bromatología. Universidad de La Laguna. Santa Cruz de Tenerife.

11. Díaz C, Henríquez P, Rodríguez E, Ribas L, Serra-Majem L: Dietary magnesium intakes and serum magnesium concentrations in the Canarian population. En: Advances in Magnesium: pharmacology, metabolism and nutrition. Capítulo 10. Escanero JF, Alda JO, Guerra M, Durlach J (eds.). Prensas Universitarias de Zaragoza, Zaragoza, 2003: 87-91.

Díaz Romero C, Henríquez Sánchez P, López Blanco F, Rodríguez Rodríguez E, Serra Majem L: Concentraciones de Na,K, Ca y P en sueros de una muestra representativa de la población canaria. *Nutr Hosp* 2002, 17:204-212.

Díaz Romero C, Henríquez Sánchez P, López Blanco F, Rodríguez Rodríguez E, Serra Majem L: Serum copper and zinc concentration in a representative sample of the Canarian population. *J Trace Elem Med Biol* 2002, 16:75-81.

Díaz Romero C, López Blanco F, Henríquez Sánchez P, Rodríguez Rodríguez E, Serra Majem L: Serum manganese concentration in a representative sample of the Canarian population. *Biol Trace Elem Res* 2001, 80:43-52.

Díaz Romero C, López Blanco F, Henríquez Sánchez P, Rodríguez Rodríguez E, Serra Majem L: Serum selenium concentration in a representative sample of the Canarian population. *Sci Total Environ* 2001, 269:65-73.

Díaz-Flores Estévez JF, Díaz-Flores Estévez F, Hernández Calzadilla C, Rodríguez Rodríguez EM, Díaz Romero C, Serra-Majem L: Application of linear discriminant analysis to the biochemical and haematological differentiation of opiate addicts from healthy subjects. A case control study. *Eur J Clin Nutr* 2003, 58:449-455.

Doran, H. (1989): "Applied Regression Analysis in Econometrics". Ed. Marcel Dekker, Inc.

Doran, H. (1989): "Applied Regression Analysis in Econometrics". Ed. Marcel Dekker, Inc.

Doug McKee (2015), An Intuitive Introduction to Multinomial Logit

E. Rodríguez Rodríguez\*, P. Henríquez Sánchez\*\*, F. López Blanco\*\*, C. Díaz Romero\* y L. Serra

Econometric at the University of Illinois (2016), Econometrics Source. Recuperado de:<http://www.econ.uiuc.edu/~morillo/links.html>

ENSANUT 2012, Recuperado de: [ensanut.insp.mx](http://ensanut.insp.mx)

ENSANUT, 2012. Recuperado de: [ensanut.insp.mx](http://ensanut.insp.mx)

Ferrer MC, Díaz Cremades J, Soria López A, Hernández Briz MJ: Encuesta nutricional de Canarias 1997-1998. Evaluación bioquímica del estado nutricional. Volumen 4, Servicio Canario de Salud, Litografía A. Romero, S.A., Tenerife, 2000.

Frone, M.R. (1997): "Regression Models for Discrete and Limited Dependent Variables"

Frone, M.R. (1997): "Regression Models for Discrete and Limited Dependent Variables".

Gujarati, D. (1997): "Econometría básica". McGraw-Hill.

Gujarati, D. (1997): "Econometría básica". McGraw-Hill. ISBN 958-600-585-2

Gutierrez JP, Rivera-Dommarco J, Shamah-Levy T, Villalpando-Hernandez S, Franco A, Cuevas-Nasu L,

Gutiérrez JP, Rivera-Dommarco J, Shamah-Levy T, Villalpando-Hernández S, Franco A, Cuevas-Nasu L.

Handbook of Econometrics (2016), Elsevier. Recuperado de: <<http://www.elsevier.com/hes/books/02/menu02.htm>>

Henríquez Sánchez P, Díaz Romero C, Rodríguez Rodríguez E, López Blanco F, Álvarez León E, Pastor Ferrer MC, Serra Majem L: Evaluación bioquímica del estado nutricional de la población canaria. Arch Latinoam Nutr 2000, 50:43-54.

Instituto Nacional de Salud Pública (MX), Cuernavaca, México, 2012.

Instituto Nacional del Cáncer de los Institutos Nacionales de la Salud de EE. UU, (2015). Recuperado de: <http://www.cancer.gov/espanol/cancer/que-es>

Instituto Nacional del Cáncer, (2015), Recuperado de: <http://www.cancer.gov/espanol/cancer/que-es>

Instituto tecnológico de Chihuahua

Instituto tecnológico de Chihuahua (2016)

International Diabetes Federation

International Diabetes Federation (2015), Recuperado de:  
<http://www.idf.org/?language=es>

Iyengar V, Woittiez J: Trace elements in human clinical specimens. Evaluation of literature data to identify reference values. Clin Chem 1988, 34:474-481.

Johnston, J. (2001): "Métodos de econometría". Ed. Vicens Vives. Barcelona.

Johnston, J. (2001): "Métodos de econometría". Ed. Vicens Vives. Barcelona.

Kennedy, P. (1998): "A Guide to Econometrics". Ed. MIT Press.

