

UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA.

UNAN-LEÓN
FACULTAD DE CIENCIAS PURAS
DEPARTAMENTO DE ESTADÍSTICA Y MATEMÁTICA



TESIS

**PARA OPTAR AL TÍTULO DE:
LICENCIADO EN ESTADÍSTICA**

**APLICACIÓN DEL METODO DE REGRESIÓN LOGÍSTICA EN UN
ESTUDIO DE PLANIFICACIÓN FAMILIAR EN LA MUJER EN
EDAD FÉRTIL EN EL DISTRITO VI DE MANAGUA Y TIPITAPA**

**AUTORES: Xochilt Varela Pérez.
Yuri Maykelly Corrales Castro.**

TUTOR: Msc. Rafael Espinoza.

León, Diciembre del 2002

AGRADECIMIENTO

Queremos expresar nuestro sinceros agradecimientos a Dios, por habernos brindado mucha sabiduría e iluminar nuestros pasos.

Por otra parte agradecemos la valiosa colaboración y apoyo a la Msc. Eliana Torrez por su tiempo y apoyo desinteresado, al Dr. Arnoldo Toruño Toruño por facilitarnos parte de nuestro trabajo para dar por concluida una etapa muy importante de nuestros estudios universitarios.

Especialmente gracias al Lic. Juan Diego Solís por su gran ayuda y colaboración en nuestro trabajo.

A nuestro excelentísimo maestro Rafael Espinoza por su enseñanza, empeño, por la calidad de nuestro trabajo de esta encomiable meta.

MUCHAS GRACIAS A TODOS

DEDICATORIA

Le dedico especialmente a Dios y a la virgen María, quien estuvo conmigo en los momentos mas difíciles y angustiosos.

A mis padres Nubia Castro y Pedro Ovidio Corrales, que con sus esfuerzos, empeños y ternura, supieron realizarme como niña, adolescente y ahora como mujer.

A mis hermanos:

Jiesner Axcel Corrales Castro.

Zamara Yunieth Corrales Castro.

Dianelis Eliane Corrales Castro.

A mis amigos, que con su apoyo, realiza la culminación de mi meta:

Xochilt Varela

Brenda Castellano

Verónica Fonseca

Freddy Sequeira

A mi excelentísimo maestro Rafael Espinoza, a mis familiares que estuvieron apoyándome y pendientes de mi en estos cinco años de estudios.

Agradezco a claustro de maestro del departamento de Estadísticas y Matemáticas, que hicieron posible la formación de una profesional mas.

YURI MAIKELLY CORRALES CASTRO

DEDICATORIA

Se la dedico en primer lugar a Dios, por regalarme sabiduría quien siempre esta a mi lado en los momentos mas difíciles y angustiosos de mi vida.

La dedico especialmente a mi madre quien a puesto sus esfuerzos e ilusiones a mi educación y formación, tanto profesional como personal y espiritual para llegar a ser una exitosa persona.

A mi hermano que con su apoyo incondicional y sus consejos a sabido desempeñar excelentemente el papel de padre, por el cual me siento orgullosa de ser su hermana

Con amor a mi hermana que con sus duros consejos emprendedores y su abnegados amor me ha ayudado a poder a llegar a culminar una meta mas de mi vida.

A mis tías: Victoria Reyes y Nubia Martínez que con su cariño y apoyo condicional han sabido sustituir a mi madre en sus ausencias con gran empeño y excelentes cuidados hacia mi persona.

Xochilt Varela Pérez.

INDICE

	Pág.
I. INTRODUCCIÓN.....	1
II. OBJETIVOS.....	2
III. MARCO TEÓRICO.....	3
3.1. Modelo de Regresión Logística.	
3.2. Tipo de Modelo Aplicado al Estudio.	
3.3. Hipótesis del Modelo de Regresión Logística Múltiple.	
3.4. Hipótesis Operativas a Contrastar.	
3.5. Obtención de las Estimaciones por medio del método de Máxima Verosimilitud.	
3.6. Contraste de Hipótesis.	
3.7. Evaluación Estadística de Wald.	
3.8. Evaluación Estadística $-2\ln(l_0)$.	
3.9. Función Logística.	
3.10. Interpretación de los Parámetros del Modelo Odds Ratios.	
3.11. Análisis de los Odds Ratios en Regresión Logística.	
IV. DISEÑO METODOLÓGICO.....	12
V. RESULTADOS.....	14
VI. CONCLUSIONES.....	19
VII. RECOMENDACIONES.....	20
VIII. BIBLIOGRAFÍA.....	21
IX. ANEXOS.....	22

I. INTRODUCCIÓN

La regresión logística es uno de los instrumentos estadístico mas expresivos y versátil de que se dispone para el análisis de datos en clínica y epidemiología. Es una técnica de análisis multivariante en la que la variable dependiente o variable repuesta es una variable dicotómica y la variable o variables independientes pueden ser cualitativa o cuantitativa.

