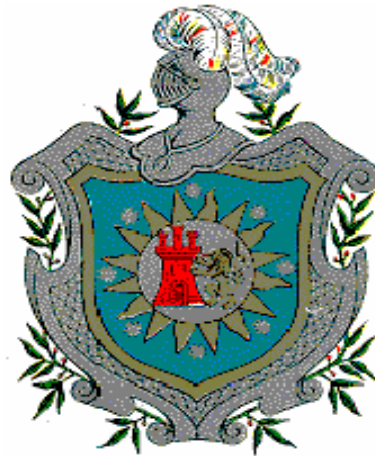




Universidad Nacional Autónoma de Nicaragua (UNAN-León)

**UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA
UNAN-LEON
FACULTAD DE CIENCIAS PURAS
DEPARTAMENTO DE ESTADISTICA-MATEMATICA**



TESIS

Para optar al título de:

LICENCIADA EN ESTADISTICA

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.

Autores: Bra. Verónica del Carmen Zapata Mojica.

Bra. Enia Sofía Barahona.

Tutor: Msc. Rafael Espinoza.

Profesor Titular

Dpto. Estadística-Matemática

León, enero de 2006



DEDICATORIA

A Dios que me guía e ilumina el camino que debo seguir y que con su amor me llena de bendiciones.

A mi madre: Maria Auxiliadora Mojica, que me dio la vida y su apoyo abnegado en mis años de preparación y de dificultad.

A mi esposo: Harold Centeno, que me brinda comprensión, seguridad y ayuda mutua para alcanzar las metas propuestas en mi vida.

Verónica del Carmen Zapata Mojica.



DEDICATORIA

Dedico esta monografía a Dios todo poderoso y a la santísima virgen María por ser la luz que me guío e iluminó durante mis años de estudio y darme la sabiduría y discernimiento para superar los momentos difíciles.

A mi madre: Ligia Sofía Barahona, por su apoyo, esfuerzo y sacrificio abnegado por escucharme y aconsejarme en mis momentos de alegría y de tristeza y por ser el pilar fundamental de mi dedicación en mis estudios universitarios.

Madre éste logro es por ti y para ti con mucho amor y cariño.

Enia Sofía Barahona.



AGRADECIMIENTOS

Las integrantes de este estudio expresamos nuestro agradecimiento a aquellas personas que colaboraron en la realización y logro de este trabajo. Especialmente a: nuestro tutor Msc. Rafael Espinoza por su apoyo científico, moral y por sus valiosas sugerencias desde el diseño del estudio hasta su culminación. A los miembros del centro de investigación en demografía y salud (CIDS) Por facilitarnos la base de datos en que se basaron las inferencias de este estudio.





INDICE

| | |
|----------------------|------|
| Contenido | Pág. |
| Dedicatoria | |
| Agradecimientos | |
| I Introducción | 1 |
| II Justificación | 2 |
| III Objetivos | 3 |
| IV Marco teórico | 4 |
| V Material y métodos | 15 |
| VI Resultados | 18 |
| VII Conclusiones | 21 |
| VIII Recomendaciones | 22 |
| IX Referencias | 23 |
| X Anexos | 24 |



INTRODUCCION

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

En el año 2004 el centro de investigación en demografía y salud (CIDS), de la UNAN-León realizaron en las áreas urbanas de las ciudades de Estelí, León y Juigalpa una encuesta a jóvenes de 14 a 25 años de edad para determinar cuales son los niveles de base en la población de estudio-seguimiento en relación a sus conocimientos, actitudes y normas sociales.

El propósito de este trabajo es investigar la influencia de las variables edad, sexo, estado civil, nivel académico, religión, quien debería decidir si usan o no condón, quien tomo la decisión de usar o no usar condón, tomo la decisión el hombre de usar o no condón en el uso de anticonceptivo. Para la determinación de estas variables influyentes, la técnica estadística aplicada para resolver este problema es la regresión logística. Lo que se pretende en este trabajo monográfico es dar a conocer bajo que condiciones se aplica esta técnica, sus fundamentos teóricos prácticos, así como la interpretación de las salidas proporcionada por paquetes estadísticos en la aplicación de este método, Por lo que este trabajo permitirá servir de apoyo para futuros estudios en donde la técnica estadística a aplicar sea la regresión logística.



JUSTIFICACIÓN

Es bien conocido que la estadística constituye una herramienta básica para estudios epidemiológicos y es de gran utilidad para los tomadores de decisiones, profesionales de salud y educadores que necesitan estar concientes de las debilidades de las actuales fuentes de datos, y de las limitaciones de los análisis descriptivos que carecen de indicadores que reflejan una relación causa-efecto entre una variable dependiente y dos o mas variables independientes además, el análisis de regresión logística hace posible controlar los factores de confusión . Por lo tanto con este estudio se pretende brindar información útil a ser considerada por tomadores de decisiones en la elaboración de programas de prevención de enfermedades de transmisión sexual y de embarazos no deseados.



