

UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA, LEÓN

FACULTAD DE CIENCIAS Y TECNOLOGIA

DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA



TESIS PARA OPTAR AL TÍTULO DE:
INGENIERIA ESTADISTICA

APLICACIÓN DEL METODO DE REGRESION LOGISTICA BINARIA EN UN
ESTUDIO DE DISLIPIDEMIA EN PACIENTES QUE ASISTIERON AL
LABORATORIO DE BIOQUIMICA DE LA FACULTAD DE CIENCIAS MEDICAS
DE LA UNAN LEON EN EL AÑO 2009.

Autores:

Bra. Lisdamur Peralta.

Bra. Karla Yodanis González Rivas.

Tutor:

Msc. Rafael Espinoza.

DEDICATORIA

Primeramente a **Dios** por haberme dado salud, inteligencia y dedicación para poder cumplir todos mis objetivos y metas planteadas, por su gran amor y misericordia infinita.

A mi mamá **Arcadia Obando** quien a pesar de las circunstancias y dificultades de la vida me apoyo siempre para poder culminar mis estudios y ser una profesional, porque a pesar de la distancia que se interpuso ante nosotras nunca me dejo sola y estuvo presente con su amor para darme ánimos, por enseñarme a nunca darme por vencida y levantarme cada vez que tropezaba y caía. TE AMO MAMÀ

LISDAMUR PERALTA

DEDICATORIA

Primeramente a **Dios mi padre celestial**, por haberme permitido llegar hasta éste momento, por darme salud, sabiduría e inteligencia para lograr mis objetivos, además de su infinita bondad y amor, Por los logros alcanzados.

A mi papá, **CARLOS JOSE GONZALEZ REYES**, al cual amo profundamente y del cual estoy orgullosa. Quien me enseñó desde pequeña a luchar para alcanzar mis metas y sueños.

A mi mamá, **FÁTIMA RIVAS VANEGAS**, a la cual amo muchísimo, quien ha sido, es y será para mí, ejemplo de amor, entrega, fortaleza y voluntad que junto a mi padre han sabido educarme y han logrado mi triunfo han sido los pilares para mi vida.

KARLA YODANIS GONZALEZ RIVAS.

AGRADECIMIENTO

A **Dios**, por su infinita bondad y misericordia de darme la oportunidad de realizar mis estudios y poder culminarlos, por ser mi amigo fiel quien siempre me dio la esperanza y confianza en mí misma para así no dejarme vencer en los momentos más difíciles en el trayecto del pesado, duro y difícil camino hacia el éxito.

A mi mamá **Arcadia Obando** que con su amor, confianza y sacrificio me obsequio la mejor herencia que puede existir, mi profesión. A mi abuela **Concepción Obando** quien siempre confió en mi persona y siempre estuvo a mi lado durante los momentos más difíciles que enfrente durante todo el camino de mis estudios, quien nunca permitió que me diera por vencida y siempre me dijo “Tu puedes, confió en Ti”. Gracias a mis dos madrecitas por su buena educación y amor hacia mi persona, por ser mis fuentes de inspiración y deseo de superación.

A todos mis maestros quienes compartieron conmigo su conocimiento, quienes me enseñaron a desarrollar mis habilidades académicas logrando así de mí ser un mejor estudiante, agradezco en especial al **Maestro Rafael Espinoza** por aceptar realizar esta tesis bajo su dirección. Su apoyo y confianza en mi trabajo y su capacidad para guiar mis ideas han sido un aporte invaluable. Muchas gracias Maestro Espinoza.

LISDAMUR PERALTA

AGRADECIMIENTO

Doy infinitas gracias a Dios, quien me dio la fe, la fortaleza, la salud y la esperanza para terminar este trabajo con éxito.

Agradezco a todos los profesores, por sus valiosas enseñanzas. En especial, al **Maestro RAFAEL ESPINOZA**, tutor de esta tesis, por su predisposición permanente e incondicional en aclarar nuestras dudas y por sus substanciales sugerencias durante la redacción de esta tesis, por su valiosa colaboración y buena voluntad.

A mis padres: **Carlos José González Reyes y Fátima de la Concepción Rivas Vanegas**, sabiendo que jamás existirán palabras ni manera de agradecerles, porque gracias a su cariño, guía y apoyo he llegado a realizar uno de los anhelos más grandes de mi vida, terminar mis estudios profesionales, que constituyen el legado más grande que pudiera recibir y por lo cual les viviré eternamente agradecida. Deseo expresarles que mis ideales, esfuerzos y logros han sido también suyos.

A mis hermanos quien los amo mucho: **Fátima, y Carlos González Rivas**, quienes han estado a mi lado, por su sacrificio, comprensión y confianza, por su amor y amistad incondicional, porque sin su apoyo no hubiera sido posible la culminación de mi carrera profesional.

KARLA YODANIS GONZALEZ RIVAS.

INDICE.

I. INTRODUCCIÓN.....	1
II. ANTECEDENTES.....	2
III. HIPÓTESIS.....	3
IV. PLANTEAMIENTO DEL PROBLEMA.....	4
V. OBJETIVOS.....	5
VI. MARCO TEÓRICO.....	6
VII. DISEÑO METODOLOGICO.....	188
VIII. RESULTADOS.....	221
IX. DISCUSIÓN.....	322
X. CONCLUSIÓN.....	333
XI. RECOMENDACIONES.....	344
XII. BIBLIOGRAFÍA.....	355
XIII. WEB GRAFÍA.....	366

I. INTRODUCCIÓN

La Regresión Logística surgió durante la década de 1960 como alternativa al procedimiento de estimación de los mínimos cuadrados ordinarios, tradicionalmente usado en el modelo de regresión lineal y su objetivo era estimar la probabilidad de ocurrencia de un evento como función de un conjunto de variables predictoras. No obstante, su uso comenzó a popularizarse a partir de los años setenta cuando comenzó a implementarse en el software estadístico. El nacimiento de la regresión logística, por otro lado, estuvo condicionado por un contexto histórico dominado por una crisis en las herramientas estadísticas orientadas a la predicción. El Modelo de Regresión Logística es el método estándar para el análisis de regresión de datos dicotómicos en muchas áreas del conocimiento, especialmente en las ciencias de la salud. El análisis de Regresión Logística resulta idóneo para estudiar la relación entre una o más variables independientes de forma conjunta en función de la variable dependiente de tipo dicotómica ($Y=Si$ o $Y=No$), es decir que solo admite dos categorías que definen opciones o característica mutuamente excluyentes u opuestas.

