MODELO PARA ESTIMAR EL RIESGO DE SUFRIR INSUFICIENCIA RENAL CRÓNICA DE LA POBLACIÓN AFILIADA A EMSSANAR EN EL MUNICIPIO DE PASTO

MILTON ENRIQUE SANCHEZ DELGADO

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
SAN JUAN DE PASTO
2008

MODELO PARA ESTIMAR EL RIESGO DE SUFRIR INSUFICIENCIA RENAL CRÓNICA DE LA POBLACIÓN AFILIADA A EMSSANAR EN EL MUNICIPIO DE PASTO

MILTON ENRIQUE SANCHEZ DELGADO

Trabajo de GRADO

DIRECTOR
JORGE HUMBERTO MAYORGA

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
SAN JUAN DE PASTO
2008

Nota de Aceptación

_____
_____
_____
_____
_____
_____
_____

_____

Director

San Juan de Pasto, octubre de 2008

DEDICATORIA

Con el permiso de Dios, quiero dedicar este trabajo a mi padre, siempre serás mi guía y ejemplo, tu espíritu luchador, tu honestidad y cariño me acompañaran siempre. A mi amada esposa por ser siempre la cómplice de mis locuras, su apoyo incondicional y Amor me retan a ser cada día mejor. A mis hijos Luís Felipe y Alejandro ojala algún día pueda ser como ustedes. A mi querida Madre, tu amor infinito me impulsa a llegar cada vez más lejos.

A mis entrañables amigos Ángela, Mónica, Freddy, Carlos, Germán, Marco, Héctor, por compartir conmigo esta aventura y darme la oportunidad de aprender de ustedes, también por enseñarme que no hay límites cuando se comparten esfuerzos y que la mediocridad de algunos es la sabiduría de otros.

TABLA DE CONTENIDO

LISTA DE CUADROS.

LISTA DE FIGURAS

INTRODUCCIÓN

Uno de los problemas que enfrenta una persona en el ámbito mundial es la falta de recursos para cubrir las necesidades básicas de salud. Esto es especialmente cierto en países en vías de desarrollo como es el caso de Colombia y en general de los países latinoamericanos.

Con la formulación de la Ley 100 de 1993, en nuestro país se estableció un modelo de administración y prestación de servicios de salud revolucionario y social, que pretende imponer una serie de principios rectores orientados **al aseguramiento universal de la población**, independientemente de la capacidad de pago de los ciudadanos que reciben un amplio plan de beneficios, el Plan Obligatorio de Salud (POS), cubierto por entidades aseguradoras que reciben un monto de dinero estándar por usuario denominado Unidad de Pago por Capitación – UPC.

La UPC, debe ser administrada eficientemente ya que a diferencia de los seguros tradicionales en los cuales se mide desde el inicio un riesgo de ocurrencia del evento asegurado, del cual depende el valor de la póliza y la discrecionalidad de venderla o no; en el modelo de aseguramiento no se aplican dichos parámetros, ya que la ley exige que cada entidad aseguradora deberá afiliar a cualquier persona sea cual sea su estado de salud siempre y cuando cumpla con los requisitos definidos según el régimen de afiliación (subsidiado y contributivo).

Frente a una UPC cada vez mas insuficiente para cubrir el plan de beneficios definido por la Resolución 5261 de 1994 pero que vía tutela se ha hecho ilimitado, surge la necesidad de lograr el equilibrio económico por parte de las entidades aseguradoras. Por esta razón surge la necesidad de identificar el estado inicial de salud de las personas que se afilian de tal manera que se pueda focalizar atenciones específicas que conlleven por una parte a impactar positivamente en la salud pública y por otra garantizar el equilibrio financiero del sistema.

Este proyecto pretende a través de la utilización de la regresión logística, calcular la probabilidad que tiene un individuo de enfermar según sus factores de riesgo de Insuficiencia renal Crónica (IRC), con el fin de focalizar las acciones en salud necesarias que busquen minimizar la posibilidad de ocurrencia y de esta manera contribuir a mejorar la calidad de vida de las personas y garantizar la viabilidad financiera del sistema de salud.

1. MARCO TEÓRICO

1.1 RIESGO

El término **"riesgo"** puede adoptar diferentes connotaciones dentro del campo de la salud. Para la OMS, riesgo es la probabilidad de que un evento adverso para la salud ocurra, o bien, cualquier factor que aumente dicha probabilidad (ref informe OMS 2002). Aún siendo ampliamente usada, esta definición puede resultar problemática, puesto que homologa dos conceptos diferentes: una probabilidad de ocurrencia, y un factor que condiciona tal probabilidad.

