Universidad Nacional Autónoma de Nicaragua – León
UNAN - León

Facultad de Ciencias y Tecnología

Departamento de Matemática y Estadística

TESIS

Para optar al título de:

Licenciada en Estadística

Aplicación del Método de Regresión Logística Multinomial en un estudio de colesterol en pacientes que asistieron al laboratorio de Bioquímica de la facultad de Ciencias Médicas de la UNAN León, en el año 2000.

Autores:

Bra. Valeria Luciana Cortés Calero
Bra. Angela Magdalena Hernández Rueda

Tutor:

Msc. Rafael Espinoza Montenegro

León, Noviembre de 2009

AGRADECIMIENTOS

Agradecemos eternamente a nuestro Dios, que nos ha guiado a lo largo de las dificultades, triunfos, fracasos y sobre todo por el habernos hecho posible el coronar nuestra carrera.

A nuestros padres, que sin su apoyo incondicional no hubiese sido posible la culminación de nuestra preparación profesional.

Reconocemos especialmente el apoyo por nuestro tutor y maestro Msc. Rafael Espinoza Montenegro que con su valiosa colaboración nos ha sabido orientar a lo largo de este difícil trabajo.

Al doctor Efrén Castellón por haber colaborado con la fuente de información.

DEDICATRORIA

Dedico de manera muy especial este trabajo a esa persona que supo creer en mí, por su apoyo incondicional, por su amor inquebrantable, por sus desvelos, por el tiempo que ha invertido en mi educación y por sus sabios consejos a mi madre Victoria Virginia Rueda.

Con mucho cariño al Lic. Felipe Martínez Juárez por haber sido mi maestro, mi compañero y un amigo incondicional.

ANGELA MAGDALENA HERNANDEZ RUEDA

DEDICATORIA

A mis padres Yasmina Calero y Raúl Cortez que me dieron la vida, su apoyo incondicional en el transcurso de todos mis estudios, se esforzaron siempre por brindarme una mejor vida y la mejor preparación.

A mis hermanas Laura, Jazmina y Carla que siempre han sido amigas incondicionales.

A mi amigo incondicional Kelvin Quiroz

VALERIA LUCIANA CORTEZ CALERO

INDICE

1. INTRODUCCION

En estudios estadísticos se pueden encontrar diferentes tipos de datos, en donde surgen variables de distintas naturaleza. Para el análisis de datos médicos referentes al área de epidemiología y ciencias sociales la regresión logística es un instrumento estadístico frecuentemente utilizado, sin embargo su limitación se presenta cuando la variable respuesta está compuesta por más de dos categorías, de allí la necesidad de implementar el método de Regresión Logística Multinomial (RLM) el cual se utiliza cuando la variable respuesta en estudio es nominal y consiste de más de dos categorías de respuestas.

Lo que se pretende con este trabajo es dar a conocer bajo que condiciones se aplica esta técnica, sus fundamentos teóricos – prácticos, así como la interpretación de las salidas proporcionadas por un paquete estadístico.

La aplicación del método de RLM se realizará en base a un estudio de colesterol realizado en el laboratorio de Bioquímica por el Dr. Efrén Castellón, jefe de departamento de la facultad de Ciencias Médicas de la UNAN León, en el año 2000. Debido a que las variables dependientes se caracterizan por ser variables categóricas con más de dos niveles de respuestas se aplicará la RLM con el objetivo de determinar la influencia de las variables edad, hábito de fumado, práctica de ejercicio e índice de masa corporal (IMC), en los niveles del Colesterol total y colesterol malo (ldl). Así mismo se determinará la probabilidad para cada uno de los niveles de riesgo obtenidos del Colesterol total y Ldl condicionadas a un determinado perfil de las variables influyentes.

Como consecuencia de que es muy poco el domino y aplicación de esta técnica este trabajo permitirá servir de apoyo para futuros estudios que no solamente sean dirigidos al área de la salud sino también en temas de gran importancia como: económicos, políticos, sociales, naturales, etc. en donde surja la necesidad de aplicar la técnica de RLM.

2. OBJETIVOS

Objetivo General:

- ✓ Aplicar el método de regresión logística multinomial en un estudio de colesterol realizado a pacientes que asistieron al laboratorio de Bioquímica de la facultad de Ciencias Médicas de la UNAN León, en el año 2000.

Objetivos Específicos:

- ✓ Definir los modelos logísticos multinomiales para el Colesterol total y colesterol malo (ldl) en función de las variables involucradas en el estudio.
- ✓ Describir las variables edad, hábito de fumado, practica de ejercicio e índice de masa corporal involucradas en el estudio.
- ✓ Estimar los parámetros de cada modelo logístico multinomial definido.
- ✓ Construir los modelos de regresión logística multinomial para cada una de las k-1 categorías de la variable respuesta.
- ✓ Interpretar el exponencial de los parámetros estimados en cada modelo (Odds Ratio).
- ✓ Estimar las probabilidades de cada nivel de riesgo del Colesterol total y ldl condicionadas a ciertas características de la persona en estudio.