Debido a la relevancia que tiene este método lo aplicaremos en un estudio de Dislipidemia en el Departamento de León para el año 2009, con datos obtenidos del laboratorio de bioquímica de la Facultad de Ciencias Médicas, proporcionados por el Dr. Efrén Castellón. La aplicación del modelo Logístico permitirá estudiar la relación entre la variable Dislipidemia como variable dependiente y las variables, Sexo, Fuma, Ejercicio, Licor, Diabete e Hipertensión Arterial (HTA) como variables independientes. La realización de este trabajo monográfico permitirá contribuir con los conocimientos básicos acerca del Método de Regresión Logística a individuos de diferentes ámbitos sociales, especialmente a especialistas en ciencias de la Salud.

II. ANTECEDENTES.

En Nicaragua, en la Ciudad de León se encuentran algunos estudios realizados sobre regresión logística, pero enfocados en diferentes temáticas tales como:

1. Aplicación del método de Regresión Logística Multinomial en un estudio de Colesterol en la Facultad de Ciencias Médicas de la UNAN-LEON en el año 2000. (UNAN-LEON, Biblioteca Central 2009).
2. Aplicación del método de Regresión Logística en un estudio Nutricional en los municipios de Tonalá, Puerto Morazán y el Viejo.
3. 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 León, Estelí y Juigalpa.
4. 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.

Al realizar la revisión de los documentos plasmados anteriormente, tomamos información que consideramos importantes para nuestro estudio; enfocándonos en los conceptos básicos de Regresión Logística, principalmente en la estimación de los parámetros para este método, cada uno de estos documentos presentan diferentes enfoques, pero todos orientados al mismo propósito.

III. HIPÓTESIS

“El Modelo de Regresión Logística es una aplicación adecuada para determinar los factores que influyen en el padecimiento de Dislipidemia”.

IV. PLANTEAMIENTO DEL PROBLEMA

La dislipidemia es un conjunto de patologías caracterizadas por alteraciones en las concentraciones de los lípidos sanguíneos (grasa en la sangre), componentes de las lipoproteínas circulantes, a un nivel que significa un riesgo para la salud. Es un término genérico para denominar cualquier situación clínica en la cual existan concentraciones anormales de colesterol: colesterol total (Col-total), colesterol de alta densidad (Col-HDL), colesterol de baja densidad (Col-LDL) o triglicéridos (TG).

Las dislipidemia constituyen un factor de riesgo mayor y modificable de enfermedades cardiovasculares (CV), especialmente de la enfermedad coronaria (EC). Niveles muy altos de TG, especialmente cuando hay hiperquilomicronemia, han sido señalados como de riesgo en la patogenia de la pancreatitis aguda

Motivado por el riesgo que implica esta enfermedad en los pacientes y la forma de manifestación se hace necesaria identificar los factores asociados a la Dislipidemia en pacientes que visitaron el laboratorio de bioquímica en la Facultad de Ciencias Médicas de la UNAN-León. Se intentaran identificar los factores mediante la estimación de los parámetros del modelo de Regresión Logística, el cual permite la relación Dislipidemia como variable dependiente (Y) con los factores (sexo, Fuma, Diabete, Hipertensión Arterial, ejercicio, Licor) como covariables (X).

De acuerdo a este planteamiento vale preguntar ¿Cuáles son los factores que inciden en la Dislipidemia en los pacientes que asistieron al laboratorio de bioquímica de la UNAN-León para el año 2009? ¿Cuál es la relación que hay entre los factores? ¿Cuál es el modelo que describe el comportamiento?

V. OBJETIVOS

Objetivo general.

- Aplicar el método de regresión logística binaria en un estudio de dislipidemia, en pacientes que asistieron al laboratorio de bioquímica en la Facultad de Ciencias Médicas de la UNAN-León en el año 2009.

Objetivo específicos.

- Determinar en lo posible un modelo de regresión logística binaria al estudio de dislipidemia.
- Estimar e interpretar los parámetros que describen el efecto de los factores que afectan la Dislipidemia, calibrando el modelo ajustado.

VI. MARCO TEÓRICO

La Regresión Logística es una de las herramientas estadística que se utiliza para el análisis de los datos en investigaciones clínicas y epidemiológicas. La Regresión Logística permite relacionar una variable dependiente categórica Y con varias variables independientes cuantitativas o binarias X

El propósito de la Regresión Logística es producir una ecuación matemática que relacione la probabilidad de un determinado resultado con un conjunto concreto de factores de riesgo.

El objetivo que ataña esta técnica es de modelar cómo influye en la probabilidad de aparición de un suceso habitualmente dicotómico la presencia o no de diversos factores de riesgo (variables independientes). También puede ser usado para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías. Denotaremos a la variable dependiente Y, que refleja la ocurrencia o no del suceso y las variables independientes X que indican los factores de riesgo. Puesto que Y es dicotómica admitimos que asume los dos valores siguientes:

$Y= 1$, si el hecho ocurre.

$Y=0$, si el hecho no ocurre.

Lo que se procura mediante la Regresión Logística es en principio expresar la probabilidad de que ocurra el hecho en función de ciertas variables, que se presume relevante o influyente. Así pues, la Regresión Logística será el método de elección en todas aquellas investigaciones cuyos predictores sean una mezcla de variables categóricas y cuantitativas.

Modelo de regresión logística

Al construir el modelo de Regresión Logística las variables explicativas (también llamadas covariables) pueden ser de cualquier naturaleza: dicotómicas, ordinales, continuas o nominales. La forma analítica en que la probabilidad objeto de interés se vincula con las variables explicativas es:

$$P(Y = 1/x_1, x_2, \dots, x_k) = \frac{1}{1 + e^{(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k)}}$$

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$: son los parámetros del modelo.

Exp: es la función exponencial.

$P(Y = 1/x_1, x_2, \dots, x_k)$ es la probabilidad de que el suceso ocurra.

$\mathbf{x} = [x_1, x_2, \dots, x_k]$: El vector de variables independientes o covariables.

En lo subsiguiente consideraremos las estimaciones a partir de β_0, β_1 ; pues el resto de β_i se deduce igual que β_1 .

Estimación de los parámetros del modelo.

Función de verosimilitud.

Los parámetros β_0, β_1 del modelo de Regresión Logística se estiman con el método de máxima verosimilitud, porque al ser los residuales dicotómicos (con probabilidades binomiales) se obtienen estimaciones más eficientes. El valor de la función de verosimilitud $L(\beta)$ de un modelo estimado representa la probabilidad de reproducir los datos de la muestra a partir de dicho modelo.