El origen de la regresión logística se remonta en la década de los sesenta con el trabajo de Walker y Duncan (1967) en que se aborda el tema de estimar la probabilidad de ocurrencia de cierto acontecimiento en función de varias variables. Su uso se universaliza desde principio de los ochenta debido especialmente a las facilidades informáticas con que se cuenta desde entonces.

En el año 2001 el departamento de salud preventiva de la UNAN- León en conjunto con el programa Alemán PROSIM realizaron en el distrito VI de Managua y en Tipitapa una encuesta entre mujeres de edad fértil para estudiar los siguientes aspectos: El uso de anticonceptivo, practicas relativas al examen del papanicolaou, maltrato de la mujer y planificación familiar. En este contexto utilizaremos el método de Regresión Logística para investigar factores influyentes como (cuantos hijos cree usted que es adecuado tener, quien debe tomar la decisión de usar anticonceptivo, edad cumplida de las encuestadas) sobre el uso de la inyección como anticonceptivo, la cual es una variable nominal cualitativa.

El método nos permitirá conocer las probabilidades de ocurrencia de uso de la inyección como anticonceptivo en presencia o ausencia de los factores en estudio. Esperamos que este trabajo facilite la toma de decisiones para las autoridades pertinentes y que además sirva de promoción para el uso de este método y su uso práctico en la ciencia de salud.

II. OBJETIVOS

OBJETIVO GENERAL

- ☞ Aplicar el método de regresión logística, en un estudio de planificación familiar en la mujer en edad fértil en el distrito VI de Managua y Tipitapa.

OBJETIVOS ESPECÍFICOS

- ☞ Describir el comportamiento de las variables involucradas en el estudio.
- ☞ Estimar los parámetros del modelo logístico, tomando como variable dependiente el uso de inyección como anticonceptivo.
- ☞ Construir el modelo de regresión logística.
- ☞ Interpretar los coeficientes en el modelo (Odds Ratio).
- ☞ Evaluar la calidad del modelo.

III. MARCO TEÓRICO

3.1 Modelo de Regresión Logística.

El método de regresión logística (RL) es uno de los instrumentos estadísticos más expresivos y versátil de que se dispone para el análisis de datos en clínica y salud pública. El objetivo primordial que resuelve esta técnica es el modelar como influye en la probabilidad de ocurrencia de un suceso, habitualmente dicotómico, la presencia o no de diversos factores de riesgos (variables independientes). También puede ser usada para estimar la probabilidad de aparición de cada uno de las posibilidades de un suceso con más de dos categorías.

La regresión logística se utiliza en modelos estadísticos en el que intervienen una o más variables independientes (cuantitativas o cualitativas) en donde la variable dependiente es tipo dicotómica donde dicha variable refleja la ocurrencia o no del suceso. Puesto que (Y) es dicotómica admitimos que asume únicamente dos valores (0 y 1), , generalmente se acepta de que

$Y= 1$ Si el hecho ocurre.

$Y= 0$ Si el hecho no ocurre.

Al construir el modelo de regresión logística la variable dependiente es de tipo dicotómico y las variables explicativas pueden ser de cualquier naturaleza. Lo que se procura mediante la regresión logística es en principio expresar la probabilidad de que ocurra el evento en función de ciertas variables que se presume relevantes o influyentes.

Con modelos de regresión logística se investigan factores causales de una determinada enfermedad o evento, a sí como determinar el valor productivo en la presencia de ciertas características. El modelo de regresión logística esta dado por la siguiente formula.

$$P(Y = 1/X_1, X_2, \dots, X_k) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_k X_k)}$$

3.2 Tipo de modelo aplicado al estudio

El modelo aplicado a nuestro estudio, es el modelo de regresión logística múltiple. Este tipo de modelo es aquel donde la ocurrencia de un determinado suceso se ve influenciado por mas de una variable independiente.

Las variables independientes en este modelo pueden ser cuantitativas y cualitativas, si estas son cualitativas con C - categorías se le hará el tratamiento C-1 variables dummy a fin de que todas las posibilidades queden bien representadas en el modelo.

Modelo Matemático

$$P(y = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_k X_k)}$$

β_0 : Es el termino independiente o constante y puede ser ig al a cero .

β_1 : Coeficiente d regresión y debe ser significativamente distinto de cero.

X_1 : Denota la variable independiente significativamente influyente.