3. MARCO TEÓRICO

Los modelos logit para respuestas multinomiales son usados cuando la variable respuesta en estudio consiste en más de dos categorías, estos modelos describen efectos de un conjunto de variables explicativas sobre una variable respuesta. A diferencia de otros modelos (como los modelos loglineales) usados para variables categóricas, los modelos logit no describen modelos de asociación e interacción entre variables explicativas.

En el estudio de modelos logit multinomiales se distinguen dos casos, dependiendo de la naturaleza de la variable respuesta. El primero la formulación de modelo para respuestas nominales en los cuales se usa un modelo logit binario separado para cada par de categoría de respuesta y el segundo, modelo para respuestas ordinales que usan logit de probabilidades respuestas acumulativa u otra función índice para estas probabilidades acumulativas.

En este trabajo explicaremos específicamente el caso para la formulación de modelo para respuestas nominales.

3.1 **Modelo Logit Multinomial**

3.1.1 Distribución Multinomial

Supongamos una muestra aleatoria simple de tamaño fijo n que es de clasificación cruzada según las variables categóricas. Una distribución previa del conteo de celda es entonces la distribución multinomial especificada por las r*c probabilidades de celdas $\{\pi_j\}$. La función de probabilidad multinomial es

$$P\left(x_1, x_2, \dots, x_J\right) = \left(\frac{n!}{x_1!\ x_2!\ \dots\ x_J!}\right)\pi_1^{x_1}\pi_2^{x_2}\ \dots\ \pi_J^{x_J}$$

3.1.2 Respuestas Nominales: Modelos Logit con categorías de referencia

Sea Y una respuesta categórica con J categorías. Los modelos logit multicategóricos (también llamados Politómicos) para variables respuestas nominal, describen simultáneamente el log odds para todos los pares de categorías $\binom{J}{2}$. Dado una cierta elección de J-1 de éstos, el resto son redundantes.

3.1.3 Modelo Logit para categoría de línea base

Trataremos los conteos en las categorías de la variable Y como multinomiales con probabilidades $\{\pi_{1(x)}, \dots , \pi_{j\ (x)}\}$, en donde π_j (x) = P (Y= j/x), en un punto fijo x para las variables explicativas, denota la probabilidad de que la respuesta caiga en la categoría j de la variable dependiente, con $\sum_j \pi_i$ (x) = 1.

El simple acercamiento a datos multinomiales es para designar una de las categorías como una línea base o de referencia, calcular el log del odds para todas las otras categorías relativas a la línea base, y luego dejar que el log del odds sea una función lineal de las variables explicativas, a menudo se escoge la última categoría como la línea base. El modelo

$$\log \frac{\pi_{j\,(x)}}{\pi_{J\,(x)}} = \propto_j + \beta_j' X\ ,\quad \text{j} = 1, \dots , \text{J} - 1 \qquad (2.1)$$

describe simultáneamente los efectos de x sobre esos J–1 logits. Los efectos varían de acuerdo al par de respuesta con la línea base. Estas J -1 ecuaciones determinan parámetros para logits con otros pares de categorías de respuesta, ya que

$$\log \frac{\pi_{a\ (x)}}{\pi_{b\ (x)}} = \log \frac{\pi_{a\ (x)}}{\pi_{J\ (x)}} - \log \frac{\pi_{b\ (x)}}{\pi_{J\ (x)}}$$

3.1.4 Estimación de Probabilidades Respuestas

La ecuación que expresa los modelos logit multinomiales directamente en términos de probabilidades respuestas $\{\pi_{j\ (x)}\}$ es

$$\pi_{j\ (x)} = \frac{\exp\left(\alpha_j + \beta_j^{'}\ x\right)}{1+\sum_{h=1}^{J-1}\exp\left(\alpha_h + \beta_h^{'}\ x\right)} \qquad (2.2)$$

con $\alpha_J = 0$ y $\beta_J = 0$ esto se sigue de (2.1), usando el hecho que (2.1) también se cumple j = J haciendo $\alpha_j = 0$ y $\beta_j = 0$. El denominador de (2.2) es el mismo para cada j. Los numeradores para varios j suman el denominador, entonces $\sum_j \pi_{j\ (x)} = 1$. Para J = 2 (2.2) se simplifica a la formula usada para regresión logística binaria.