La función de verosimilitud tiene como fin estimar $\beta = (\beta_0, \beta_1)$ y analizar el comportamiento del modelo estimado se toma una muestra aleatoria de tamaño n dada por $(x_i, y_i)_{i=1,2,\dots,n}$ donde el valor de las variables independientes es

$x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ e y_i Variable dependiente pertenece al intervalo $[0, 1]$ es el valor observado de Y en el i -ésimo elemento de la muestra.

Vamos a desarrollar la estimación de los parámetros para el caso de una variable.

Sea una muestra de n elementos, donde se ha observado la variable respuesta Y (que sólo puede tomar dos valores: cero y uno) las variables X_i .

La función de probabilidad de una observación cualquiera es:

$$p(y = 1/x) = \pi(x_i) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

$$p(y = 0/x) = 1 - \pi(x_i)$$

Por lo tanto

$$p(y/x) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

Por lo tanto la función de probabilidad de la muestra es:

$$p(y_1, y_2, \dots, y_n) = \prod_i \pi(x_i)^{y_i} * [1 - \pi(x_i)]^{1-y_i}$$

Función de verosimilitud

$$L(\beta) = \sum_{i=1}^n [y_i * \text{Log}(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))]$$

Derivamos β con respecto a β_0

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_0} &= \sum_{i=1}^n \left(\frac{y_i}{\pi(x_i)} * \pi'(x_i) + \frac{(1 - y_i)}{(1 - \pi(x_i))} \right) [1 - \pi(x_i)]' \\ &= \sum_{i=1}^n \frac{y_i}{\pi(x_i)} * \pi'(x_i) + \frac{(1 - y_i)}{1 - \pi(x_i)} (-\pi'(x_i)) \end{aligned}$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \pi'(x_i) \left[\frac{y_i}{\pi(x_i)} - \frac{(1-y_i)}{1-\pi(x_i)} \right]$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \pi'(x_i) \left[\frac{(1-\pi(x_i))y_i - \pi(x_i)(1-y_i)}{(\pi(x_i)(1-\pi(x_i)))} \right]$$

Procedemos a derivar $\pi(x_i)$ con respecto a β_0

$$\begin{aligned} \pi'(x_i) &= -1(1 + e^{-\beta_0 - \beta_i x_i})^{-2} * (-1 * e^{-\beta_0 - \beta_i x_i}) \\ &= \frac{e^{-\beta_0 - \beta_i x_i}}{(1 + e^{-\beta_0 - \beta_i x_i})^2} \\ &= \frac{e^{-\beta_0 - \beta_i x_i}}{(1 + e^{-\beta_0 - \beta_i x_i})} * \frac{1}{1 + e^{-\beta_0 - \beta_i x_i}} \\ &= \pi(x_i) * [1 - \pi(x_i)] \end{aligned}$$

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_0} &= \sum_{i=1}^n (1 - \pi(x_i)) \pi(x_i) \left[\frac{y_i - \pi(x_i)y_i - \pi(x_i) + \pi(x_i)y_i}{\pi(x_i)(1 - \pi(x_i))} \right] \\ &= \sum_{i=1}^n (y_i - \pi(x_i)) = 0 \quad (0) \end{aligned}$$

En función de β_1 la estimación es la siguiente

$$\pi(x_i) = \left(\frac{1}{1 + e^{-\beta_0 - \beta_1 x_1}} \right) = (1 + e^{-\beta_0 - \beta_1 x_1})^{-1}$$

Derivamos con respecto a β_1

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(\frac{y_i}{\pi(x_i)} * \pi'(x_i) + \frac{(1 - y_i)}{(1 - \pi(x_i))} \right) [1 - \pi(x_i)]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \frac{y_i}{\pi(x_i)} * \pi'(x_i) - \frac{(1 - y_i)}{1 - \pi(x_i)} (-\pi'(x_i))$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \pi'(x_i) \left[\frac{y_i}{\pi(x_i)} - \frac{(1 - y_i)}{1 - \pi(x_i)} \right]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \pi'(x_i) \left[\frac{(1 - \pi(x_i))y_i - (1 - y_i)\pi(x_i)}{\pi(x_i)(1 - \pi(x_i))} \right]$$

$$[\pi'(x_i) = x_i - 1(1 + e^{-\beta_0 - \beta_1 x_1})^{-2} * (-1 * e^{-\beta_0 - \beta_1 x_1})]$$

$$= \frac{x_i e^{-\beta_0 - \beta_1 x_1}}{(1 + e^{-\beta_0 - \beta_1 x_1})^2} = \frac{x_i e^{-\beta_0 - \beta_1 x_1}}{(1 + e^{-\beta_0 - \beta_1 x_1})} * \frac{1}{(1 + e^{-\beta_0 - \beta_1 x_1})}$$

$$= x_i [1 - \pi(x_i)]\pi(x_i)]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i [1 - \pi(x_i)]\pi(x_i) \left[\frac{y_i - \pi(x_i)y_i - \pi(x_i) + \pi(x_i)y_i}{\pi(x_i)(1 - \pi(x_i))} \right]$$

$$\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i [1 - \pi(x_i)] \pi(x_i) \left[\frac{y_i - \pi(x_i)}{\pi(x_i) (1 - \pi(x_i))} \right]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i [1 - \pi(x_i)] \pi(x_i) \left[\frac{(1 - \pi(x_i)) y_i - (1 - y_i) \pi(x_i)}{\pi(x_i) (1 - \pi(x_i))} \right]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i (1 - \pi(x_i)) \pi(x_i) \left[\frac{(1 - \pi(x_i)) y_i - (1 - y_i) \pi(x_i)}{\pi(x_i) (1 - \pi(x_i))} \right]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n x_i (1 - \pi(x_i)) \pi(x_i) \left[\frac{y_i - \pi(x_i) y_i - \pi(x_i) + \pi(x_i) y_i}{\pi(x_i) (1 - \pi(x_i))} \right]$$

$$\sum_{i=1}^n x_i (y_i - \pi(x_i)) = 0 \quad (1)$$

Este sistema de ecuaciones (0) y (1) son no lineales para β_0 y β_1 , no tienen una solución directa, requieren de métodos especiales para su solución, como el método numérico de Newton Rapson.

Bondad de ajuste del modelo estimado.

En un modelo de Regresión Logística estimado por el método de máxima verosimilitud, el ajuste global se describe con estadísticos derivados de la verosimilitud del modelo. Entre los que se encuentra el -2LL (0), -2LL (β), el R^2 de Cox y Snell y el R^2 de Nagelkerke cada uno de ellos nos explica la proporción de incertidumbre de los datos ajustados por el modelo.

-2LL (0) indica la verosimilitud que solo contiene el término constante.