En primera instancia parece claro que el riesgo es un concepto relacionado con los procesos de causalidad. Tales procesos implican la existencia de una causa o determinante, un efecto o consecuencia, y una conexión causal entre los dos, es decir, un nexo que conduce de la causa al efecto. Desde esta perspectiva, el riesgo puede verse como una forma de cuantificar la incertidumbre que surge del conocimiento imperfecto de los mecanismos que ligan una causa y su efecto.

La acepción de la OMS expuesta anteriormente identifica el riesgo con la causa, y también con una expresión de la conexión entre causa y efecto (la probabilidad). Otras definiciones comúnmente encontradas de la palabra riesgo incluyen su identificación con el efecto, la expresión de una incertidumbre (cuantificada o no), y la manifestación de una vulnerabilidad (ref reporte OMS 2002).

Para la Epidemiología Moderna, el riesgo es la probabilidad de que un evento relacionado con la salud ocurra en un individuo durante un periodo de tiempo determinado, dado que no existen otros factores que compitan con esta probabilidad (ref Kleinbaum). A menudo, al riesgo promedio de una población o conjunto de personas se le llama incidencia acumulada (ref Rothman - Epidemiology: an introduction). Esta será desde ahora en adelante la definición operativa de riesgo para este documento.

Entonces, el riesgo o incidencia acumulada, será la probabilidad promedio de que ocurra un evento de importancia para la salud en una población de individuos susceptibles, en un periodo de tiempo determinado, dado que esos individuos no pueden morir o desaparecer de la población fuente debido a otras causas (ref Rothman).

Matemáticamente se puede expresar como la siguiente proporción:

$$R_t = \frac{I}{N'}$$

Donde R_t es el riesgo para un periodo de tiempo D_t determinado, I es el número de casos incidentes (casos nuevos del evento estudiado) que aparecen durante el mismo periodo y N' es el número de sujetos con posibilidad de adquirir el evento al principio del periodo de observación (individuos susceptibles) (ref Kleinbaum). Como toda probabilidad, sus valores oscilan entre 0 (probabilidad nula) y 1 (probabilidad absoluta).

Esta fórmula provee una herramienta útil para cuantificar el riesgo en una población o cohorte cerrada, una vez los casos del evento de interés ya han ocurrido. No obstante, a menudo las condiciones de observación de una población no permiten la medición directa del riesgo mediante esta expresión sencilla. Esto ocurre, por ejemplo, cuando la población se comporta como una cohorte dinámica, o bien cuando la asunción de que no existen otros desenlaces y causas competentes no se cumple. En estos casos es posible utilizar tratamientos matemáticos que, introduciendo correcciones para este tipo de situaciones, logran proporcionar un estimativo no sesgado del riesgo. Una exposición detallada de estos métodos puede encontrarse en otras fuentes (ref Kleinbaum, Breslow & Day, Rothman).

También es posible que el interés del observador no esté centrado en cuantificar los casos que ya ocurrieron sino en tener un estimador de los casos que ocurrirán en un futuro relativamente cercano. En esta situación, es posible estimar el número de casos que ocurrirán mediante el uso de un modelo basado en el riesgo observado en el pasado, y el número de individuos susceptibles en el momento inicial del periodo para el cual se desea hacer la estimación, bajo la asunción de que la distribución poblacional de los factores que determinan dicho riesgo no se modificará durante el periodo. Parte del objetivo de este trabajo es proponer un modelo de este tipo.

Factor de Riesgo

La Epidemiología es una disciplina que, entre otros aspectos, se ocupa del estudio de las causas de las enfermedades (ref Kleinbaum). Aunque puede pensarse que la Epidemiología tiene un corte de pensamiento determinista, de hecho la imposibilidad de identificar leyes naturales absolutas y universales que expliquen la ocurrencia de la enfermedad implica la existencia de cierto grado de incertidumbre sobre los desenlaces que seguirán a una distribución específica de causas, incertidumbre que es fundamental e inherente a la disciplina.

Adicionalmente, la incapacidad de percibir los mecanismos causales (ref Hume en Rothman) y por tanto de diferenciar la causalidad de otras formas de determinación, como la autodeterminación cuantitativa, la interdependencia mutua o la determinación mecánica, (ref Weed, Douglas) y el conocimiento imperfecto de todos los posibles determinantes de la enfermedad, llevan a cambiar el término causa, por otro menos estricto: factor de riesgo.