Se puede demostrar que a través de las probabilidades respuesta estimadas por la formula (2.2) y cumpliéndose con $\alpha_J = 0$ y $\beta_J = 0$ se llega al modelo (2.1) de la siguiente manera:

$$\log\frac{\pi_j(x)}{\pi_J(x)} = \log\frac{\exp\left(\alpha_j + \beta_j' x\right)/1+\sum_{h=1}^{J-1}\exp\left(\alpha_h + \beta_h' x\right)}{\exp\left(\alpha_J + \beta_J' x\right)/1+\sum_{h=1}^{J-1}\exp\left(\alpha_h + \beta_h' x\right)}$$

$$= \log\frac{e^{\left(\alpha_j + \beta_j^{'}\ x\right)}}{e^{\left(\alpha_J + \beta_J^{'}\ x\right)}}$$

$$= \log e^{(\alpha_j - \alpha_J) + (\beta_j^{'} - \beta_J^{'})\ x}$$

$$= \alpha_j + \beta_j' x$$

3.1.5 Ajuste del Modelo Logit para Categoría de Línea Base

El ajuste de máxima verosimilitud de modelos logit multinomial maximiza la probabilidad sujeta a $\{\pi_{j\ (x)}\}$ satisfaciendo simultáneamente las J – 1 ecuaciones que especifica el modelo. Para i = 1,..., n, $y_i = (y_{i1}, \ldots, y_{iJ})$ representa el proceso multinomial para el sujeto i, donde $y_{ij} = 1$, cuando la respuesta está en la categoría j y $y_{ij} = 0$ en otro caso. Así $\sum_j y_{ij} = 1$. Sea $x_i = (x_{i1}, \ldots, x_{ip})$' denota los valores de la

variable explicativa para el sujeto i. Sea $\beta_j = (\beta_{j1}, \dots, \beta_{jp})'$ denota los parámetros para el j-ésimo logit.

Asumamos n observaciones independientes, entonces el logaritmo de la función de verosimilitud es:

$$\log \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_j \ (x_i)^{y_{ij}}\right]$$

$$= \log \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \left(\frac{\exp(\alpha_j + \beta_j')}{1 + \sum_{h=1}^{J-1} exp\,(\alpha_h + \beta_h' x)}\right)^{y_{ij}}\right]$$

$$= \sum_{i=1}^{n} \{ \sum_{j=1}^{J-1} y_{ij} \ (\alpha_j + \beta_j' x_i) - \log\left[1 + \sum_{j=1}^{J-1} exp\,(\alpha_j + \beta_j' x_i)\right]$$

$$= \sum_{j=1}^{J-1} \left[\alpha_j \left(\sum_{i=1}^{n} y_{ij}\right) + \sum_{i=1}^{n} y_{ij} \left(\beta_j' \ x_i\right)\right] - \sum_{i=1}^{n} \log\left[1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \beta_j' \ x_i)\right]$$

Teniendo en cuenta que:

$$\sum_{i=1}^{n} y_{ij} \ (\beta_j' \ x_i) = \sum_{i=1}^{n} y_{ij} \ \sum_{k=1}^{p} \beta_{jk} \ x_{ik} = \sum_{k=1}^{p} \beta_{jk} \left(\sum_{i=1}^{n} x_{ik} \ y_{ij}\right)$$

Entonces:

$$= \sum_{j=1}^{J-1} \left[\alpha_j \left(\sum_{i=1}^{n} y_{ij}\right) + \sum_{k=1}^{p} \beta_{jk} \left(\sum_{i=1}^{n} x_{ik} y_{ij}\right)\right] -$$

$$\sum_{i=1}^{n} log\left[1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \beta_j' x_i)\right]$$

Para maximizar la función de verosimilitud se aplica el logaritmo neperiano a la función, luego se deriva parcialmente con respecto al parámetro a estimar y se iguala a cero, obteniéndose un sistema de ecuaciones que se resolverán por el método aproximativo de Newton-Raphson.

El método Newton- Raphson produce los parámetros estimados de máxima verosimilitud. Sus errores estándares aproximados son raíces cuadradas de los elementos diagonales de la inversa de la matriz de la información.

Muchos software estadísticos pueden ajustar modelos logit multinomial, pero algunos pueden ajustar solamente modelos de regresión logística binaria. Una alternativa de ajuste es ajustar modelos logit binarios separadamente para las J – 1 pares de respuestas: el modelo (2.1) solo para j = 1, usando solamente observaciones en la categoría 1 y J de la variable respuesta para obtener estimaciones de α_1 y β_1; usando solamente las categorías 2 y J para obtener estimaciones de α_2 y β_2 ; de esta manera obtener J – 1 ajuste separados de modelos logit. Un modelo logit ajustado usando solamente dos categorías de respuesta es el mismo como un modelo logit regular ajustado, condicional a la clasificación en una de esas categorías. Por ejemplo el j – ésimo logit de la categoría de referencia es un logit de probabilidad condicional

$$\log \frac{\pi_j\,(x)/(\pi_j\,(x)+\pi_J\,(x))}{\pi_J\,(x)/(\pi_j\,(x)+\pi_{J\,(x)})} = \log \frac{\pi_j\,(x)}{\pi_J\,(x)}$$

Las estimaciones de ajustes separados difieren de las estimaciones de máxima verosimilitud para ajustes simultáneos de J – 1 logits. Ellos son menos eficientes, tendiendo a tener más grandes errores estándar. Sin embargo, Begg y Gray (1984) mostraron que la pérdida de eficiencia es menor cuando la categoría de respuesta teniendo la más alta prevalencia es la de referencia.