-2LL (β) indica la verosimilitud del modelo con las variables predictoras.

$$Desviance = -2 \left[\log \left(\frac{p(y/\hat{\vartheta}_0)}{p(y/\vartheta_s)} \right) \right]$$

Este valor puede ser contemplado como un descriptivo de la bondad de ajuste de este modelo, que cuanto más próximo a cero sea, mejor es el ajuste del modelo. El valor R^2 de Cox y Snell viene dado por:

$$R^2 = 1 - \left[\frac{L(0)}{L(\beta)} \right]^{2/n}$$

En donde $L(0)$ es la verosimilitud del modelo con solo la constante. La $L(\beta)$ es la verosimilitud del modelo considerado. Por otra parte, este índice tiene inconveniente de no alcanzar el valor 1 (100%) cuando el modelo reproduce exactamente los datos, en efecto si $L(\beta) = 1$ el $R^2_{max} = 1 - [L(0)]^{2/n}$.

Por este motivo Nagelkerke propone el índice corregido de la siguiente forma:

$$R_c^2 = \frac{R^2}{R^2_{max}}$$

El cual vale 1 si el modelo explica el (100%) de la incertidumbre de los datos.

Significancia de los parámetros del Modelo.

- ❖ Evaluación del estadístico de Wald.
- ❖ Evaluación del estadístico $-2LL(l_0)$

Evaluación del estadístico de Wald.

La significancia de los parámetros se estudia con el cociente:

$$Z = \frac{\beta}{SE(\beta)}$$

Donde β es el coeficiente de regresión logística estimado.

$SE(\beta)$ es el error del cociente.

El estadístico Z Sigue una ley normal estandarizada con el cuadro del cociente llamado estadístico de Wald el cual sigue una ley de ji-cuadrado con 1 grado de libertad.

$$\chi^2_{\text{Wald}} = \left[\frac{\beta}{\text{SE}(\beta)} \right]^2 \sim \chi^2_1$$

Si el valor experimental de Wald es menor que el nivel de significancia fijado por el contraste rechazaremos la hipótesis nula y concluiremos diciendo que la variable independiente influye en la probabilidad de las características de la variable dependiente.

Este estadístico se usa en caso de que se esté discutiendo si la supresión de una variable especificada reduce significativamente el grado de explicación en el modelo. Diferentes estudios (Hauck y Donner 1977: Jeannigs, 1986) han puesto de manifiesto la falta de potencia de esta prueba cuando el valor del parámetro β se aleja de cero y recomienda que en su lugar se aplique una prueba de razón de verosimilitud (LR). La prueba de razón de similitud consiste en comparar el logaritmo de la verosimilitud del modelo estimado con las P variables predictoras L(M) con el modelo que solo tiene la constante L(0). A través del cociente:

$$\text{Desvianza} = -2\ln LR = -2\ln \left(\frac{L(0)}{L(M)} \right) = -2LL(0) - [-2LL(M)] \rightarrow \chi^2_1$$

La prueba de significancia global indica si el modelo estimado es significativo.

Calibración del modelo:

Prueba de Omnibus:

La prueba de ómnibus es el nombre general que se refiere a un conjunto o una prueba global y en la mayoría de los casos ensayo ómnibus se llama en otras expresiones tales como: F-test o prueba de Chi-cuadrado que evalúa la hipótesis nula de los coeficientes (β) de todos los términos (excepto la constante) incluidos en el modelo son cero. El estadístico ji cuadrado para este contraste es la

diferencia entre el valor de -2LL para el modelo solo con la constante y el valor de -2LL para el modelo actual.

Prueba de bondad de ajuste de Hosmer y Lemeshow.

La calibración del modelo es un aspecto del ajuste que consiste en valorar la concordancia entre las probabilidades observadas en la muestra (P_i) y las predichas por el modelo (π_i). Esta valoración se puede realizar con el estadístico Ji-Cuadrado de bondad de ajuste. Si la muestra está formada por “n” sujetos tendremos J \leq n estimaciones (π_i) diferentes. En efecto. Si el modelo contiene p variables predictoras $x = x_1; x_2; \dots x_p$, cada combinación x_j diferente de valores de los predictores dará una estimación diferente de π_j de manera que J tomara el máximo valor n si ninguna combinación se repite. Sin embargo puesto que en la mayor parte de los estudios el número de J de combinaciones diferentes es grande, Hosmer y Lemeshow proponen ordenar “n” sujeto según las predicciones π_i y dividirlos en g=10 grupos de aproximadamente igual tamaño (llamados deciles de riesgos). Esta división de grupo de igual tamaño es especialmente adecuada cuando muchas de las probabilidades estimadas π_i son pequeñas. Si se comparan las probabilidades predichas y observadas en estos g=10 grupos de aproximadamente el mismo tamaño (llamados deciles de riesgo). Estas divisiones en grupos de igual tamaño es especialmente adecuada cuando muchas de las probabilidades estimadas (π_i) son pequeñas.

No obstante los grupos no acostumbran a tener exactamente el mismo tamaño porque los sujetos con la misma combinación de valores de las variables independientes se sitúan en el mismo grupo.

Si se comparan las probabilidades predichas y observadas en estos g=10 grupos con el estadístico ji-cuadrado de bondad de ajuste:

$$\chi^2 = \sum_{j=1}^2 \sum_{i=1}^g \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \rightarrow \chi_{\theta}^2$$

Hosmer y Lemeshow han demostrado mediante simulaciones que la significancia del estadístico ji cuadrado de Wald puede ser calculada en la ley de Ji-cuadrado con $g-2=8$ grados de la libertad si la mayor parte de las frecuencias esperadas son superiores a 5 ($e_{ij} > 5$) y ningunas de ellas es inferior a 1.

Interpretación de los coeficientes de regresión logísticas.

El uso de los coeficientes de regresión siempre depende del tipo de estudio que se utilice. Si se trata de un estudio de cohorte las probabilidades estiman incidencias acumuladas y puede ser interpretado como una medida de riesgo. Si es un estudio transversal, las probabilidades estiman prevalencia y si se trata de un estudio de caso y controles las probabilidades estiman una proporción que carecen de interés porque están afectadas por el número de casos y controles.

Seleccionados ya los coeficientes tienen una estimación directa con este estudio a través de la transformación en Odds del modelo:

$$Odds = \frac{\pi_x}{1 - \pi_x} = e^{\beta_0} e^{\beta_1 x_1} \dots e^{\beta_k x_k}$$

Si una variable predictora x se incrementa en una unidad ($x \rightarrow +1$) la razón de Odds es:

$$OR = \frac{Odds_{x+1}}{Odds_x} = e^{\beta_i}$$

Que es general e independiente del número de variables predictoras X del modelo. Si se mantienen constantes todas las variables predictoras excepto las variables X_i su término aparece tanto en el numerador como en el denominador del anterior cociente y se podrá simplificar de manera que OR será igual al e^{β_i} . Así pues e^{β_i} es el factor por el cual se multiplica los Odds de la respuesta (Y) cuando la variable predictora X_i se incrementa en una unidad. Un valor superior que 1 ($\beta > 0$) indica que el factor X incrementa la probabilidad de respuesta $Y=1$. Un

valor menor que 1 ($\beta > 0$) indica que el factor X decreta la probabilidad de respuesta $Y=1$.