OBJETIVOS

OBJETIVO GENERAL.

- Aplicar el método 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.

OBJETIVOS ESPECIFICOS

- Describir las variables involucradas en el estudio.
- Estimar los parámetros del modelo logístico.
- Interpretación de los parámetros estimados.
- Determinar la bondad de ajuste del modelo.
- Calibrar el modelo ajustado.
- Obtener el modelo global.



MARCO TEORICO

REGRESION LOGISTICA

La regresión logística es uno de los instrumentos estadísticos más expresivos y versátiles de que se dispone para el análisis de datos en clínica y epidemiología.

La regresión logística permite la relación de una variable dependiente categórica Y con varias variables independientes cuantitativas y/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 resuelve esta técnica es de modelar como 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 mas de dos categorías.

Denotaremos a la variable dependiente Y, que refleja la ocurrencia o no del suceso y a la variable independiente X, que indica 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 presumen relevantes o influyentes. 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 llamada 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.

e: es la función exponencial.

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

x_1, x_2, \dots, x_k : Denotan las variables independientes.

ESTIMACION DE LOS PARAMETROS DEL MODELO

Los parámetros $\beta_0, \beta_1, \dots, \beta_k$ del modelo de regresión logística se estiman con el método de máxima verosimilitud por que al ser los residuales dicotómicos se obtienen estimaciones mas eficientes que con el método de mínimos cuadrados.

El valor de la función de verosimilitud $L(\beta)$ de un modelo estimado representa la probabilidad de reproducir los datos a partir de dicho modelo:

$$\zeta(x_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (1)$$

Las funciones de verosimilitud se expresan en forma de un producto de probabilidades que dependen del valor de los parámetros a estimar:

$$L(\beta) = \prod_{i=1}^n \zeta(x_i) \quad (2)$$

Donde el logaritmo neperiano de la verosimilitud:

$$L(\beta) = \ln[L(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (3)$$

El método de estimación que se usa en el caso multivariante es el de máxima verosimilitud la función de verosimilitud es idéntica a la de la ecuación (2) con la diferencia que ahora se definen $\pi(x)$ que está expresada como:



$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Donde $g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ las ecuaciones resultantes son igual a cero:

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

Y

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \text{ Donde } j = 1, 2, \dots, p.$$

La solución de este sistema tiene que ser con métodos aproximativos. Así la ecuación de verosimilitud requiere especial tratamiento en donde β estimado denota la solución de ecuaciones lineales.

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 encuentran el $-2LL(0)$, $-2LL(\beta)$, 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 ajustado por el modelo.

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

$-2LL(\beta)$ indica la verosimilitud del modelo con las variables predictoras.

Este valor puede ser contemplado como un descriptivo de la bondad de ajuste de este modelo, que cuanto mas 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 sólo la constante.

$L(\beta)$ es la verosimilitud del modelo considerado.

Por otra parte, este índice tiene el 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:

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

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 cociente de regresión.

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

Sigue una ley normal estandarizada con el cuadrado del cociente llamado estadístico de wald que sigue una ley de ji-cuadrado con 1 grado de libertad.

$$\chi^2_{\text{wald}} = \left[\frac{\beta}{SE(\beta)} \right]^2 \longrightarrow \chi_1^2$$

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 utiliza en caso que se este 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 recomiendan que en su lugar se aplique una prueba de razón de verosimilitud (LR). La prueba de razón de verosimilitud 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:

$$-2 \ln LR = -2 \ln \frac{L(0)}{L(M)} = -2LL(0) - [-2LL(M)] \rightarrow \chi_1^2$$

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

CALIBRACIÓN DEL MODELO: 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” observaciones sujetas tendremos $J \leq n$ estimaciones (π_i) diferentes. Sin embargo, puesto que en la mayor parte de los estudios el número J de combinaciones diferente es grande, Hosmer y Lemeshow proponen ordenar los “n” sujetos según las predicciones π_i y dividirlos en $g = 10$ grupos de aproximadamente igual tamaño (llamados deciles de riesgo). Esta división de grupos 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 con el estadístico Ji – cuadrado de bondad de ajuste:

$$\chi^2 = \sum_{J=1}^2 \sum_{i=1}^g \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \longrightarrow \chi_8^2$$



La significancia de este estadístico puede ser calculada con la ley de Ji – cuadrado con $g - 2 = 8$ grados de libertad si la mayor parte de las frecuencias esperadas son superiores a 5 ($e_{ij} > 5$) y ninguna de ellas es inferior a 1.

INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESION LOGISTICA

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 prevalencias y si se trata de un estudio de casos y controles las probabilidades estiman una proporción que carece 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 la variable predictora X se incrementa en una unidad ($X \rightarrow X+1$) 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 constante todas las variables predictoras excepto la variable X sus términos aparecerán tanto en el numerador como en el denominador del anterior cociente y se podrán simplificar de manera que OR será igual al e^{β} . Así pues, e^{β} es el factor por el cual se multiplica la odds de la respuesta (Y) cuando la variable predictora X 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 decrementa la probabilidad de respuesta $Y=1$.



El intervalo de confianza $1-\alpha$ de RO que mide el efecto de un incremento de la variable X:

$$\hat{OR} \cdot e^{\pm Z_{\alpha/2} \hat{se}(\beta)}$$

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

exp.: es la base del logaritmo natural.

Z: denota el nivel de significancia fijado.

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 el 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 cuando 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.

Intervalo de confianza para la estimación del riesgo relativo es:

$$\hat{RR} \cdot e^{\pm z_{\alpha/2} \hat{se}(\beta)}$$

RR: es la estimación puntual del riesgo relativo.

e: es la base del logaritmo natural.

Error estándar = $\hat{se}(\beta)$



La interpretación del riesgo relativo se efectúa de la siguiente manera: si el riesgo relativo es igual a uno, el factor de riesgo no tiene influencia en la aparición de la enfermedad. Si al construir el intervalo de confianza no incluye el valor 1 concluiremos que el riesgo es estadísticamente significativo.

En cambio si el riesgo relativo fuese menor que 1 y también su intervalo de confianza, estaremos ante la presencia de un factor de protección.

RAZON DE DOS ODDS

Se define como odds ratio: a la razón entre los dos odds correspondientes a un suceso bajo cierta condición entre los que les corresponden bajo otra.

El odds ratio constituye otra forma de cuantificar la asociación entre dos variables dicotómicas y no varía aunque se cambie el orden de las categorías en cualquiera de las variables lo que no ocurre con el riesgo relativo. Los odds ratio (OR) expresan la probabilidad o riesgo de padecer una enfermedad estando expuesto a cierto factor y la probabilidad de padecer la enfermedad en ausencia del factor.

$$OR = \frac{PRF}{PRN} = \frac{\frac{P_F^{(E)}}{1 - P_F^{(E)}}}{\frac{P_{\bar{F}}^{(E)}}{1 - P_{\bar{F}}^{(E)}}}$$

PRF: Predominio de la enfermedad en presencia del factor (F); este 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 (F).

PRN: Predominio de la enfermedad en ausencia del factor (\bar{F}); este es igual a la razón entre la probabilidad de padecer la enfermedad en ausencia del factor (\bar{F}) y la probabilidad de no padecer la enfermedad en ausencia del factor (\bar{F}).



El OR indica la asociación y la influencia entre la o las variables independientes (factores de riesgo) y la dependiente (enfermedad o evento). Constituye además la medición de riesgo más utilizada en las Ciencias de la Salud. Si no existe relación entre el factor y la enfermedad el valor del OR es igual a 1; se dice que es factor de riesgo si el OR es mayor que 1; y si el OR es menor que 1 entonces se considera factor de protección.

PREDICCIÓN CON EL MODELO

La ecuación de regresión logística se expresa como:

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

Esta ecuación permite estimar las probabilidades de que ocurra el hecho para un determinado patrón de valores de las variables predictoras X_i . Esta probabilidad siempre depende del tipo de estudio que se utilice.

TIPOS DE ESTUDIOS EPIDEMIOLOGICOS

Estudio de Cohorte(o de seguimiento)

Un estudio de cohorte es un estudio epidemiológico, observacional, analítico, longitudinal prospectivo, en el que los individuos que componen los grupos de estudio se seleccionan en función de la presencia de una determinada característica o exposición. Estos individuos no tienen la enfermedad de interés y son seguidos durante un cierto periodo de tiempo para observar la frecuencia con que la enfermedad aparece en cada uno de los grupos. Con este estudio se obtienen estimaciones más directas y una mejor valoración del estado de exposición. También se llama estudio de seguimiento, de proyección o de incidencia. La razón entre las probabilidades (riesgo) estimadas para dos patrones



diferentes x_1 y x_0 de valores de las variables predictoras permite estimar directamente el riesgo relativo (RR):

$$RR = \frac{\hat{P}_r(Y = 1/X_1)}{\hat{P}_r(Y = 1/X_0)}$$

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):

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

Estudio de Casos y Controles

En este tipo de estudio partiendo de la enfermedad se estudian sus antecedentes en el que se seleccionan un grupo de personas con una enfermedad (u otra

Variable de interés) y se compara con un grupo que no tenga la enfermedad, denominados casos y controles. En este diseño no se puede calcular el riesgo relativo pero si el odds ratio:

$$OR = e^{\beta_i}$$

VARIABLES CATEGORICAS.

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 cual de las categorías de la variable 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.

LOS TRES PRINCIPALES METODOS DE CODIFICACION DE LAS VARIABLES CATEGORICAS.