4. DISEÑO METODOLOGICO

Los datos de este trabajo son secundarios ya que fueron proporcionados por el Dr. Efrén Castellón, quien realizó un estudio en la facultad de Ciencias Médicas en pacientes que asistieron al laboratorio de Bioquímica en el año 2000.

De la información recogida en el estudio solo fueron seleccionadas las variables colesterol malo (ldl) y colesterol total (colt1). Luego fueron seleccionadas aquellas variables explicativas que pueden ser factores de riesgo para las variables respuestas seleccionadas tales como: edad, habito de fumado, practica ejercicio e índice de masa corporal.

Operacionalización de las variables:

Variable	Etiqueta de la variable	Codificación
Edad	Edad	
Fuma	Habito de fumado	0: No fuma 1: Si fuma
Ejercicio	Practica ejercicio	0: no ejercicio 1: ejercicio
IMC	Índice de masa corporal	1:Obeso 2: Sobrepeso 3: Bajo peso 4: Normal
Colt1	Colesterol1	1: Deseable 2: Limite 3: Elevado
Ldl1	Niveles de riesgo	1: Riesgo bajo 2: Riesgo moderado 3: Riesgo elevado

Medición del riesgo

La medición del riesgo se realizó a través del Odds Ratio. Se considera factor de protección cuando el valor es menor de "1" y factor de riesgo cuando el valor es mayor a uno. En el caso del que valor fuera "1" está variable no esta influyendo en la variable respuesta.

Software

Para procesamiento de los datos y codificación de las variables se utilizó el paquete estadístico SPSS versión 15.0 y para el levantado de texto utilizamos Microsoft Word 2007.

5. RESULTADOS

Modelo 1.

Modelo de regresión multinomial para colesterol total (variable respuesta); ejercicio, edad e índice de masa corporal (variables explicativas).

$$\log \frac{\pi_j}{\pi_{Des}} = \alpha_j + \beta_1 Edad + \beta_2 Ejercicio + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3 \qquad j = 1,2$$

Donde: π_j = es la probabilidad de estar en el nivel j;

π_{Des} = es la probabilidad de estar en el nivel deseable.

Descripción de las variables en estudio

El promedio de las edades de las personas en estudio fue de 39 años.

En la tabla (1) encontramos que el 68.1% de las personas en estudio presentaron el colesterol en un nivel deseable, 21% en el nivel limite y el 11% en el nivel elevado.

El 30.9% de las personas en estudio practican ejercicio.

En la variable índice de masa corporal el 6.0% de las personas en estudio presentaron un bajo peso, el 34.4% un sobre peso y un 25.5% eran obesos.

Tabla Nº 1 Resumen de las variables

Variable	Niveles de la variable	N	Porcentaje
Colesterol1	Deseable	409	68.1%
	limite	126	21.0%
	elevado	66	11.0%
Practica ejercicio	no ejercicio	415	69.1%
	ejercicio	186	30.9%
Índice de masa corporal	Bajo peso	36	6.0%
	Normal	205	34.1%
	Sobre peso	207	34.4%
	Obeso	153	25.5%
Total		601	100.0%

Contraste global del modelo

Para realizar el contraste global del modelo se plantean las siguientes hipótesis

H_0: las variables explicativas no influyen significativamente en el modelo.

H_1: al menos una de las variables explicativas influye significativamente en el modelo.

Para contrastar las hipótesis se calculan dos modelos.

V1: Denota verosimilitud inicial, describe un modelo que no se somete a control las variables explicativas y simplemente se adapta una intercepción para predecir la variable resultado.

-2ln (V1)= 730.164

V2: Denota verosimilitud final, describe un modelo que incluye las variables explicativas especificadas.

-2ln (V2)= 660.800

El valor de ji-cuadrada con 10 grados de libertad es igual a 69.364 con un nivel de significancia de 0.000.

Comprobando este valor con el valor crítico de una distribución ji-cuadrada con 10 grados de libertad, rechazamos la hipótesis nula del modelo y concluimos que al menos uno de los coeficientes de regresión en el modelo es diferente de cero.