El intervalo de confianza del 100 (1- α) % para RO mide el efecto de un incremento de la variable X:

$$\widehat{OR} \cdot \exp^{\pm Z_{\alpha/2} \widehat{se}(\beta_i)}$$

OR: es la estimación puntual de la razón de odds.

Exp: es la base del logaritmo natural.

Z: es el cuartil normal dada la significancia fijado α

Se (β): es el error estándar del coeficiente β

Riesgo relativo

Este coeficiente, nos indica cuanto más probable es desarrollar la enfermedad si se está en primer caso, que si se está en el segundo.

$$RR = \frac{P_A^{(E)}}{P_B^{(E)}}$$

Esta razón expresa el riesgo relativo de padecer la enfermedad E cuando se está en la condición A respecto a cuándo se está en la condición B. el riesgo relativo mide la fuerza de la asociación entre la exposición y la enfermedad; indica la probabilidad de que se desarrolle la enfermedad en los expuestos a un factor de riesgo en relación al grupo de los no expuestos.

Intervalos de confianza para la estimación del riesgo relativo.

$$\widehat{RR} \cdot \exp^{\pm Z_{\alpha/2} \widehat{se}(\beta_i)}$$

RR: es la estimación puntual del riesgo relativo.

Exp: es la base del logaritmo natural.

Error estándar= $se(\beta)$

La interpretación del riesgo relativo se efectúa de la siguiente manera: si el riesgo relativo es igual a uno, el factor del riesgo no tiene influencia con la aparición de la enfermedad. Si al construir el intervalo de confianza no incluye el valor uno concluiremos que el riesgo es estadísticamente significativo. En cambio si el riesgo relativo fuese menor que uno y también su intervalo de confianza, estaremos ante la presencia de un factor de protección ante la ocurrencia de un suceso.

VII. DISEÑO METODOLOGICO.

El estudio se realizó durante el período de enero y noviembre del año 2009, en los pacientes que asisten al laboratorio de bioquímica de la facultad de ciencias médica de la universidad nacional autónoma de Nicaragua.

Estudio Transversal

En este tipo se estudia la exposición y la enfermedad en una población bien definida en un momento determinado. Y se utiliza fundamentalmente para conocer la prevalencia de una enfermedad o de un factor de riesgo. Su cociente permite estimar la razón de prevalencia (PR):

$$\widehat{PR} = \frac{\widehat{P}_r(Y = 1/X_1)}{\widehat{P}_r(Y = 1/X_0)}$$

VARIABLES CATEGÓRICAS.

La regresión logística permite tratar variables predictoras categóricas de idéntica forma que la regresión múltiple. Cuando alguna de las variables explicativas es de índole nominal de más de dos categorías (politómicas), para incluirlas en el modelo hay que darles un tratamiento especial. Si estamos en presencia de una variable nominal con C categorías debemos incluirlas en el modelo de regresión logística como variable categórica de manera que a partir de ella se crean C-1 variables dicotómicas llamadas dummy o ficticias. Al crear las variables dummy se debe precisar con cuál de las categorías de la variables original interesa comparar el resto y esa será la llamada categoría de referencia. En general, el exponencial del coeficiente correspondiente a una de las variables dummy, estima la magnitud en que varía el riesgo de que ocurra el suceso.

Método de codificación de las variables categóricas.

Codificación respecto a una categoría de referencia. Se utiliza siempre que se tenga sentido clínico considerar una de las categorías de referencia o cuando deseamos estimar razones de Odds. El coeficiente de regresión b de cada variable ficticia refleja el contraste entre la categoría que representa la variable ficticia y la categoría de referencia.

Tipo de estudio.

Estudio analítico de corte transversal.

Área o población de estudio.

En este estudio se definieron casos, solamente a aquellos pacientes que asistieron al laboratorio de bioquímica de la facultad de ciencias médicas UNAN-León para el año 2009, con un total de 301 pacientes.

Periodo de estudio.

Noviembre 2012- Noviembre 2013

Fuentes de información.

Laboratorio de bioquímica de la facultad de ciencias médicas UNAN-LEON,

Selección del modelo de Regresión Logística.

En primer paso consistió en seleccionar en la base de datos original, todas aquellas variables independientes o explicativas que estuvieron más relacionadas en el desenlace de la variable dependiente. "caso de dislipidemia". El segundo consistió en realizar análisis univariado para identificar aquellas variables independientes que pudieran influenciar en la variable dependiente. Luego se procedió a recodificar las variables originales en variables dicotómicas y realizar el análisis bivariado sobre la variable dependiente (se asigna $Y=1$ si padece la enfermedad y $Y=0$: en caso contrario). En el modelo logístico solamente fueron incluidas aquellas variables independientes con significancia estadística.

Operacionalización de las variables.