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.

Codificación respecto al promedio de las categorías.

Se usa cuando se desea utilizar como referencia el efecto conjunto definido por la respuesta promedio de todos los sujetos del estudio. El coeficiente de regresión b de cada variable ficticia refleja el contraste entre la categoría que representa la variable ficticia y el promedio del conjunto de las k categorías.

Codificación mediante polinomios ortogonales se utiliza solo para variables ordinales:

El coeficiente de regresión b de la primera variable ficticia refleja la tendencia lineal del conjunto de las categorías. El coeficiente b de la segunda variable ficticia refleja la tendencia cuadrática y así sucesivamente.



MATERIAL Y MÉTODO

Tipo de estudio:

Estudio analítico de corte transversal.

Área o población de estudio:

Una encuesta poblacional fue llevada a cabo durante el periodo comprendido entre septiembre y diciembre del 2004, en los municipios de Estelí, León, Juigalpa.

Definición del caso:

En este estudio, se definieron casos solamente a aquellos jóvenes entre 14 a 25 años de edad que usaron como anticonceptivo el condón en sus relaciones.

Fuente de datos:

La base de datos fue proporcionada por el centro de investigación en demografía y salud (CIDS).

Selección del modelo de Regresión Logística:

En primer paso consistió en seleccionar de la base de datos original, todas aquellas variables independientes o explicativas que estuvieron mas relacionadas en el desenlace de la variable dependiente “uso de anticonceptivo” .El segundo consiste 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 con la variable dependiente (se asigno 1 si y 0 en caso contrario). En el modelo logístico solamente fueron incluidas aquellas variables independientes con significancia estadística.



Operacionalización de las Variables

| VARIABLES | CONCEPTO | ESCALAS |
|------------------|---|---|
| Edad | ¿Cuántos años cumplidos tenés? | |
| Rsexo | Sexo de el/la entrevistado(a) | masculino femenino |
| Rcivil | ¿Actualmente estás casado(a), acompañado(a), con pareja pero no viven juntos, jalando o no? | Con pareja(casado, acompañado,) Sin pareja(jalando, no jalando) |
| Rnivacad | ¿Cuál fue el último grado o año que aprobaste? | Ningún nivel, primaria, secundaria, técnico básico, técnico medio Técnico superior universitario o mas |
| Religión | ¿Prácticas alguna religión? | No si |
| Rdccn | ¿Quién debería decidir si usan o no condón? | Hombre, mujer, no sabe Ambos |
| Tduc | ¿Quién tomó la decisión de usar o no usar condón? | Yo, mi pareja, no lo hablamos Ambos |
| Tomdh | Tomó decisión el hombre de usar o no condón. | si no |



Pruebas de significancia estadística:

Como prueba de significancia estadística se utilizó la prueba de Wald y el intervalo de confianza del 95% (IC95%) en la prueba de Wald se consideró que había significancia estadística con un valor de $p \leq 0.05$. En caso que el IC del 95% incluyera el valor de "1" entonces la variable no fue estadísticamente significativa, en caso que dicho intervalo excluyera el valor de "1" entonces se considera significativa la variable.

Medición del riesgo:

La medición del riesgo se realizó a través del odd ratio. Se considera factor de protección cuando todos los valores incluidos en el IC 95% fueron menores de uno. Si todos los valores del 95% son mayores que uno entonces se dice que la variable considerada es un factor de riesgo.

Software:

Para introducir, procesar los datos y codificar las variables se utilizó el paquete estadístico SPSS versión 11.5 para Windows, y para el levantado de texto utilizamos el programa Microsoft Word.



RESULTADOS

La población en estudio fue de 1986 jóvenes con edades comprendidas entre 14 y 25 años, obteniéndose un promedio de edad de 20 años. El 55.5% practica alguna religión. En cuanto al nivel académico el 79.7% están cursando la educación media y el 20.3% cursa la educación superior. La distribución por sexo es de 56.8% de hombres y 43.2% de mujeres. El 56.2% tiene pareja y el 43.8% no tiene pareja. En cuanto a quien debería decidir si usan o no el condón como anticonceptivo el 75% dijo que fueran ambos y el 25% opinó que fuera una de los dos un 50.4% dijo que debería de ser uno de los dos quien tome la decisión de usar o no el condón y el 49.6% dice que deberían de ser ambos. El 56.3% dijo que es el hombre el que toma la decisión de usar o no el condón como anticonceptivo y un 52.2% responden que si usan el condón como método anticonceptivo.