Estimación de los parámetros en el modelo (1)

En la estimación de los parámetros se usa el método Newton- Raphson que produce los parámetros estimados de máxima verosimilitud, en nuestro estudio se obtuvieron las siguientes estimaciones:

- **Para j =1(Limite)**

$$\log \frac{\pi_{\text{Lim}}}{\pi_{\text{Des}}} = \alpha_j + \beta_1 Edad + \beta_2 Ejercicio + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3$$

$$\hat{\alpha} = \text{-}2.677, \widehat{\beta_1} = 0.032, \widehat{\beta_2} = \text{-}0.110, \widehat{\beta_3} = 0.628, \widehat{\beta_4} = 0.303, \widehat{\beta_5} = \text{-}0.865$$

El ajuste del modelo nos permite apreciar a través del test de Wald que las variables edad (0.000) y el IMC obeso (0.026) resultaron significativas para el modelo que contrasta el colesterol a un nivel limite con relación al deseable y las variables ejercicio (0.635) y el IMC en un nivel de bajo peso (0.185) y sobrepeso (0.261) no resultaron significativas para el modelo.

Construcción del modelo

$$\log \frac{\pi_{\text{Lim}}}{\pi_{\text{Des}}} = \text{-}2.677 + 0.032 \text{ Edad} - 0.110 \text{ Ejercicio} + 0.628\, IMC_1 + 0.303 IMC_2 - 0.865 IMC_3$$

Interpretación del exponencial de los parámetros del modelo

La tabla 2 (ver Anexos) contiene una columna etiquetada Exp (B), en la cual se presentan los Odds Ratio (OR) estimados para las variables en estudio que son:

Para la variable edad de las personas en estudio el OR es igual a 1.033, lo cual indica que por cada año cumplido aumenta el riesgo de tener un colesterol al nivel límite en comparación con el nivel deseable en 0.033.

Para la variable IMC obeso el OR es igual a 1.874, lo que nos dice que las personas obesas tienen aproximadamente 2 veces mas riesgo que las personas con IMC normal de tener colesterol a un nivel limite en comparación con el nivel deseable.

- **Para j =2(Elevado)**

$$\log \frac{\pi_{\text{Ele}}}{\pi_{\text{Des}}} = \alpha_j + \beta_1 Edad + \beta_2 Ejercicio + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3$$

$$\hat{\alpha} = \text{-}4.220, \widehat{\beta_1} = 0.034, \widehat{\beta_2} = 0.327, \widehat{\beta_3} = 1.131, \widehat{\beta_4} = 1.073, \widehat{\beta_5} = \text{-}0.845$$

Las variables edad (0.000), el IMC obeso (0.006) e IMC sobre peso (0.006) resultaron significativas, para el modelo que contrasta el colesterol a un nivel elevado con relación al deseable y la variable ejercicio (0.313) y el IMC de bajo peso (0.436) no resultaron significativas.

Construcción de lo modelo

$$\log \frac{\pi_{\text{Ele}}}{\pi_{\text{Des}}} = \text{-}4.220 + 0.034\text{Edad} + 0.327\text{Ejercicio} + 1.131 IMC_1 + 1.073 IMC_2 \text{ - } 0.845\, IMC_3$$

Interpretación del exponencial de los parámetros del modelo

Para la variable edad de las personas en estudio el OR es igual a 1.035, lo cual indica que por cada año cumplido aumenta el riesgo de tener un colesterol al nivel elevado en comparación con el deseable en 0.035.

El IMC obeso presenta un OR igual a 3.099, esto nos indica que las personas obesas son 3 veces mas propensas de estar en el nivel elevado en comparación con el nivel deseable y para IMC de sobrepeso el OR es 2.925, lo que nos dice que estas personas son aproximadamente 3 veces más propensas que las de IMC normal de estar en el nivel elevado en comparación al deseable.

Estimación de las probabilidades para cada nivel de riesgo del Colesterol total

Las probabilidades respuestas para cada nivel del colesterol total se calculan a partir de:

$$\pi_{j\ (x)} = \frac{\exp\left(\alpha_j + \beta_j^{'}\ x\right)}{1 + \sum_{h=1}^{J-1} \exp\left(\alpha_h + \beta_h^{'}\ x\right)} \quad \text{para j =1,2 y 3}$$

Perfil Sano

Para una persona de 37 años, hace ejercicio y tiene peso normal las probabilidades respuestas son:

Niveles	Probabilidades
Deseable	0.78
Limite	0.18
Elevado	0.04

Perfil Medio

Para una persona de 45 años que no hace ejercicio y tiene bajo peso las probabilidades respuestas son:

Niveles	Probabilidades
Deseable	0.87
Limite	0.10
Elevado	0.04

Para una persona de 61 años, no hace ejercicio y tiene sobre peso las probabilidades respuestas son:

Niveles	Probabilidades
Deseable	0.48
Limite	0.29
Elevado	0.23

Perfil Enfermo

Para una persona que tiene 64 años, no hace ejercicio y es obeso las probabilidades respuestas estimadas son:

Niveles	Probabilidades
Deseable	0.40
Limite	0.37
Elevado	0.22

Modelo 2.