Se utiliza la expresión factor de riesgo para referirse a una variable que se considera relacionada con la probabilidad de que un individuo desarrolle o no una enfermedad, cuya acción se da antes de la iniciación del proceso patológico en la historia natural de la enfermedad (ref Kleinbaum). Habitualmente se utiliza esta expresión con connotaciones peyorativas, refiriéndose únicamente a aquellos factores que se relacionan con un daño a la salud, distinguiéndolos de los factores protectores, es decir, aquellos que se relacionan con un efecto benéfico sobre la salud. Aunque en general el término riesgo, y por tanto factor de riesgo, puede usarse indiferentemente en ambas situaciones, en la medida de lo posible utilizaremos la expresión factores determinantes (o simplemente determinantes) para referirnos al conjunto de variables relacionadas con la enfermedad tanto en sentido protector como causal (ref Miettinen). Nótese que conforme a la definición de riesgo dada anteriormente es correcto usar el término factor de riesgo para referirse a aquel que disminuye la probabilidad de ocurrencia de la enfermedad.

Los factores determinantes se deben diferenciar de los factores promotores y de los factores de detección. Los primeros son aquellos que actúan una vez se ha completado el proceso etiológico, acelerando el inicio de los signos y síntomas, y por tanto la detección de la enfermedad. Los segundos aumentan la probabilidad de detección pero no modifican el momento de aparición de los signos y síntomas que marca el final del proceso patológico. Por tanto, los factores promotores tienen relevancia etiológica, mientras que los factores de detección no (ref Kleinbaum).

También suele usarse la expresión indicador o marcador de riesgo para referirse a un concepto semejante al de factores de detección. Son indicadores de riesgo aquellas variables que sin tener necesariamente una relación etiológica con la enfermedad en estudio, permiten predecir el riesgo debido a su asociación con otros factores que efectivamente actúan como determinantes de la enfermedad. Es oportuno aclarar que algunos autores utilizan el concepto indicador de riesgo en el sentido que se ha señalado aquí para factor de riesgo (ref Miettinen).

Finalmente, suelen llamarse factores pronósticos aquellos que modifican el curso clínico de la enfermedad, aumentando o disminuyendo la probabilidad de presentación de un desenlace particular (ref Kleinbaum). Para considerar que una variable es factor de riesgo generalmente se acepta que debe cumplir con tres características:

1. Covaría con el efecto estudiado, es decir la frecuencia relativa del efecto varía entre los distintos niveles del factor de riesgo.
2. El factor de riesgo precede temporalmente al efecto.
3. La covariación no es explicada enteramente por el error aleatorio o sistemático (ref McMahon).

De todo lo anterior puede derivarse como hecho destacable que el riesgo no es una probabilidad única que rige de manera natural para todos los individuos en el universo. Su valor presenta variaciones entre individuos específicos, dependiendo de la distribución de los determinantes del riesgo en la población de donde vienen dichos individuos (ref Miettinen). De ahí que el riesgo sea una probabilidad condicional, y los factores de riesgo su condición.

Por último, resaltamos que en términos de la investigación etiológica en Epidemiología, a menudo se homologan los conceptos de factor de riesgo y exposición. No obstante, la exposición debería entenderse preferiblemente como una caracterización detallada del contacto entre un individuo y un factor de riesgo (ref Armstrong). En cualquier caso, el desconocimiento de todos los posibles niveles de exposición de todos los posibles factores de riesgo para los individuos en una población, genera una limitación para la estimación del riesgo. En otras palabras, el riesgo promedio calculado para una población expuesta a una distribución de factores de riesgo determinada será diferente al de otra población o subconjunto de la misma que tenga una distribución de factores diferente. Este cálculo provee una estimación del riesgo individual (que no se puede medir directamente), aunque no está exenta de error (ref Rothman).

1.2 MODELO ESTADÌSTICO

1.2.1 MODELO DE REGRESIÓN LOGÍSTICA (LOGIT)[1]

En este caso, para el cálculo de la probabilidad de enfermar de IRC (Insuficiencia Renal Crónica) el modelo de regresión logística se definirá como sigue:

Se quiere identificar si una persona en el futuro desarrollará o no la enfermedad IRC, mediante una variable discreta Y_i, que se encuentra en función de un conjunto de variables independientes, los factores de riesgo identificados como predisponentes para la enfermedad.

[1] Montgomery (2000).

Las observaciones de cada persona encuestada se denotan por medio de una matriz como:

X'_{in} = (Factor de $Riesgo_{1n}$, Factor de $Riesgo_{2n}$,…, Factor de $Riesgo_{in})_{nxk}$

Donde, cada fila representará la combinación de factores de riesgo que explican el estado de salud de cada persona encuestada. Estas variables se especificarán más adelante pues se debe hacer énfasis en sus definiciones.