Selección de las variables

El interés del modelo a plantear es establecer la relación de uso o no del condón como anticonceptivo en función de las variables: edad, sexo, estado civil, nivel académico, religión, quien debería decidir si usan o no condón, quien tomó la decisión de usar o no condón, tomó la decisión el hombre de usar o no condón. Aplicando el método de regresión logística "Adelante RV" fueron seleccionadas las variables mas relevantes para obtener un modelo mas simple a través del procedimiento de máxima verosimilitud resultando incluidas en el modelo las siguientes variables: Con pareja o sin pareja (estado civil), cual fue el ultimo grado o año que aprobaste (nivel académico), quien tomo la decisión de usar o no usar condón. Dichas variables incluidas estuvieron estadísticamente relacionadas a la variable respuesta.



Estimación de los parámetros

En la estimación de los parámetros se usa el método de máxima verosimilitud en nuestro estudio se obtuvieron los siguientes valores estimados: $\hat{\beta}_0 = -0.606$, $\hat{\beta}_1 = 0.226$, $\hat{\beta}_2 = 0.372$, $\hat{\beta}_3 = 1.186$ que son los coeficientes del modelo

Bondad de ajuste del modelo estimado:

La prueba de bondad de ajuste se llevó a cabo a través de los estadísticos R^2 de cox y snell y el R^2_c de negelkerke donde el $R^2_c = 0.116$ que nos muestran que el ajuste del modelo fue bajo siendo normal en este tipo de estudio ya que solo explica un 12% de la incertidumbre de los datos.

La Prueba de ómnibus sobre los coeficientes del modelo nos muestra la significancia global que indica que el modelo estimado es significativo ($\chi^2 = 180.944$; gl = 3; p = 0.000). La prueba de razón de verosimilitud nos indica que las variables predictoras estado civil, nivel académico, quien tomó la decisión de usar o no condón son significativas en el modelo en estudio.

Significancia de los parámetros del modelo:

Es evaluada por el estadístico de wald que nos indica que las variables independientes estado civil, nivel académico, quien tomó la decisión de usar el condón influyen en la probabilidad de la variable dependiente “uso del anticonceptivo” (ver tabla 1).

Calibración del modelo: prueba de bondad de ajuste de Hosmer y Lemeshow.

En la calibración del modelo de Hosmer y Lemeshow se realizaron 3 pasos resultando en el segundo paso el estadístico ji-cuadrado casi nulo y con un grado de significancia de p = 0.99 próximo a uno, indicando un excelente ajuste.



Interpretación de los coeficientes de regresión logística.

Tabla # 1

| Factores de Riesgo | Coeficiente | Error Estándar | X ² Wald | Valor de P | Odds Ratio | Intervalo de Confianza (95%) | |
|--|-------------|----------------|---------------------|------------|------------|------------------------------|----------|
| | | | | | | Inferior | Superior |
| Estado civil | 1.186 | .096 | 151.266 | .000 | 3.274 | 2.710 | 3.954 |
| Nivel académico | .372 | .119 | 9.755 | .002 | 1.451 | 1.149 | 1.833 |
| Quien tomó la decisión de usar o no condón | .226 | .095 | 5.618 | .018 | 1.254 | 1.040 | 1.512 |
| Constante | -.606 | .078 | 59.952 | .000 | .546 | | |

Los jóvenes que tienen pareja tuvieron un OR = 3.27 (IC 95%: 2.71-3.95), o sea que los jóvenes comprendidos en este grupo tienen tres veces mayor probabilidad de usar el condón como anticonceptivo que los que no tienen pareja. El nivel académico tuvo un OR = 1.45 (IC 95%: 1.15-1.83), es decir que los jóvenes que cursan el ciclo básico tienen un mayor probabilidad de usar el método anticonceptivo que los jóvenes que llevan estudios superiores. En cuanto a, quien tomo la decisión de usar o no usar condón se muestra un OR = 1.25 (IC 95%: 1.04-1.51), donde las personas que opinaron que deberían ser ambos tienen mayor probabilidad de uso del método anticonceptivo que cuando lo decide uno solo.

Modelo de regresión logística:

Con respecto al valor predictivo del modelo este fue de 63.8% que es la capacidad discriminante del modelo lo que nos indica que el modelo de regresión logística es adecuado en el estudio. Así por lo tanto el modelo resultante es.

$$P(Y = 1/X) = \frac{1}{1 + \exp^{[-0.606 + 0.226TDUC + 0.372RNIVACAD + 1.186RCIVIL]}}$$



CONCLUSIONES

- Las variables explicativas en el estudio fueron: estado civil, nivel académico y quien tomó la decisión de usar o no usar el condón.
- La variable más representativa en nuestro estudio fue estado civil con un OR = 3.274, o sea que los jóvenes con pareja estable tiene tres veces mayor riesgo a usar el método anticonceptivo que los que no tienen pareja.
- Como los coeficientes $\beta_1 = 1.186$, $\beta_2 = 0.372$, $\beta_3 = 0.226$ son positivos, las variables aumentan la probabilidad del suceso que estamos estudiando (uso del anticonceptivo).
- La bondad de ajuste del modelo es baja, siendo normal en este tipo de estudios.
- La calibración del modelo es significativa, indicando un ajuste adecuado.
- El modelo resultante es satisfactorio a pesar que los R^2 de la prueba de bondad de ajuste fueron bajos.

$$P(Y = 1/X) = \frac{1}{1 + \exp^{[-0.606 + 0.226TDUC + 0.372RNIVACAD + 1.186RCIVIL]}}$$



RECOMENDACIONES

- Sugerimos la aplicación del modelo de regresión logística para abordar los problemas de salud o de otra clase ya que constituyen un poderoso instrumento para analizar e interpretar la medición del riesgo y el perfil de dichos problemas, que de otra manera sería imposible de determinar.