$P(Y=1)$: Es la probabilidad que la variable dependiente tome un determinado valor.

Exp: Denota la base del logaritmo neperiano que es una de las constantes mas utilizadas.

3.3 Hipótesis del M.R.L.M.

Las hipótesis conceptuales a contrastar en este modelo son:

H_0 : Las variables independientes no influyen significativamente sobre la variable dependiente.

H_1 : Las variables independientes si influyen significativamente sobre la variable dependiente

3.4 Las hipótesis operativas a contrastar son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0$$

3.5 Obtención de las estimaciones por medio del método de máxima verosimilitud.

El problema que aborda la RL es de naturaleza tal que cada sujeto tiene o no la condición representada por la variable respuesta “Y”, el modelo atribuye al individuo una probabilidad P_i de que tenga la condición y la probabilidad complementaria $1-P_i$ de que no la tenga.

Llamemos λ_i a la probabilidad de que atribuye el método a la condición en que realmente se encuentre el sujeto i -ésimo. Si $Y_i = 1$ (El individuo efectivamente tiene la condición) entonces, como P_i es la probabilidad que el método atribuye a esa condición, se tiene que $\lambda_i = P_i$. Si, por el contrario, se cumple que $Y_i = 0$, la probabilidad estimada que no tenga la condición es $1-P_i$, es decir, $\lambda_i = 1-P_i$.

La probabilidad λ_i es una función de P_i y de Y_i , es fácil convencerse que dicha función puede expresarse de modo compacto:

$$\lambda_i = P_i^{Y_i} (1-P_i)^{1-Y_i}$$

Se definen V (La llamada función de verosimilitud) como el producto de los λ_i para toda la muestra. O sea, V se puede expresar del modo siguiente:

$$V = \prod_{i=1}^n P_i^{y_i} (1-P_i)^{1-y_i}$$

Nótese que lo ideal sería que tuviera $P_i=1$ siempre que $Y_i=1$ y $P_i=0$ siempre que $Y_i=0$ (vale decir, el método sería perfecto siempre que se cumpliera que el sujeto tiene el rasgo en estudio, entonces el procedimiento lo clasificará en dicha categoría y viceversa).

En tal caso, se tendría $\lambda_i=1$ para todo i , Y , consecuentemente la V sería igual a 1, de aquí se deduce que el máximo valor teóricamente alcanzable por V asciende también a uno.

Suponiendo fijos los valores de “ Y ”, X_1, X_2, \dots, X_k , el valor de P_i depende de quienes sean: a, b_1, b_2, \dots, b_k . Para cada conjunto de $k+1$ valores que se elijan, para ellos la magnitud de V será también diferente aunque siempre menor que uno.

Dicho en resumen para una matriz de datos fijos, V es una función de: a, b_1, \dots, b_k cuyo valor máximo es uno. De lo que se trata, es hallar aquellos valores de a, b_1, b_2, \dots, b_k . Para los cuales V se hace máximo (lo más próximo posible a la unidad).

La forma condicional para la obtención del máximo verosímil:

$$L_u = \prod_{i=1}^{m_i} P(x_i) \prod_{i=m_i+1}^n [1 - P(x_i)]$$

$P(x)$ es igual al modelo logístico:

$$P(x) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} = \frac{1}{1 + e^{-z}}$$

3.6 Contraste de hipótesis :

Las hipótesis conceptuales a contrastar en este caso son:

H_0 : Las variables independientes no influyen significativamente sobre la variable dependiente.

H_1 : Las variables independientes si influyen significativamente sobre la variable dependiente.

Para el contraste de estas hipótesis, se describen dos procedimientos distintos:

- Evaluación estadística de wald.
- Evaluación estadística del estadístico $-2\ln(I_0)$.

3.7 Evaluación estadística de wald.

Un estadístico muy utilizado para evaluar la significación estadística del M.R.L. es, el que resulta de dividir el cuadrado de b_1 por el cuadrado del error estándar de b_1 , a este estadístico lo denominaremos estadístico de wald.

$$WALD = \frac{(b_1)^2}{(EEb_1)^2}$$

Donde b_1 es el coeficiente de regresión logística muestral y $EE(b_1)$ es el error estándar de b_1 . Este estadístico se distribuye según una χ^2 cuadrado con $k-1$ grados de libertad. Si el valor experimental de wald es tal que la probabilidad de obtenerlo por azar, bajo los supuestos de la hipótesis nula, es menor que el nivel de significancia fijado en el contraste, rechazará la hipótesis nula y concluiremos diciendo que la variable independiente influye en la probabilidad de las características de la variable dependiente. El estadístico de wald se utiliza en caso que se este discutiendo si la supresión de una sola variable especifica reduce significativamente el grado de explicación en el modelo.