Los coeficientes asociados a las variables independientes se denotaran como un vector fila:

$$\beta'_i = (\beta_1, \beta_2, ..., \beta_n)_{Kx1}$$

y la variable independiente Y_i será:

$$Y_i = \begin{cases} 1 & \text{con probabilidad } P(Y = 1/X_i) = \pi(X_i) \\ 0 & \text{con probabilidad } P(Y = 0/X_i) = 1 - \pi(X_i) \end{cases}$$

Es decir,

$$Y_i = \begin{cases} 1, & \text{si la respuesta es éxito} \\ 0, & \text{si la respuesta es fracaso} \end{cases}$$

donde, Y_i toma el valor 1 si la persona encuestada desarrolla la enfermedad IRC, con probabilidad $\boldsymbol{\pi}(x_i)$ y 0 si la persona afiliada no desarrolla la enfermedad IRC con una probabilidad de $1 - \boldsymbol{\pi}(x_i)$. Adicionalmente se debe cumplir que $0 \leq \boldsymbol{\pi}_i \leq 1$.

Supóngase que el modelo tiene la forma:

$$y_i = x'_i\beta + \varepsilon_i$$

donde $E(\xi_i)=0$ y el valor esperado de la variable respuesta es:

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i)$$
$$= \pi_i$$

Esto implica que

$$E(y_i) = x_i^{'}\beta = \pi_i$$

Esto significa que la respuesta esperada, determinada con la función de respuesta $E(y_i) = x'_i\beta$ no es más que la probabilidad de que la variable respuesta tenga valor de 1.

Adicionalmente, se debe observar que si la variable respuesta es binaria, entonces los términos de error ε_i sólo pueden tomar dos valores:

$$\varepsilon_i = 1 - x_i^{'}\beta \quad \text{cuando } y_i = 1$$
$$\varepsilon_i = -x_i^{'}\beta \quad \text{cuando } y_i = 0$$

Es decir, los errores no son normales y su varianza no es constante:

$$\begin{aligned}\sigma_y^2 &= E(y_i - E(y_i))^2 \\ &= (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i) \\ &= \pi_i (1 - \pi_i)\end{aligned}$$

Equivalente a:

$$\sigma_y^2 = E(y_i)\left[1 - E(y_i)\right]$$

porque $E(y_i) = x'_i\beta = \pi_i$, lo que indica que la varianza de las observaciones es una función de la media.

Como forma funcional para $E(y_i)$, se usa la función de respuesta logística que garantiza valores en el intervalo (0, 1) y tiene la forma:

$$E(y) = \frac{e^{(x^{'}\beta)}}{1 + e^{(x^{'}\beta)}}$$

la función de respuesta logística se puede linealizar con facilidad.
Un enfoque consiste en definir la estructural del modelo en términos de una función de la media de la función de respuesta. Sea

$$\eta = x_i^{'}\beta$$

y el predictor lineal, estando definida η por la transformación

$$\eta = \ln \frac{\pi}{1-\pi}$$

a esta transformación se le conoce como: la razón *odds*, que se define como razón entre la probabilidad de que una persona desarrolle la enfermedad y la probabilidad de que no.

Para estimar los parámetros del predictor lineal $x_i\beta$, se usa el método de máxima verosimilitud. Como cada observación de la muestra sigue la distribución Bernoulli, la distribución de probabilidades de cada observación será:

$$f(y_i) = \pi_i^{\,y_i}(1-\pi_i)^{1-y_i}, \qquad i = 1,2,...,n$$

y cada observación y_i toma el valor 0 o 1.

Como las observaciones son independientes, la función de verosimilitud será:

$$\begin{aligned} L(y_1, y_2, ..., y_n, \beta) &= \prod_{i=1}^{n} f(y_i) \\ &= \prod_{i=1}^{n} \pi_i^{\,y_i}(1-\pi_i)^{1-y_i} \end{aligned}$$

o por facilidad, el logaritmo de la función de verosimilitud:

$$\begin{aligned} \ln L(y_1, y_2, ..., y_n, \beta) &= \ln \prod_{i=1}^{n} f(y_i) \\ &= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi_i}{1-\pi_i} \right) \right] + \sum_{i=1}^{n} \ln(1-\pi_i) \end{aligned}$$