Modelo de regresión multinomial para colesterol malo (variable respuesta); ejercicio, habito de fumado e índice de masa corporal (variables explicativas).

$$\log \frac{\pi_j}{\pi_{Baj}} = \alpha_j + \beta_1 Ejercicio + \beta_2 Fuma + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3 \quad j = 1,2$$

Donde: π_j = es la probabilidad de estar en el riesgo j;

π_{Baj} = es la probabilidad de estar en el riesgo bajo.

Descripción de las variables en estudio

La tabla (3) nos describe el porcentaje que representan las personas en estudios, presentaron los niveles de riesgo de la siguiente manera: 64.9 % en el riesgo bajo, 21.1% en el riesgo moderado y el 14% en el riesgo elevado.

El 30.9% de las personas en estudio practicaban ejercicio.
El 87.4% de las personas en estudio no fumaban.

En la variable índice de masa corporal el 6% de las personas tuvieron un bajo peso, el 34.1% un peso normal, un 34.4% presentaron sobrepeso y el 25.5% eran obesos.

Tabla Nº 3 Resumen de la variables

Variable	Niveles de la variable	N	Porcentaje
Niveles de riesgo	riesgo bajo	390	64,9%
	riesgo moderado	127	21,1%
	riesgo elevado	84	14,0%
Practica ejercicio	no ejercicio	415	69,1%
	Ejercicio	186	30,9%
Habito de fumado	no fuma	525	87,4%
	Fuma	76	12,6%
índice de masa corporal	Obeso	153	25.5%
	Sobre peso	207	34.4%
	Bajo peso	36	6.0%
	Normal	205	34.1%
Total		601	100,0%

Contraste global del modelo

Para realizar el contraste global del modelo se plantean las siguientes hipótesis

H_0: las variables explicativas no influyen significativamente en el modelo.

H_1: al menos una de las variables explicativas influye significativamente en el modelo.

Para contrastar las hipótesis se calculan dos modelos.

V1: Denota verosimilitud inicial, describe un modelo que no se somete a control las variables explicativas y simplemente se adapta una intercepción para predecir la variable resultado.

-2ln (V1)= 148.865

V2: Denota verosimilitud final, describe un modelo que incluye las variables explicativas especificadas.

-2ln (V2)= 106.776

El valor de ji-cuadrada con 10 grados de libertad es igual a 42.188 con un nivel de significancia de 0.000.

Comprobando este valor con el valor critico de una distribución ji-cuadrada con 10 grados de libertad, rechazamos la hipótesis nula del modelo y concluimos que al menos uno de los coeficientes de regresión en el modelo es diferente de cero.

Estimación de los parámetros:

En la estimación de los parámetros se usa el método Newton- Raphson que produce los parámetros estimados de máxima verosimilitud, en nuestro estudio se obtuvieron las siguientes estimaciones:

- **Para j =1**

$$\log \frac{\pi_{\text{Mod}}}{\pi_{\text{Baj}}} = \alpha_j + \beta_1 Ejercicio + \beta_2 Fuma + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3$$

$\hat{\alpha}$ = -1.971, $\widehat{\beta_1} = 0.583$, $\widehat{\beta_2} = 0.112$ y $\widehat{\beta_3} = 0.909$, $\widehat{\beta_4}$ =0.3169 y $\widehat{\beta_4}$ = -0.566

El ajuste del modelo nos permite apreciar a través del test de Wald que las variables ejercicio y el IMC obeso resultaron significativa con 0.018 y 0.001 respectivamente, para el modelo que contrasta el riesgo moderado con relación al bajo y las variables fuma (0.734), IMC en sobrepeso (0.222) y IMC bajo peso (0.317) no resultaron significativas.

Construcción del modelo

$$\log \frac{\pi_{\text{Mod}}}{\pi_{\text{Baj}}} = -1.971 + 0.583\text{Ejercicio} + 0.112\text{Fuma} + 0.909IMC_1 + 0.316IMC_2 - 0.566IMC_3$$

Interpretación de los parámetros del modelo

La tabla 4 (ver Anexos) contiene una columna etiquetada Exp (B), en la cual se presentan los Odds Ratio (OR) estimados para las variables en estudio que son:

La variable ejercicio presentó un OR de 1.791 esto nos dice que las personas que hacen ejercicio son aproximadamente 2 veces mas probables de estar en el riesgo moderado en comparación con el bajo.