3.8 Evaluación del estadístico $-2\ln(l_0)$:

La evaluación de este estadístico, significa menos dos veces el logaritmo de la verosimilitud. Este es un número menor 1 y un modelo perfecto $L_0 = 1$ y $L_0(1) = 0$, entonces $-2\ln(L_0) = 0$. Además se distribuye como una ji cuadrada con $n-k-1$ grados de libertad.

Las hipótesis operativas a contrastar son:

- $H_0: L_0 = 1$
- $H_1: L_0 \neq 1$

Aceptar la nula nos indica que el modelo es significativo y por lo tanto el coeficiente β_1 es distinto de cero y el rechazar la hipótesis nula indica que β_1 es igual a cero y el modelo no es significativo.

En caso de aceptar la hipótesis nula, entonces tendremos que comprobar después mediante los estadísticos Z cual o cuales de los coeficientes deben de estar incluidos en el modelo. Recordemos que en el modelo logístico todos los coeficientes deben de ser significativos.

3.9 Función logística.

La función logística en el modelo es el principal objetivo en el caso multivariante.

Sea $Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$, para obtener el modelo logístico de la función logística, denotaremos parámetros β_1 asociados a la variable X_1 , mas β_2 asociados a la variable X_2 y así sucesivamente hasta el K -ésimo parámetro asociado a la K -ésima variable, donde las X_k denotan las variables independientes de interés y, α es el término constante y los β_i son los términos constante representando a los parámetros desconocidos.

En conclusión ,si Z denota las combinaciones de las variables independientes, sustituimos la suma lineal de Z en $f(z)$, es decir :

$$F(Z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + \exp^{(-\alpha - \sum \beta_i X_i)}} ; i = 1, \dots, k$$

En donde $f(z)$ es la función logística a utilizar.

Para explicar un poco más la regresión logística, hablaremos de la función logística $f(z)$ en donde el rango es, $0 \leq f(z) \leq 1$, esto es lo que hace tal popular a la técnica de R.L puesto que la $f(z)$ solo puede tomar valores.

El modelo se utiliza para describir una probabilidad, la cual es siempre un número entre cero y uno. Así que para el M.R.L. nunca podrá tener un valor estimado mayor que 1 ó menor que 0.

Por lo tanto el modelo es definido como logístico si la expresión para la probabilidad, dada X_i variables independientes es igual a:

$$P(Y=1/X_1, X_2, \dots, X_k) = \frac{1}{1 + \exp^{(-\alpha - \sum_{i=1}^k \beta_i X_i)}} , \quad i=1, 2, \dots, k$$

En donde α y los β_i denota los parámetros desconocidos que se estimarán basados en los datos obtenidos de los X_i y de las variables dependientes para un grupo de sujetos. Para una conveniencia denotaremos a $P(Y=1/X_1, X_2, \dots, X_k)$ como $P(Y)$.

$$P(Y=1/X_1, X_2, \dots, X_k) = \frac{1}{1 + \exp\left(-\alpha - \sum_{i=1}^k \beta_i X_i\right)}$$

3.10 Interpretación de los parámetros del modelo odds ration.

Se define el llamado odds ration a la razón de los odds correspondiente a un suceso bajo cierta condición entre los que les corresponde bajo otra.

P.R.F: Predominio de la enfermedad en presencia del factor (F), éste es igual a la razón entre la probabilidad de padecer la enfermedad en presencia del factor (F) y la probabilidad de no padecer la enfermedad en presencia del factor.

P.R.N: Predominio de la enfermedad en ausencia del factor (F), éste es igual a la razón entre la probabilidad de padecer la enfermedad en ausencia del factor (F) y la probabilidad de no padecer la enfermedad en ausencia del factor.

Los O.R indican la asociación y la influencia o no del factor y la enfermedad o cuanto mayor riesgo tiene una persona de padecer una determinada enfermedad si presente una determinada característica con respecto a otra persona con otra característica.

Los O.R son la razón de predominio mas utilizada en ciencia de la salud en problemas o estudios de casos y controles.

Si no existe relación entre el factor y la enfermedad el valor de los O.R es estadísticamente igual a uno, si el factor es de riesgo, el O.R es mayor que uno, si el factor es de protección o favorecedor el O.R es menor que uno.

En análisis los O.R es la medida simple de asociación directamente estimada de un modelo logístico en estudio de casos y controles.

3.11 Análisis de los O.R en regresión logística.

En R.L la medida d asociación más utilizada es el O.R por su sencillez de cálculo.