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



REFERENCIA.

➤ **Excursión a la regresión logística en ciencias de la salud.**

Luis Carlos Silva Aycaguer, Ed. Díaz de Santos, Madrid 1995.

➤ **Métodos Multivariantes en Bioestadística**

Víctor Abraira Santos, Alberto Pérez de Vargas Luque, Ed. Centro de Estudios Ramón Areces.

➤ **Análisis Multivariante: Modelo de Regresión**

José M. Doménech Massons

➤ **Metodología de la investigación : Determinación de factores de riesgo**

Pita Fernández S., Vila Alonso MT., Carpena Montero J., Unidad de epidemiología, clínica y bioestadística, Complejo Hospitalario Juan Canalejo, A Coruña (España) Cad Aten Primaria 1997.

➤ **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 la Causalidad de Lesiones por Accidentes de Tráfico en niños menores de 15 años del municipio de León, durante 1995 (Tesis).**

Gloria Esmeralda Duarte Chávez., Juan Antonio Pérez Rivera, Est. 378.2 D812a 2002 c.1.

➤ **Paquetes Estadísticos SPSS 8.0 Bases Teóricas. Practicas propuestas, resueltas y comentadas.**

Quintín Martín Martín., Maria Teresa Cabero Morán., Ramón Ardanuy Albajar, 005.3 M 382p c.2.



ANEXOS



Tabla de frecuencia

Cuantos años cumplidos tenés

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|-------|------------|------------|-------------------|----------------------|
| Válidos | 14 | 74 | 3,7 | 3,7 | 3,7 |
| | 15 | 83 | 4,2 | 4,2 | 7,9 |
| | 16 | 123 | 6,2 | 6,2 | 14,1 |
| | 17 | 159 | 8,0 | 8,0 | 22,1 |
| | 18 | 183 | 9,2 | 9,2 | 31,3 |
| | 19 | 183 | 9,2 | 9,2 | 40,5 |
| | 20 | 205 | 10,3 | 10,3 | 50,9 |
| | 21 | 204 | 10,3 | 10,3 | 61,1 |
| | 22 | 202 | 10,2 | 10,2 | 71,3 |
| | 23 | 200 | 10,1 | 10,1 | 81,4 |
| | 24 | 204 | 10,3 | 10,3 | 91,6 |
| | 25 | 166 | 8,4 | 8,4 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Prácticas alguna religión

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|-------|------------|------------|-------------------|----------------------|
| Válidos | no | 884 | 44,5 | 44,5 | 44,5 |
| | si | 1102 | 55,5 | 55,5 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Actualmente estas casado(a), acompañado(a), con pareja pero no viven juntos, jalando o no

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|------------|------------|------------|-------------------|----------------------|
| Válidos | con pareja | 1116 | 56,2 | 56,2 | 56,2 |
| | sin pareja | 870 | 43,8 | 43,8 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |



Cual fue el ultimo grado o año que aprobaste.

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|-------------------------|------------|------------|-------------------|----------------------|
| Válidos | anal-prim-secun-tecnico | 1582 | 79,7 | 79,7 | 79,7 |
| | tec.sup-universitario | 404 | 20,3 | 20,3 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Sexo de el/la entrevistado(a)

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|--------|------------|------------|-------------------|----------------------|
| Válidos | hombre | 1128 | 56,8 | 56,8 | 56,8 |
| | mujer | 858 | 43,2 | 43,2 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Quien tomó la decision de usar o no usar condón.

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|----------------------------------|------------|------------|-------------------|----------------------|
| Válidos | yo, mi pareja, no lo hablamos | 1001 | 50,4 | 50,4 | 50,4 |
| | ambos | 985 | 49,6 | 49,6 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Quien deberia decidir si usan o no condón.