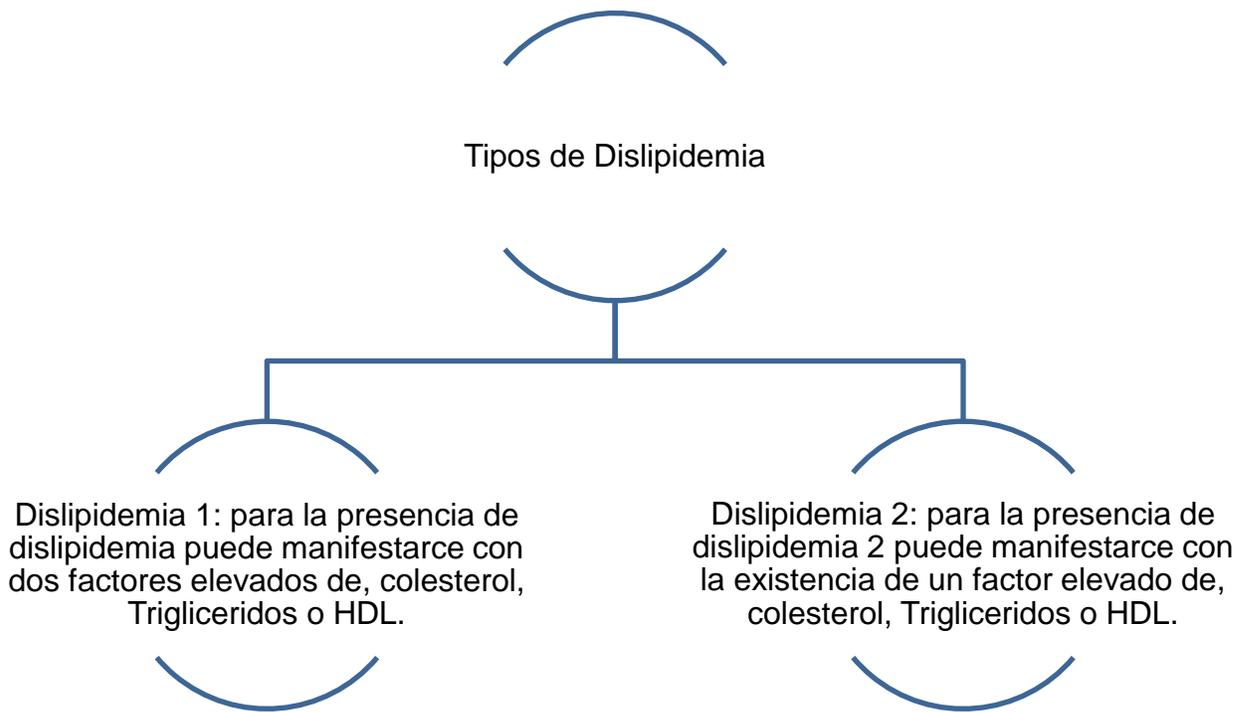
Variables	Definición	Factores	Codificación
Sociodemográficas.	Es la agrupación de algunas características personales.	Edad	Años
		Sexo	0: Femenino 1: Masculino
		Fuma: habito de fumado	0: No fuma 1: Fuma
		HTA: Hipertensión Arterial. Incremento de la presión sanguínea de los pacientes	0: Hipertenso 1: Normal
		Diabética: trastorno metabólico.	0: No 1: Si
		Ejercicio: actividad física que realizan los pacientes.	0: No 1: Si
		Licor: consumo de licor.	0: No toma 1: Toma
		Dislipidemia 1:	0: No tiene 1: Tiene
		Dislipidemia 2:	0: No tiene 1: Tiene
		Nivel de riesgo del índice de masa corporal.	0: Alto riesgo 1: Bajo riesgo

Parámetros que fueron incluidos en el cálculo de Dislipidemia 1 y Dislipidemia 2 para el procesamiento y análisis de los datos.

Factores que inciden en la enfermedad Dislipidemia	Dislipidemia 1				Dislipidemia 2		
Colesterol (CT) : 0: Normal: ≤ 200 1: Elevado: >200	P	P	P	A	P	A	A
Triglicéridos (TAG): 0: Normal: ≤ 200 1: Alterado:>200	P	P	A	P	A	P	A
HDL 0: Normal:≤40 1: Riesgo:>40	P	A	P	P	A	A	P

Donde: P= presencia del factor.

A= ausencia de un factor.



VIII. RESULTADOS.

La población en estudio fue de 301 pacientes que fueron atendidos en el laboratorio de Bioquímica en la Facultad de Ciencias Médicas UNAN-León en el año 2009. El 4.98% de la muestra presentó Dislipidemia 1, y un 47.18% Dislipidemia 2.

Al realizar el análisis mediante Regresión logística para determinar los factores que influyen en Dislipidemia y planteando el modelo general para cada uno de los casos obtuvimos:

Modelo 1. Dislipidemia 1

$$P(y=1/x) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 - \beta_4 X_4 - \beta_5 X_5 - \beta_6 X_6)}$$

Inicialmente se consideraron todas las covariables y se aplicó el método introducir, obteniendo como resultado que sólo la covariable Diabetes, resulta ser ligeramente significativa, como puede apreciarse en la tabla No.1

Tabla Nº 1**Variables en la ecuación**

Paso 1(a)	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
						Odds	Inferior	Superior
X ₁ Sexo	.227	.670	.115	1	.735	1.255	.338	4.663
X ₂ Licor	-.787	.674	1.362	1	.243	.455	.121	1.707
X ₃ Fuma	-.972	.661	2.162	1	.141	.378	.104	1.382
X ₄ HTA	-.860	.578	2.216	1	.137	.423	.136	1.313
X ₅ Diabética	-1.084	.642	2.849	1	.091	.338	.096	1.191
X ₆ Ejercicio	.072	.549	.017	1	.896	1.074	.367	3.149
X ₀ Constante	-.397	1.013	.153	1	.695	.672		

a Variable(s) introducida(s) en el paso 1: Sexo, Licor, Fuma, HTA, Diabética, Ejercicio.

Esto nos llevó a aplicar el método Adelante de Wald, para seleccionar de forma definitiva las covariables que debía de estar en el modelo de Dislipidemia1. Los resultados obtenidos al aplicar el método de Wald fueron que solamente las covariables Fuma y Diabetes resultaron significativas (0.033 y 0.053), como puede apreciarse en la tabla No. 2, obteniéndose un OR de 0.285 y 0.295 respectivamente, resultando ambas covariables factores de protección para el caso de Dislipidemia 1 debido que el OR son valores menores que uno.

Tabla N°2 método delante de wald

Variabes en la ecuación

		B	E.T.	Wald	Gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
							Odds	Inferior	Superior
Paso 1(a)	Fuma	-1.123	.574	3.835	1	.050	.325	.106	1.001
	Constante	-2.079	.474	19.218	1	.000	.125		
Paso 2(b)	Fuma	-1.255	.588	4.554	1	.033	.285	.090	.903
	Diabética	-1.219	.630	3.743	1	.053	.295	.086	1.016
	Constante	-.982	.725	1.836	1	.175	.374		

a Variable(s) introducida(s) en el paso 1: Fuma.

El modelo logístico resultante para Dislipidemia 1 es el siguiente:

$$P(Y = 1/X) = \frac{1}{1 + \exp^{[0.982+1.255*fuma+1.219*Diabetes]}}$$

Validación del modelo de Dislipidemia 1.**Pruebas Omnibus sobre los coeficientes del modelo**

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	3.361	1	.067
	Bloque	3.361	1	.067
	Modelo	3.361	1	.067
Paso 2	Paso	3.193	1	.074
	Bloque	6.554	2	.038
	Modelo	6.554	2	.038

La Prueba de Omnibus sobre los coeficientes del modelo muestra la significancia estadística (0.038) indica que el modelo con las nuevas covariables de Fuma y Diabetes, mejora el ajuste de forma significativa, esto nos indica que las covariables tomadas en conjunto, contribuyen efectivamente a explicar las modificaciones que producen en la variable respuesta.