Si tenemos un modelo de R.L significativo, en el que una o más variables independientes es de tipo dicotómica, con valores cero y uno, el EXP del coeficiente es el O.R correspondiente al riesgo o protección que implica un aumento unitario de las variables independientes.

$O.R = \exp(b_i)$, que es una medida del riesgo relativo inherente a poseer la condición X_i a los efectos de padecer el problema $Y=1$.

IV. DISEÑO METODOLÓGICO

Los datos de esta encuesta fueron proporcionados por el Dr. Arnoldo Toruño T. coordinador del proyecto Alemán PROSIM quienes realizaron el estudio en el distrito VI de Managua y Tipitapa, por lo que los datos son secundarios.

De la información proporcionada por el estudio sólo se tomó lo relacionado al uso de anticonceptivos específicamente El **uso de la inyección como anticonceptivo**". Luego se seleccionaron, aquellas variables independientes o explicativas relacionadas con el uso de la inyección tales como: N105 años (edad cumplida de las mujeres encuestadas), cuan201 (Numero de hijos que usted cree que es adecuado tener), quien217 (quienes toman la decisión de tomar anticonceptivo) que tuvieran mas relacionadas en el desenlace de la variable dependiente "Uso de la inyección " como anticonceptivo.

Variables en el Estudio:

Variable	Etiqueta de la Variable	Codificación de la Variable
n 105 años	Edad cumplida de las mujeres en estudio	
n 207 inye	Tipo de método para evitar embarazos	0: No usa Inyección. 1: Usa Inyección
Cuan 201	Numero de hijos que usted cree que es adecuado tener	0: Numero de Hijos 1: Lo que Dios Mande 2: No sabe / No responde.
Quien 217	Quienes toman la decisión de tomar anticonceptivos	0: Si la toma el esposo. 1: Si la toma la esposa 2: Si la toma la Pareja.

Software.

Para introducir y procesar los datos y codificación de las variables se utilizó el programa estadístico spss versión 7.5 para Windows, y para el levantado de texto utilizamos Microsoft Word.

V. RESULTADOS

DESCRIPCIÓN DE LAS VARIABLES EN ESTUDIO.

Encontramos que el promedio de las edades de las mujeres en estudio fueron alrededor de 28 años por otro lado, en promedio han tenido 3 hijos por cada entrevistada.

En cuanto al uso de la inyección como anticonceptivo un 64% de las mujeres lo usan. En cuanto a la decisión de cuantos hijos quieren tener un 97.7% tiene definida la cantidad a tener, un 0.9% respondió lo Dios mande y sólo un 1.4% no sabe o no respondió (ver tabla No. 1).

Tabla No.1

Número de hijos que usted cre adecuado

		Frecuenci	Porcentaj	Porcentaj válido	Porcentaj acumulad
Válido	Número de	682	97.6	97.7	97.7
	Lo que Dios	6	.9	.9	98.6
	No sabe /No	10	1.4	1.4	100.0
	Total	698	99.9	100.0	
Perdido	Sistem	1	.1		
Total		699	100.0		

En cuanto a la toma de decisión de tomar anticonceptivos un 57% respondió que la toman en pareja, un 37.6% las mujeres y sólo un 4.4% los hombres. (ver tabla No.2)

Tabla No.2

Quienes toman la decisión de tomar

		Frecuenc	Porcenta	Porcenta válid	Porcenta acumula
Válido	Si la toma él	31	4.4	4.6	4.6
	Si la toma la	25	36.	37.	42.
	Si la toma la	38	55.	57.	100.
	Tota	67	96.	100.	
Perdid	Sistem	26	3.7		
Tota		69	100.		

En la tabla No. 3 puede observarse que las mujeres entre 15 y 24 años son las que mayor uso hacen de la inyección como anticonceptivo y el más bajo lo representan las mujeres de 45 a 49 años con 27.

Tabla No. 3

Tabla de contingencia. Uso de la inyección como anticonceptivo con relación a las edades de las mujeres en estudio.

USO DE LA INYECCION COMO ANTICONCEPTIVO			
EDAD	USA INYECCION	NO US INYECCION	TOTAL
15- 24	119	214	333
25- 34	66	110	176
35- 44	53	82	135
45- 49	27	27	54
TOTAL	265	433	698

Podemos observar que en el grupo de edades de 15-24 años son las que usan con mas frecuencia la inyección y también se observa que las mujeres que con menos frecuencia usan la inyección como anticonceptivo son las mujeres de 45-49 años de edad, observando que al aumentar la edad de la mujer es menor el uso de la inyección como anticonceptivo.