Para la variable IMC obeso el OR es 2.483, esto indica que las personas obesas son 2.5 veces más propensas de estar en riesgo moderado en relación al bajo que las personas con IMC normal.

- **Para j =2**

$$\log \frac{\pi_{Ele}}{\pi_{Baj}} = \alpha_j + \beta_1 Ejercicio + \beta_2 Fuma + \beta_3 IMC_1 + \beta_4 IMC_2 + \beta_5 IMC_3$$

$\hat{\alpha}$ = -2.060, $\widehat{\beta_1}$ = -0.011, $\widehat{\beta_2}$ = -0.305, $\widehat{\beta_3}$ = 1.387, $\widehat{\beta_4}$ = 1.073 y $\widehat{\beta_5}$ =-0.356

Para el modelo que contrasta el riesgo elevado con relación al bajo solamente la variable IMC en los niveles obeso y sobrepeso resultó significativa con 0.000 y 0.001 respectivamente, en cambio las variables ejercicio (0.967), fuma (0.368) y IMC en bajo peso (0.648) no resultaron significativas.

Construcción del modelo

$$\log \frac{\pi_{Ele}}{\pi_{Baj}} = \text{-2.060} - 0.011\text{Ejercicio} - 0.305\text{Fuma} + 1.387IMC_1 + 1.073IMC_2 - 0.356IMC_3$$

Interpretación de los parámetros del modelo

En el modelo que compara el riesgo elevado en comparación con el bajo el OR para IMC obeso es 4.005 esto muestra un factor de riesgo de 4 veces mas probable de estar en el riesgo elevado en comparación con el bajo.

El IMC sobre peso tiene un OR de 2.925, esto indica que las personas con sobre peso son 3 veces mas propensos de estar en el nivel elevado en comparación con el bajo.

Estimación de las probabilidades para cada nivel de riesgo del Colesterol malo

Las probabilidades respuestas para cada nivel de riesgo del colesterol malo (ldl) se calculan a partir de:

$$\pi_{j\ (x)} = \frac{\exp\left(\alpha_j + \beta_j^{'}\ x\right)}{1 + \sum_{h=1}^{J-1} \exp\left(\alpha_h + \beta_h^{'}\ x\right)} \quad \text{para j} = 1,2 \text{ y } 3$$

Perfil Sano:

Para una persona que hace ejercicio, no fuma y tiene un peso normal las probabilidades respuestas son:

Niveles de riesgo	Probabilidades
Riesgo bajo	0.80
Riesgo moderado	0.12
Riesgo elevado	0.08

Perfil Medio:

Para una persona que hace ejercicio, no fuma y es de bajo peso las probabilidades repuestas son:

Niveles de riesgo	Probabilidades
Riesgo bajo	0.87
Riesgo moderado	0.08
Riesgo elevado	0.06

Para una persona que no hace ejercicio, no fuma y tiene sobre peso las probabilidades respuestas son:

Niveles de riesgo	Probabilidades
Riesgo bajo	0.60
Riesgo moderado	0.23
Riesgo elevado	0.16

Perfil Enfermo:

Para una persona que no hace ejercicio, fuma y es obeso las probabilidades respuestas son:

Niveles de riesgo	Probabilidades
Riesgo bajo	0.47
Riesgo moderado	0.29
Riesgo elevado	0.24

6. CONCLUSIONES

- Los modelos obtenidos para el Colesterol total fueron:

$$\log \frac{\pi_{Lim}}{\pi_{Des}} = -2.677 + 0.032\ \text{Edad} - 0.110\ \text{Ejercicio} + 0.628\ IMC_1 + 0.303 IMC_2 - 0.865 IMC_3$$

$$\log \frac{\pi_{Ele}}{\pi_{Des}} = -4.220 + 0.034\text{Edad} + 0.327\text{Ejercicio} + 1.131 IMC_1 + 1.073 IMC_2 - 0.845\ IMC_3$$

- Las variables más significativas encontradas en el estudio para el modelo Nº 1 fue edad la cual resultó significativa en ambos contrastes con un OR de 1.033 y 1.035 respectivamente, lo cual indica que por cada año cumplido aumenta el riesgo de tener un colesterol a un nivel superior al de referencia, seguida de IMC obeso con un OR de 1.874 y 3.099 que nos dice que el encontrarse en este nivel aumenta el riesgo de tener colesterol elevado y al límite en comparación al de referencia.