Kennedy, P. (1998): "A Guide to Econometrics". Ed. MIT Press. ISBN: 0262611406

Log-Linear, Logit, and Probit Models. Recuperado de:  
<<http://www2.chass.ncsu.edu/garson/pa765/logit.htm>>

McKee Doug, (2015), An intuitive introduction to Multinomial Logit

Modelos lineales generalizados (GLM) – Luis Cayuela –Universidad de

Modelos lineales generalizados, en seguros – Monserrat Guillen y Catalina Bolancé  
– Universidad de Barcelona

Monserrat Guillen y Catalina Bolancé, (2016), Universidad de Barcelona Modelos  
lineales generalizados, en seguros. Recuperado de:  
<http://www.uv.es/lejarza/mcaf/glm.pdf>

Novales, A. (1993): "Econometría". McGraw-Hill.

Novales, A. (1993): "Econometría". McGraw-Hill. ISBN 84-481-0128-6

Novartis Oncology (2016), Endocrine Diseases. Recuperado de:  
<<http://www.novartis oncology.com/es/for-patients-and-caregivers/endocrine-diseases.jsp>>

OMS: Organización Mundial de la Salud

Online Resources for Econometric Students. Recuperado por:  
<[http://www.oswego.edu/~kane/econometrics/stud\\_resources.htm](http://www.oswego.edu/~kane/econometrics/stud_resources.htm)>

Organización Mundial de la Salud (2016), Publicaciones de la Organización Mundial de la Salud. Recuperado de: <<http://www.who.int/es/>>

Pajuelo J. Estado nutricional del adulto en el Perú. Acta Médica Peruana 1992; 16: 22-32.

Pajuelo J. Estado nutricional del adulto en el Perú. Acta Médica Peruana 1992; 16: 22-32.

Pajuelo J. La obesidad en el Perú. Cuadernos de Nueva Perspectiva. Alimentación y Nutrición, Número 1. Lima 1997

Pajuelo J. La obesidad en el Perú. Cuadernos de Nueva Perspectiva. Alimentación y Nutrición, Número 1. Lima 1997

Pontificia Universidad Católica de Valparaíso (2016), Estadística. Recuperado de: <[http://www.ucv.cl/web/estadistica/cb\\_esc\\_medic.htm](http://www.ucv.cl/web/estadistica/cb_esc_medic.htm)>

Pontificia Universidad Católica de Valparaíso, Recuperado de: [http://www.ucv.cl/web/estadistica/cb\\_esc\\_medic.htm](http://www.ucv.cl/web/estadistica/cb_esc_medic.htm)

Pulido, A. (2001): "Modelos econométricos". Ed. Pirámide. Madrid.

Pulido, A. (2001): "Modelos econométricos". Ed. Pirámide. Madrid. ISBN 84-368-1534-3

Romero-Martinez M, Hernandez-Avila M. Encuesta Nacional de Salud y Nutrición 2012. Resultados Nacionales.

Romero-Martínez M, Hernández-Ávila M. Encuesta Nacional de Salud y Nutrición 2012. Resultados Nacionales.

Secretaría de Salud (México)

Secretaría de Salud (México)

Serra Majem L, Armas Navarro A, Ribas Barba L, en nombre del equipo investigador de ENCA (1997-98): Encuesta nutricional de Canarias 1997-1998. Hábitos alimentarios y consumo de alimentos. Volumen 1, Servicio Canario de Salud, Litografía A. Romero, S.A., Tenerife, 1999

Serra Majem L, Henríquez Sánchez P, López Blanco F, Álvarez León EE, Díaz Romero C, Rodríguez Rodríguez E, Pastor

SPSS View (2016), Recuperado de: <[http://www.ub.edu/aplica\\_infor/spss/cap4-1.htm](http://www.ub.edu/aplica_infor/spss/cap4-1.htm)>

Tang, W., He, H., Tu, X.M.,(2012), Applied Categorical and Count Data Analysis, CRC Press

Tang, W., He, H., Tu, X.M.,(2012), Applied Categorical and Count Data Analysis, CRC Press

The Econometrics Journal On-Line, Recuperado de:  
<http://www.feweb.vu.nl/econometriclinks/index.html>

Universidad de Granada (2016), Modelos lineales generalizados (GLM), Luis Cayuela. Recuperado de: <https://www.ugr.es/>

Universidad de Valencia (2016), Distribuciones e Inferencia para datos categóricos. Recuperado de:<<http://www.uv.es/ayala/docencia/cda/t1Distribucioneseinferenciaparadatoscategoricos.pdf>>

University of Texas (2016), Statistical Methods for Categorical Data Analysis  
Recuperado de:

<<http://www.la.utexas.edu/research/faculty/dpowers/book/htmlbook/content.html>>

Uriel, E. (1990): "Econometría: el modelo lineal". Ed. AC. Madrid.

Uriel, E. (1990): "Econometría: el modelo lineal". Ed. AC. Madrid. ISBN 84-7288-150-4.

Wooldridge, J. (2001): "Introducción a la Econometría: un enfoque moderno". Ed. Thomson Learning.