Contraste global del modelo

Entre los resultados obtenidos se computaron dos expresiones de estas funciones:

V1: Denota verosimilitud inicial

$$-2\ln (V1) = 586.4251$$

V2: Denota verosimilitud final y el resultado que arroja la RL. es igual a:

$$-2\ln (V2) = 525.075$$

El valor de ji-cuadrada con cuatro grados de libertad es igual 14.860 con un nivel de significancia 0.05 .

El estadístico de prueba o coeficiente de verosimilitud es :

$$R = (586.40251 - 525.075) / 9.57251 = 61.32751$$

Comparando este valor con el valor crítico de la ji-cuadrada con 4 grados de libertad, rechazamos la hipótesis nula del modelo lo cual nos indica que existe una fuerte influencia de las variables explicativas en las variables respuestas tomándolas en conjunto.

Tabla No 4

Resultados del modelo del uso o no uso, de la inyección como anticonceptivo.

Variables	Coeficientes	Desv. Est.	Wald	Sign.
Cuan 201			11.3745	0.0034
Cuan 201 (1)	1.874	0.8742	4.2747	0.0387
Cuan 201 (2)	-0.4172	1.4898	0.0784	0.7794
N105 años	0.0254	0.0094	7.3536	0.0067
Quien 217			11.8506	0.0027
Quien 217 (1)	1.6245	0.4807	11.4226	0.0007
Quien 217(2)	0.2881	0.2163	1.7748	0.1828

Prueba de la hipótesis sobre los coeficientes

En los resultados de la tabla número 3 bajo la columna β , encontramos el valor de los coeficientes del modelo. De ellos solamente las variables cuan201, cuan201(1), N105años, quien217, quien217 (1) aparentan ser significativamente diferente de cero usando un nivel de significancia 0.05.

Correlación parcial

En la misma tabla encontramos la correlación parcial la cual indica la contribución individual de las variables en la Regresión Logística

Observamos que por ejemplo el mayor aporte lo dan las variables cuan201, quien217, quien217 (1), n105años , y cuan201(1).

Solamente las variables cuan201(2) y quien217(2) no tienen una gran contribución al modelo.

Construcción del modelo estadístico.

Con los coeficientes de la primera columna de la tabla numero 5 del modelo podemos construir nuestro modelo de Regresión Logística el cual resulta:

$$\Pr\{Y = 1\} = \frac{1}{1 + \exp^{-1.8074(\text{cuan201}(1))+0.0254(\text{n105años})+1.6245(\text{quien217}(1))}}$$

TABLA No. 5

Resultados de los coeficientes de Regresión Logística.

Variables	Coefficiente	O.R.
CUAN 201 (1)	1.8074	0.1641
CUAN 201(2)	-0.4172	0.6589
N105 AÑOS	0.0254	1.0257
QUIEN 217 (1)	1.6245	5.0757
QUIEN 217 (2)	0.2881	1.3339

En la primera columna encontramos los valores del coeficiente del modelo logístico. Encontramos que cuando la variable cuan201 cambia de cero a uno y los valores de las otras variables independientes se mantienen constantes el uso de la inyección se va a incrementar en 1.80. Cuando la variable quien217 cambia de cero a uno se incrementará el uso de anticonceptivo en 1.62. Por el contrario cuando la variable cuan 201 cambia de 1 a 2, el uso de anticonceptivo disminuye en 0.4172.

Bajo la columna OR están los valores del factor por el cual los odds ratio cambian cuando la variable independiente incrementa en una unidad por ejemplo: Para la variable n105años el OR es igual a 1.0257 lo que indica que al aumentar en un año la edad de la mujer aumenta el uso de la inyección como anticonceptivo en 1.0257 veces más.

Para la variable Quien217(1), nos dicen que en las mujeres casadas aumenta 5.0757 veces más la decisión de el uso de anticonceptivos.

En el caso de quien 217 (2) el OR significa que en las parejas aumenta en 1.33 veces mas la decisión de usar anticonceptivos.

Las variables mas influyente en el modelo fueron: cuan201 con (OR=0.1641), N105Años(OR=1.0257), Quien217(1)(OR=5.0757).

Validación del modelo.

En esta tabla observamos que 265 mujeres que no usan la inyección fueron clasificadas correctamente por el modelo lo cual significa un 94.64 % y 16 mujeres que si la usan también fueron clasificadas correctamente. En total 280 sobre 423 son correctamente clasificadas por el modelo.

Tabla No.6

Clasificación del uso de la inyección como anticonceptivo

Valores Observados	Valores Predichos		
	No usa la inyección	Si usa la inyección	Porcentaje correcto
Si la usa	265	515	94.64%
No la usa	127	16	94.64%
Total	392	531	189.28%

VI. CONCLUSIONES

☞ Los grupos etarios de las mujeres en edad fértil de 15 a 19 años fueron las que más utilizaron el método la inyección como anticonceptivo.