- Los modelos obtenidos para el Colesterol malo (ldl) fueron:

$$\log \frac{\pi_{Mod}}{\pi_{Baj}} = -1.971 + 0.583\text{Ejercicio} + 0.112\text{Fuma} + 0.909 IMC_1 + 0.316 IMC_2 - 0.566 IMC_2$$

$$\log \frac{\pi_{Ele}}{\pi_{Baj}} = -2.060 - 0.011\text{Ejercicio} - 0.305\text{Fuma} + 1.387 IMC_1 + 1.073 IMC_2 - 0.356 IMC_3$$

- La variable más significativa encontrada para el modelo Nº2 fue el IMC obeso el cual resultó ser significativo en ambos contrastes con un OR de 2.483 y 4.005, lo que nos dice que el estar en un nivel obeso es un factor de riesgo de tener colesterol moderado y elevado en comparación con el bajo, respectivamente.

7. RECOMENDACIONES

- Profundizar en el estudio de regresión logística para variables categóricas a través de la Regresión Logística Ordinal.

8. REFERENCIAS BIBLIOGRAFICAS

- Agresti, A (1990,2002) Categorical Data Analysis. New York: John Wiley and sons.
- De Leeuw, Jan, Meijer, Erik. Handbook of multilevel Analysis

- Aplicación del método de regresión logística en un estudio de planificación familiar en la mujer en edad fértil en el distrito VI de Managua y Tipitapa (Tesis).
  Xóchilt Varela Pérez, Yuri Maykelly Corrales Castro, EST 378.2 v 293ª 2002.

- Aplicación de Regresión Logística en un estudio de uso de anticonceptivos en jóvenes de 14 a 25 años de edad en las áreas urbanas de las ciudades de Estelí, León y Juigalpa.
  Verónica del Carmen Zapata Mojica, Enia Sofía Barahona, EST 378.2Z35a 2006.

ANEXOS

Tabla Nº2 Resultados de los coeficientes de Regresión Logística Multinomial

colesterol 1(a)		B	Error típico	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
limite	Intersección	-2.677	.350	58.531	1	.000			
	edad	.032	.007	20.025	1	.000	1.033	1.018	1.048
	ejercicio	-.110	.231	.225	1	.635	.896	.570	1.409
	IMC=1	.628	.282	4.962	1	.026	1.874	1.078	3.257
	IMC=2	.303	.270	1.262	1	.261	1.354	.798	2.299
	IMC=3	-.865	.652	1.761	1	.185	.421	.117	1.511
elevado	Intersección	-4.220	.532	62.951	1	.000			
	edad	.034	.009	12.978	1	.000	1.035	1.016	1.054
	ejercicio	.327	.324	1.018	1	.313	1.386	.735	2.615
	IMC=1	1.131	.413	7.486	1	.006	3.099	1.378	6.970
	IMC=2	1.073	.393	7.458	1	.006	2.925	1.354	6.318
	IMC=3	-.845	1.083	.608	1	.436	.430	.051	3.592

a La categoría de referencia es: Deseable.

Tabla Nº4 Resultados de los coeficientes de Regresión Logística Multinomial

Niveles de riesgo(a)		B	Error típico	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
riesgo moderado	Intersección	-1.971	.387	25.905	1	.000			
	Fuma	.112	.329	.115	1	.734	1.118	.587	2.131
	Ejercicio	.583	.245	5.644	1	.018	1.791	1.107	2.895
	IMC=1	.909	.264	11.856	1	.001	2.483	1.480	4.167
	IMC=2	.316	.259	1.490	1	.222	1.372	.826	2.280
	IMC=3	-.566	.566	1.002	1	.317	.568	.187	1.720
riesgo elevado	Intersección	-2.060	.423	23.744	1	.000			
	Fuma	-.305	.339	.810	1	.368	.737	.380	1.432
	Ejercicio	-.011	.262	.002	1	.967	.989	.593	1.652
	IMC=1	1.387	.346	16.059	1	.000	4.005	2.032	7.894
	IMC=2	1.073	.329	10.654	1	.001	2.925	1.535	5.572
	IMC=3	-.356	.779	.209	1	.648	.700	.152	3.227

a La categoría de referencia es: riesgo bajo.

Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	730,164			
Final	660,800	69,364	10	,000

Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	660,800(a)	,000	0	.
edad	688,315	27,515	2	,000
ejercic	662,369	1,569	2	,456
IMC2	681,538	20,738	6	,002

Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo: -2 log verosimilitud	Contrastes de la razón de verosimilitud: Chi-cuadrado	gl	Sig.
Sólo la intersección	148,965			
Final	106,776	42,188	10	,000

Contrastes de la razón de verosimilitud

Efecto	Criterio de ajuste del modelo: -2 log verosimilitud del modelo reducido	Contrastes de la razón de verosimilitud: Chi-cuadrado	gl	Sig.
Intersección	106,776(a)	,000	0	.
ejercic	113,114	6,338	2	,042
fuma	107,877	1,101	2	,577
IMC2	140,178	33,401	6	,000