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|------------------------|------------|------------|-------------------|----------------------|
| Válidos | hombre, mujer, no sabe | 497 | 25,0 | 25,0 | 25,0 |
| | ambos | 1489 | 75,0 | 75,0 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |

Tomó la decisión el hombre de usar o no condón

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|-----------|------------|------------|-------------------|----------------------|
| Válidos | otros | 867 | 43,7 | 43,7 | 43,7 |
| | el hombre | 1119 | 56,3 | 56,3 | 100,0 |
| | Total | 1986 | 100,0 | 100,0 | |



Regresión logística

Resumen del procesamiento de los casos

| Casos no ponderados ^a | | N | Porcentaje |
|----------------------------------|--------------------------|------|------------|
| Casos seleccionados | Incluidos en el análisis | 1986 | 100,0 |
| | Casos perdidos | 0 | ,0 |
| | Total | 1986 | 100,0 |
| Casos no seleccionados | | 0 | ,0 |
| Total | | 1986 | 100,0 |

a. Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

Codificación de la variable dependiente

| Valor original | Valor interno |
|----------------|---------------|
| no usa | 0 |
| si usa | 1 |

Bloque 0: Bloque inicial

Tabla de clasificación ^{a,b}

| Observado | | | Pronosticado | | |
|-----------|----------|-------------------|--------------|--------|---------------------|
| | | | U_CONDON | | Porcentaje correcto |
| | | | no usa | si usa | |
| Paso 0 | U_CONDON | no usa | 0 | 949 | ,0 |
| | | si usa | 0 | 1037 | 100,0 |
| | | Porcentaje global | | | 52,2 |

a. En el modelo se incluye una constante.

b. El valor de corte es ,500

Variables en la ecuación

| | B | E.T. | Wald | gl | Sig. | Exp(B) |
|------------------|------|------|-------|----|------|--------|
| Paso 0 Constante | ,089 | ,045 | 3,897 | 1 | ,048 | 1,093 |



VARIABLES QUE NO ESTÁN EN LA ECUACIÓN

| | | | Puntuación | gl | Sig. |
|-----------------------|-----------|----------|------------|----|------|
| Paso 0 | Variables | EDAD | 26,643 | 1 | ,000 |
| | | RSEXO | 32,633 | 1 | ,000 |
| | | RCIVIL | 160,051 | 1 | ,000 |
| | | RNIVACAD | 7,203 | 1 | ,007 |
| | | RELIGION | 1,294 | 1 | ,255 |
| | | TDUC | 17,589 | 1 | ,000 |
| | | RDCCN | 8,743 | 1 | ,003 |
| | | TOMDH | 1,918 | 1 | ,166 |
| Estadísticos globales | | | 180,805 | 8 | ,000 |

BLOQUE 1: MÉTODO = POR PASOS HACIA ADELANTE (RAZÓN DE VEROSIMILITUD)

PRUEBAS OMNIBUS SOBRE LOS COEFICIENTES DEL MODELO

| | | Chi-cuadrado | gl | Sig. |
|--------|--------|--------------|----|------|
| Paso 1 | Paso | 162,814 | 1 | ,000 |
| | Bloque | 162,814 | 1 | ,000 |
| | Modelo | 162,814 | 1 | ,000 |
| Paso 2 | Paso | 12,518 | 1 | ,000 |
| | Bloque | 175,332 | 2 | ,000 |
| | Modelo | 175,332 | 2 | ,000 |
| Paso 3 | Paso | 5,612 | 1 | ,018 |
| | Bloque | 180,944 | 3 | ,000 |
| | Modelo | 180,944 | 3 | ,000 |



Resumen de los modelos

| Paso | -2 log de la verosimilitud | R cuadrado de Cox y Snell | R cuadrado de Nagelkerke |
|------|----------------------------|---------------------------|--------------------------|
| 1 | 2586,466 ^a | ,079 | ,105 |
| 2 | 2573,948 ^a | ,084 | ,113 |
| 3 | 2568,336 ^a | ,087 | ,116 |

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

Prueba de Hosmer y Lemeshow

| Paso | Chi-cuadrado | gl | Sig. |
|------|--------------|----|------|
| 1 | ,000 | 0 | . |
| 2 | ,025 | 2 | ,987 |
| 3 | 2,784 | 4 | ,595 |

Tabla de contingencias para la prueba de Hosmer y Lemeshow

| | | U_CONDON = no usa | | U_CONDON = si usa | | Total |
|--------|---|-------------------|----------|-------------------|----------|-------|
| | | Observado | Esperado | Observado | Esperado | |
| Paso 1 | 1 | 673 | 673,000 | 443 | 443,000 | 1116 |
| | 2 | 276 | 276,000 | 594 | 594,000 | 870 |
| Paso 2 | 1 | 543 | 542,361 | 324 | 324,639 | 867 |
| | 2 | 130 | 130,639 | 119 | 118,361 | 249 |
| | 3 | 237 | 237,639 | 478 | 477,361 | 715 |
| | 4 | 39 | 38,361 | 116 | 116,639 | 155 |
| Paso 3 | 1 | 345 | 340,324 | 181 | 185,676 | 526 |
| | 2 | 198 | 202,480 | 143 | 138,520 | 341 |
| | 3 | 130 | 130,196 | 119 | 118,804 | 249 |
| | 4 | 108 | 117,732 | 220 | 210,268 | 328 |
| | 5 | 129 | 119,464 | 258 | 267,536 | 387 |
| | 6 | 39 | 38,804 | 116 | 116,196 | 155 |