☞ El Modelo obtenido fue el siguiente:

$$\Pr\{Y = 1\} = \frac{1}{1 + \exp^{-1.8074(\text{cuan201})(1) + 0.0254(\text{n105años}) + 1.6245(\text{quien217})(1)}}$$

☞ Resumen de los odds que influyeron en el modelo son: caun201 (O.R = 0.1641), n105años (O.R = 1.0257), quien217 (O.R = 5.0757).

☞ El modelo resultante es satisfactorio puesto que presenta un porcentaje de predicciones correctas de 0.664 %.

VII. RECOMENDACIONES

- ☞ Profundizar el estudio con mayores recursos hacia el conocimiento de este tema.

- ☞ Aplicar el método de regresión logística en nuevos estudios ya que es de vital importancia para el análisis de datos por que da una mejor contribución al estudio realizado.

VIII. Bibliografía.

1. Botanco Mo. Herrera Cl, Laguna JM. Análisis de Regresión Logística.
2. Dean AG, et al. Epi Info versión 6.04 : A word processing database, and statistical program for public health. Atlanta. Cent Dis. Control 1996.
3. Ewritt Bs. The Cambridge Dictionary of statistics in the medical Sciences New York Cambridge University Press 1995.
4. Ferran M. SPSS para Windows: Programación y Análisis. Estadístico Madrid. McGraw-Hill, 1996: 313-335.
5. Hosmer Dw. Lemeshow S. Applied logistic regression. New York: John Wiley S. Sons. 1989.
6. Hosmer Dw. Tuber S, Lemeshow S. The importance of assessing.
7. Piura López julio , Introducción a la Metodología de la Investigación Científica, et, El Amanecer S. A Managua-Nic. 1994.
8. Cooperación Alemana Al Desarrollo PROSIM Salud Sexual y Reproductiva en dos poblaciones de Nic. 1997

ANEXOS

Regresión logística

Dependent Variable Encoding:

Original Value	Internal Value
0	0
1	1

	Value	Freq	Parameter Coding	
			(1)	(2)
QUIEN217				
Si la toma él esposo.	.00	22	.667	-.333
Si la toma la esposa.	1.00	174	-.333	.667
Si la toma la pareja.	2.00	227	-.333	-.333
CUAN201				
Número de hijos.	.00	413	.667	-.333
Lo que Dios mande.	1.00	4	-.333	.667
No sabe /No responde.	2.00	6	-.333	-.333

Dependent Variable.. N207INYE Tipo de metodo para evitar un embarazo.(Inyecciones)

Beginning Block Number 0. Initial Log Likelihood Function

-2 Log Likelihood 586.40251

No terms in the model.

Classification Table for N207INYE
The Cut Value is .50

Observed		Predicted	Percent Correct				
			No las usa N	Si usa inyeccion I	S		
No las usa	N	I	0	I	280	I	.00%
Si usa inyeccion	S	I	0	I	143	I	100.00%
Overall							33.81%

Beginning Block Number 1. Method: Backward Stepwise (COND)

Variable(s) Entered on Step Number

- 1.. CUAN201 Número de hijos que usted cre adecuado tener.
- N105AÑOS Edad cumplida de las mujeres en estudio.
- N312CUAN Número de hijos que ha tenido la mujer.
- QUIEN217 Quienes toman la decisión de tomar anticonceptivos.

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 523.795
Goodness of Fit 420.113
Cox & Snell - R² .138
Nagelkerke - R² .183

	Chi-Square	df	Significance
Model	62.608	6	.0000
Block	62.608	6	.0000
Step	62.608	6	.0000

----- Hosmer and Lemeshow Goodness-of-Fit Test-----

Group	N207INYE = No las usa		N207INYE = Si usa inyeccion		Total
	Observed	Expected	Observed	Expected	
1	38.000	33.004	5.000	9.996	43.000
2	29.000	31.071	13.000	10.929	42.000
3	31.000	30.344	11.000	11.656	42.000
4	32.000	29.745	10.000	12.255	42.000
5	32.000	28.846	10.000	13.154	42.000
6	30.000	29.261	14.000	14.739	44.000
7	20.000	27.603	23.000	15.397	43.000
8	21.000	26.015	21.000	15.985	42.000
9	26.000	24.214	16.000	17.786	42.000
10	21.000	14.658	20.000	26.342	41.000