Para evaluar de forma global la validez del modelo, complementando lo anterior tenemos el RESUMEN DE LOS MODELOS.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	115.851(a)	.011	.034
2	112.658(a)	.022	.066

a La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

La R cuadrado de Cox y Snell con un valor de (0.022) nos indica que solo en 2.2% de la variación de la variable dependiente es explicada por las covariables incluidas en el modelo.

Para evaluar la bondad de ajuste del modelo de regresión Logística para el caso de Dislipidemia 1 se basa en la prueba de prueba de Hosmer y Lemeshow, que plantea la hipótesis nula de que el modelo se ajusta los datos, como podemos ver en la siguiente tabla nos da un resultado de no significancia (.564), es decir el modelo se ajusta a los datos).

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	.000	0	.
2	.333	1	.564

Pronóstico. Nos permitió evaluar el ajuste del modelo de regresión comparando los valores predichos con los valores observados, como puede observarse en la siguiente tabla el modelo es pronosticado con un 95%, lo que significa que es bastante confiable

Tabla de clasificación(a)

Observado			Pronosticado		
			Dislipidemia_1		Porcentaje correcto
			No tiene	Tiene	
Paso 1	Dislipidemia_1	No tiene	286	0	100.0
		Tiene	15	0	.0
	Porcentaje global				95.0
Paso 2	Dislipidemia_1	No tiene	286	0	100.0
		Tiene	15	0	.0
	Porcentaje global				95.0

a El valor de corte es .500.

Ejemplos a utilizar en el modelo de Dislipidemia 1

- ✓ Aplicación del Modelo Especifico para el caso de Dislipidemia 1:

1. Cuál es la probabilidad de que una persona que fuma, y es diabética presente Dislipidemia 1.

$$P(Y = 1/X) = \frac{1}{1 + \exp[\beta_0 - \beta_2 \text{fuma} - \beta_4 \text{Diabete}]}$$

Donde Y=1: la probabilidad de que padezca Dislipidemia 1

Sustituyendo las variables por sus respectivos valores nos queda:

$$P(Y = 1/X) = \frac{1}{1 + \exp[-(-0.982) - (-1.255)(1) - (-1.219)(1)]}$$

$$P(Y = 1/X) = \frac{1}{1 + \exp[3.456]} = \frac{1}{1 + 31.68996281}$$

$$P(Y = 1/X) = 0.030590429$$

La probabilidad de que una persona que fuma y es diabética padezca de Dislipidemia 1 es de 0.03006 (3.06%).

2. Cuál es la probabilidad de que una persona que fuma, y no es diabética presente Dislipidemia 1.

$$P(Y = 1/X) = \frac{1}{1 + \exp[-(-0.982) - (-1.255)(1) - (-1.219)(0)]}$$

$$P(Y = 1/X) = \frac{1}{1 + \exp[2.237]} = \frac{1}{1 + 9.365193521}$$

$$P(Y = 1/X) = 0.096476732$$

La probabilidad de que una persona que fuma y es diabética padezca de Dislipidemia 1 es de 0.096 (9.6%).

3. Cuál es la probabilidad de que una persona que no fuma, y es diabética presente Dislipidemia 1.

$$P(Y = 1/X) = \frac{1}{1 + \exp[-(-0.982) - (-1.255)(0) - (-1.219)(1)]}$$

$$P(Y = 1/X) = \frac{1}{1 + \exp[2.201]} = \frac{1}{1 + 9.034043027}$$

$$P(Y = 1/X) = 0.099660724$$

La probabilidad de que una persona que no fuma y es diabética padezca de Dislipidemia 1 es de 0.099 (9.96%).

4. Cuál es la probabilidad de que una persona que no fuma, y no es diabética presente Dislipidemia 1.

$$P(Y = 1/X) = \frac{1}{1 + \exp[-(-0.982) - (-1.255)(0) - (-1.219)(0)]}$$

$$P(Y = 1/X) = \frac{1}{1 + \exp[0.982]} = \frac{1}{1 + 2.669790487}$$

$$P(Y = 1/X) = 0.2726$$

La probabilidad de que una persona que no fuma y no es diabética padezca de Dislipidemia 1 es de 0.27 (27.2%).

Resumen de los cuatro casos condicionados.

Probabilidad Dislipidemia	Fuma	No fuma
Diabete	0.0306	0.0997
No diabética	0.0965	0.2726

Modelo general con Dislipidemia 2.

$$P(y=1/x) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 - \beta_4 X_4 - \beta_5 X_5 - \beta_6 X_6)}$$

Inicialmente se consideraron todas las covariables y se aplicó el método introducir, obteniendo como resultado que sólo las covariables sexo y Ejercicio, resultan ser ligeramente significativa, como puede apreciarse en la tabla No.1

Tabla N°1

Variables en la ecuación

Paso 1(a)	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
						Odds	Inferior	Superior
X ₁ Sexo	1.003	.295	11.530	1	.001	2.727	1.528	4.865
X ₂ Licor	-.035	.321	.012	1	.913	.965	.514	1.812
X ₃ Fuma	-.265	.371	.509	1	.476	.768	.371	1.588
X ₄ HTA	-.036	.269	.018	1	.893	.965	.570	1.633
X ₅ Diabética	-.627	.381	2.699	1	.100	.534	.253	1.129
X ₆ Ejercicio	.440	.244	3.262	1	.071	1.553	.963	2.505
X ₀ Constante	.195	.559	.121	1	.728	1.215		

a Variable(s) introducida(s) en el paso 1: Sexo, Licor, Fuma, HTA, Diabética, Ejercicio.

Esto nos llevó a aplicar el método Adelante de Wald, para seleccionar de forma definitiva las covariables que deben de estar en el modelo de Dislipidemia 2. Los resultados obtenidos al aplicar el método de Wald obtuvimos que no hubieron cambios, confirmándonos de esta manera que las covariables que influyen en la variable respuesta son sexo y ejercicio, resultaron significativas (0.000 y 0.071), como puede apreciarse en la tabla No. 2, obteniéndose un OR de 2.727 y 1.553.

Tabla N°2

Variables en la ecuación

Paso 1(a)	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)
-----------	---	------	------	----	------	--------	------------------------

Resultados.

						Odds	Inferior	Superior
Sexo	1.043	.258	16.333	1	.000	2.837	1.711	4.704
Ejercicio	.435	.241	3.259	1	.071	1.545	.963	2.477
Constante	-.644	.185	12.055	1	.001	.525		

a Variable(s) introducida(s) en el paso 1: Sexo, Ejercicio.