Tabla de clasificación

| Observado | | | Pronosticado | | |
|-----------|-------------------|--------|--------------|--------|---------------------|
| | | | U_CONDON | | Porcentaje correcto |
| | | | no usa | si usa | |
| Paso 1 | U_CONDON | no usa | 673 | 276 | 70,9 |
| | | si usa | 443 | 594 | 57,3 |
| | Porcentaje global | | | | 63,8 |
| Paso 2 | U_CONDON | no usa | 673 | 276 | 70,9 |
| | | si usa | 443 | 594 | 57,3 |
| | Porcentaje global | | | | 63,8 |
| Paso 3 | U_CONDON | no usa | 673 | 276 | 70,9 |
| | | si usa | 443 | 594 | 57,3 |
| | Porcentaje global | | | | 63,8 |

a. El valor de corte es ,500

Variables en la ecuación

| | B | E.T. | Wald | gl | Sig. | Exp(B) | . 95,0% para EXP(| | |
|---------------------|-----------|-------|------|---------|------|--------|-------------------|----------|-------|
| | | | | | | | Inferior | Superior | |
| Paso 1 ^a | RCIVIL | 1,185 | ,095 | 155,074 | 1 | ,000 | 3,270 | 2,713 | 3,940 |
| | Constante | -,418 | ,061 | 46,717 | 1 | ,000 | ,658 | | |
| Paso 2 ^b | RCIVIL | 1,211 | ,096 | 159,455 | 1 | ,000 | 3,356 | 2,781 | 4,050 |
| | RNIVACA | ,415 | ,118 | 12,368 | 1 | ,000 | 1,514 | 1,201 | 1,907 |
| | Constante | -,513 | ,067 | 58,107 | 1 | ,000 | ,599 | | |
| Paso 3 ^c | RCIVIL | 1,186 | ,096 | 151,266 | 1 | ,000 | 3,274 | 2,710 | 3,954 |
| | RNIVACA | ,372 | ,119 | 9,755 | 1 | ,002 | 1,451 | 1,149 | 1,833 |
| | TDUC | ,226 | ,095 | 5,618 | 1 | ,018 | 1,254 | 1,040 | 1,512 |
| | Constante | -,606 | ,078 | 59,952 | 1 | ,000 | ,546 | | |

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

b.Variable(s) introducida(s) en el paso 2: RNIVACAD.

c.Variable(s) introducida(s) en el paso 3: TDUC.



Modelo si se elimina el término

| Variable | Log verosimilitud del modelo | Cambio en -2 log de la verosimilitud | gl | Sig. del cambio |
|---------------|------------------------------|--------------------------------------|----|-----------------|
| Paso 1 RCIVIL | -1374,640 | 162,814 | 1 | ,000 |
| Paso 2 RCIVIL | -1371,022 | 168,096 | 1 | ,000 |
| RNIVACAD | -1293,233 | 12,518 | 1 | ,000 |
| Paso 3 RCIVIL | -1363,573 | 158,809 | 1 | ,000 |
| RNIVACAD | -1289,095 | 9,854 | 1 | ,002 |
| TDUC | -1286,974 | 5,612 | 1 | ,018 |

Variables que no están en la ecuación

| | | | Puntuación | gl | Sig. |
|--------|-----------|-----------------------|------------|----|------|
| Paso 1 | Variables | EDAD | 1,830 | 1 | ,176 |
| | | RSEXO | ,003 | 1 | ,958 |
| | | RNIVACAD | 12,449 | 1 | ,000 |
| | | RELIGION | ,024 | 1 | ,876 |
| | | TDUC | 8,291 | 1 | ,004 |
| | | RDCCN | 7,914 | 1 | ,005 |
| | | TOMDH | 2,564 | 1 | ,109 |
| | | Estadísticos globales | 22,349 | 7 | ,002 |
| Paso 2 | Variables | EDAD | ,110 | 1 | ,741 |
| | | RSEXO | ,058 | 1 | ,809 |
| | | RELIGION | ,049 | 1 | ,826 |
| | | TDUC | 5,626 | 1 | ,018 |
| | | RDCCN | 4,561 | 1 | ,033 |
| | | TOMDH | 2,026 | 1 | ,155 |
| | | Estadísticos globales | 9,941 | 6 | ,127 |
| Paso 3 | Variables | EDAD | ,023 | 1 | ,880 |
| | | RSEXO | ,182 | 1 | ,669 |
| | | RELIGION | ,034 | 1 | ,853 |
| | | RDCCN | 2,947 | 1 | ,086 |
| | | TOMDH | 1,161 | 1 | ,281 |
| | | Estadísticos globales | 4,328 | 5 | ,503 |