	Chi-Square	df	Significance
Goodness-of-fit test	18.5499	8	.0175

Classification Table for N207INYE
The Cut Value is .50

Observed		Predicted			Percent Correct
		No las usa N	Si usa inyeccion I	S	
No las usa	N	I 262	I 18	I	93.57%
Si usa inyeccion	S	I 124	I 19	I	13.29%
Overall					66.43%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R
CUAN201			10.5313	2	.0052	.1055
CUAN201(1)	-1.8288	.8741	4.3770	1	.0364	-.0637
CUAN201(2)	-.5871	1.4898	.1553	1	.6936	.0000
N105AÑOS	.0188	.0110	2.9176	1	.0876	.0396
N312CUAN	.0570	.0506	1.2712	1	.2595	.0000
QUIEN217			11.3869	2	.0034	.1122
QUIEN217(1)	1.6003	.4823	11.0093	1	.0009	.1240
QUIEN217(2)	.2780	.2168	1.6448	1	.1997	.0000

Variable	Exp(B)	95% CI for Exp(B)	
		Lower	Upper
CUAN201(1)	.1606	.0290	.8909
CUAN201(2)	.5560	.0300	10.3088
N105AÑOS	1.0189	.9972	1.0411
N312CUAN	1.0587	.9588	1.1690
QUIEN217(1)	4.9547	1.9252	12.7517
QUIEN217(2)	1.3205	.8634	2.0195

----- Model if Term Removed -----
Based on Conditional Parameter Estimates

Term Removed	Log Likelihood	-2 Log LR	df	Significance of Log LR
CUAN201	-267.705	11.615	2	.0030
N105AÑOS	-263.400	3.005	1	.0830
N312CUAN	-262.538	1.281	1	.2577
QUIEN217	-267.960	12.125	2	.0023

Variable(s) Removed on Step Number
2.. N312CUAN Número de hijos que ha tenido la mujer.

Estimation terminated at iteration number 3 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood 525.075
Goodness of Fit 420.751
Cox & Snell - R² .135
Nagelkerke - R² .180

	Chi-Square	df	Significance
Model	61.327	5	.0000

Block	61.327	5	.0000
Step	-1.281	1	.2578

Note: A negative Chi-Square value indicates that the Chi-Square value has decreased from the previous step.

----- Hosmer and Lemeshow Goodness-of-Fit Test-----

N207INYE = No las usa			N207INYE = Si usa inyeccion			Total
Group	Observed	Expected	Observed	Expected	Expected	
1	38.000	32.943	5.000	10.057	43.000	
2	34.000	31.787	9.000	11.213	43.000	
3	22.000	28.062	17.000	10.938	39.000	
4	35.000	31.664	10.000	13.336	45.000	
5	32.000	27.346	8.000	12.654	40.000	
6	28.000	27.069	13.000	13.931	41.000	
7	23.000	28.173	21.000	15.827	44.000	
8	18.000	24.204	21.000	14.796	39.000	
9	27.000	26.628	18.000	18.372	45.000	
10	23.000	16.890	21.000	27.110	44.000	

	Chi-Square	df	Significance
Goodness-of-fit test	22.7954	8	.0036

Classification Table for N207INYE
The Cut Value is .50

Observed		Predicted			Percent Correct
		No las usa N	Si usa inyeccion I	S	
No las usa	N	I 265	I 15	I 94.64%	
Si usa inyeccion	S	I 127	I 16	I 11.19%	
				Overall 66.43%	

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R
CUAN201			11.3745	2	.0034	.1121
CUAN201(1)	-1.8074	.8742	4.2747	1	.0387	-.0623
CUAN201(2)	-.4172	1.4898	.0784	1	.7794	.0000
N105AÑOS	.0254	.0094	7.3536	1	.0067	.0955
QUIEN217			11.8506	2	.0027	.1157
QUIEN217(1)	1.6245	.4807	11.4226	1	.0007	.1268
QUIEN217(2)	.2881	.2163	1.7748	1	.1828	.0000

Variable	Exp(B)	95% CI for Exp(B)	
		Lower	Upper

CUAN201(1)	.1641	.0296	.9102
CUAN201(2)	.6589	.0355	12.2162
N105AÑOS	1.0257	1.0071	1.0447
QUIEN217(1)	5.0757	1.9786	13.0205
QUIEN217(2)	1.3339	.8731	2.0380

----- Model if Term Removed -----
Based on Conditional Parameter Estimates

Term Removed	Log Likelihood	-2 Log LR	df	Significance of Log LR
CUAN201	-268.807	12.538	2	.0019
N105AÑOS	-266.482	7.888	1	.0050
QUIEN217	-268.865	12.655	2	.0018

----- Variables not in the Equation -----
Residual Chi Square 1.286 with 1 df Sig = .2568

Variable	Score	df	Sig	R
N312CUAN	1.2857	1	.2568	.0000

No more variables can be deleted or added.