El modelo Logístico resultante para Dislipidemia 2 es el siguiente:

$$P(Y = 1/X) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 * \text{sexo} - \beta_6 * \text{Ejercicio})}$$

Validación del modelo Dislipidemia 2.

Pruebas Omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	16.220	1	.000
	Bloque	16.220	1	.000
	Modelo	16.220	1	.000

La prueba de Omnibus sobre los coeficientes del modelo nos muestra una significancia estadística (0.00) esto nos indica que el modelo con las nueva covariable Sexo y Ejercicio no ajusta un buen modelo.

Para evaluar de forma global la validez del modelo, complementando lo anterior tenemos el RESUMEN DE LOS MODELOS.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke

1	400.094(a)	.052	.070
---	------------	------	------

a La estimación ha finalizado en el número de iteración 3 porque las estimaciones de los parámetros han cambiado en menos de .001.

La R cuadrado de cox y snell con un valor de (0.052) nos indica que solo en 5.2% de la variación de la variable dependiente es explicada por las variables incluidas en el modelo.

Para evaluar la bondad de ajuste del modelo de regresión Logística para el caso de Dislipidemia 1 podemos observar la tabla de prueba de Hosmer y Lemeshow.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	.000	0	.

Para la calibración del modelo de Hosmer y Lemeshow mediante el paso de estadístico ji-cuadrado es nulo y con el grado de significancia de $p=0.00$, esto nos indica que el modelo no se ajusta a los datos.

Pronostico. Nos permitió evaluar el ajuste del modelo de regresión comparando los valores predichos con los valores observados Como puede observarse en la siguiente tabla el modelo es pronosticado con un 61%.

Tabla de clasificación(a)

Observado	Pronosticado
-----------	--------------

Resultados.

			Dislipidemia_2		Porcentaje correcto
			No tiene	Tiene	
Paso 1	Dislipidemia_2	No tiene	124	35	78.0
		Tiene	80	62	43.7
Porcentaje global					61.8

a El valor de corte es .500

IX. DISCUSIÓN.

En caso Dislipidemia 1 obtuvimos como resultados dos covariables significativas (Diabetes y Fuma) que estadísticamente está relacionada con la variable respuesta, pero a la vez son un factor de protección. Estas covariables en la realidad deberían de ser un factor de riesgo para la variable Dislipidemia 1.

El resultar estas covariables un factor de protección según especialista nos encontramos con la siguiente situación, en cuanto a la covariable diabética esto es debido a que las personas con este problema, que están bajo control o bien controlada la enfermedad, al momento de los diagnósticos puede aparecer sin Dislipidemia, esto puede ocurrir dependiendo del tiempo de la evolución de la diabetes, entre menos tiempo la persona este con la enfermedad, hay menos probabilidad de padecer de Dislipidemia, pero si el tiempo es de 10 o más años, si corre riesgo de padecerla. Con respecto a la covariables Fumar el motivo de ser un factor de protección se debe al factor tiempo de fumado, la cantidad de cigarrillo consumidos y la frecuencia con que realiza el hábito de fumar.

Para el caso de Dislipidemia 2 aplicando el método de introducir, para conocer las covariables que más influyen para el modelo 2, la cual únicamente encontramos las covariables sexo y Ejercicio, estas covariables están estadísticamente relacionadas con la variable respuesta (Dislipidemia 2) con esta se controla el aumento o no aumento de HDL; según el especialista, al aumentar el HDL en el paciente que si realiza ejercicio según especialista en estudios médicos no presenta dislipidemia tipo 2.

Estadísticamente nos encontramos que este modelo de dislipidemia 2 no se ajusta a los datos el cual no da lugar a la aplicación del mismo. Esto se debe a que los datos presentan algunos problemas, que nos afecta un poco en los resultados del modelo 2 de regresión logística y en cuanto a las covariables que resultan como factor de protección.

X. CONCLUSIÓN.

- 1) Las covariables más significativas encontradas en el estudio para el modelo de dislipidemia 1 fueron: Fuma y diabete las cuales resultaron significativas (0.033 y 0.053), con un OR de 0.285 y 0.295, resultando ambas covariables factor de protección, que real mente deberían considerarse factores de riesgo. Mediante estos resultados obtenemos el siguiente modelo logístico.

$$P(Y = 1/X) = \frac{1}{1 + \exp[(0.982) - (-1.255)fuma - (-1.219)Diabete]}$$

La validación del modelo mediante la prueba de omnibus resultando significativo, (significancia. 0.038) nos indica un ajuste adecuadodel modelo, obteniendo un pronostico del 95%, esto indica que el modelo es confiable.

Mediante la prueba de Hosmer y Lemeshow, demostramos que el modelo se ajusta a los datos, con un valor de no significancia estadística de (.564)

- 2) La prueba de Hosmer y Lemeshow sobre los coeficientes del modelo para Dislipidemia 2, con una significancia estadística de (0.00) nos refleja que no hay un buen ajuste del modelo.

XI. RECOMENDACIONES.

- ✓ Sugerimos contrastar el modelo de regresión logística con otro modelo para desarrollar temas en el ámbito de la salud o de otro tipo ya que este modelo es una herramienta que nos permite analizar e interpretar la medición de riesgo de diferentes problemas.
- ✓ Sugerimos que al momento de diseñar, el instrumento de recolección de datos sea elaborado o supervisado por un estadístico de igual manera al recolectar y ejecutar los datos.
- ✓ Extraer con detalle la variable según el grado de fumado y los padecimientos diabéticos, para diferenciar de mejor forma los factores de riesgo con los factores de protección.

XII. BIBLIOGRAFÍA.

- ❖ José M. Doménech Massons. Diseños y Estadística en ciencias de la salud Análisis multivariante (Modelos de Regresión). (Barcelona 2001)
- ❖ Luis Carlos Silvas Aycaguer. Excursión a la regresión Logística en Ciencias de la Salud. (1994)
- ❖ Albert J. Jovell. Análisis de Regresión Logística. (Barcelona1962)

XIII. WEB GRAFÍA.

- ❖ Molinero Luis M. (Alce Ingeniería). LA REGRESION LOGISTICA (I). (Enero 2001): (<http://www.seh-lilha.org/rlogis1.htm>)
- ❖ Flores Manrique. Modelos de Regresión Logísticas: (sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/flores_ml/Cap3.pdf)
- ❖ Silva Aycaguer Luis Carlos. Excursión a la regresión logística en ciencias de la salud. (1995): (Excursión a la Regresión Logística en Ciencias de la Salud).
- ❖ Sorribas Albert. Análisis de Regresión Logística: (<http://web.udl.es/Biomath/Bioestadistica/Dossiers/Doctorat/Regresion%20logistica.pdf>)