Ahora bien, sea:

$$1-\pi_i = \left[1+e^{(x_i'\beta)}\right]^{-1} \qquad \text{y} \qquad \eta = \ln\frac{\pi}{1-\pi}$$

Entonces, el logaritmo de la función de verosimilitud se puede expresar como:

$$L(\beta) = \ln L(y,\beta) = \sum_{i=1}^{n} y_i x_i'\beta - \sum_{i=1}^{n} \ln(1+e^{(x_i'\beta)})$$

Para estimar el vector β_{Kx1} que maximice $L(\beta)$:

$$\frac{\partial L(\beta)}{\partial \beta_i} = 0$$

donde i= 1,…,i

Esto es equivalente a:

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

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

De manera general, se pueden utilizar métodos numéricos para calcular los estimadores β por máxima verisimilitud, pero resulta más fácil utilizar el método de mínimos cuadrados ponderados generalizados iterados.

De donde se obtendrá el valor estimado del predictor lineal $\eta = x_i\beta$ y el valor esperado del modelo de regresión logística

$$\hat{y} = \hat{\pi} = \frac{e^{(\hat{\eta})}}{1 + e^{(\hat{\eta})}}$$
$$= \frac{e^{(x'\beta)}}{1 + e^{(x'\beta)}}$$
$$= \frac{1}{1 + e^{(-x'\beta)}}$$

1.2.2 PRUEBAS DE BONDAD DE AJUSTE Y DE LOS PARAMETROS DEL MODELO[2]

Las pruebas de hipótesis para el modelo de regresión logística se basan en pruebas de cociente de máxima verosimilitud (procedimiento para muestras grandes). Este método conduce a un estadístico conocido como desviación, que nos ayuda a determinar si el modelo propuesto describe de manera acertada o no la relación entre los datos.

***Pruebas de ajuste del modelo**:*

La desviación del modelo compara el logaritmo de verosimilitud del modelo ajustado con la verosimilitud del modelo saturado, que en este caso es un modelo donde las probabilidades $\pi(x_i)$ son totalmente irrestrictas por lo que al igualar $\pi(x_i)$ = y_i se maximizará la verosimilitud.

Función de verosimilitud del modelo ajustado:

$$\ln L(\hat{\beta}) = \sum_{i=1}^{n} y_i x_i' \hat{\beta}_i - \sum_{i=1}^{n} \ln[1 + e^{x_i' \hat{\beta}}]$$

La desviación compara el logaritmo de verosimilitud del modelo saturado con el logaritmo de verosimilitud del modelo ajustado. De manera especifica la desviación del modelo se define como:

[2] Montgomery (2004)

$$\lambda(\beta) = 2\ln L(\text{modelo saturado}) - 2\ln L(\hat{\beta})$$

$$= 2[\ell(\text{modelo saturado}) - \ell(\hat{\beta})]$$

donde ℓ representa el logaritmo de la función de verosimilitud.

Si el modelo de regresión logística es la función correcta para describir el comportamiento de los datos, y n es una muestra grande, entonces la desviación del modelo tendrá una chi-cuadrado con n-p grados de libertad. Si la desviación del modelo representa valores grandes el modelo no es correcto, mientras que si representa valores pequeños significa que el modelo ajustado que tiene menos parámetros se ajusta casi tan bien como el modelo saturado.

Si $\lambda(\beta) \leq \chi^2_{\alpha,n=p}$ el modelo ajustado es adecuado

Si $\lambda(\beta) > \chi^2_{\alpha,n=p}$ el modelo ajustado no es adecuado

Nota: La desviación se considera como en el modelo de regresión lineal en el error de la suma de cuadrados residuales dividido entre la varianza del error σ^2.

Pruebas sobre subconjuntos de parámetros:

La desviación también se utiliza para hacer pruebas de hipótesis sobre subconjuntos de parámetros del modelo.

Supóngase que se desea probar:

$$H_0 : \beta_2 = 0$$

Por consiguiente el modelo completo será: $\eta = X\beta$ y el modelo reducido será $\eta = X_1\beta_1$, y su desviación será $\lambda(\beta_1)$y la diferencia en la desviación se definirá como:

$$\lambda(\beta_2 / \beta_1) = \lambda(\beta_1) - \lambda(\beta)$$

donde $\lambda(\beta_2/\beta_1)$ tiene una distribución chi-cuadrado con r (r=n-(n-r)(n-p)) grados de libertad.

Si $\lambda(\beta_2 / \beta_1) \leq \chi^2_{\alpha,r}$ se rechaza la hipótesis nula

Si $\lambda(\beta_2 / \beta_1) > \chi^2_{\alpha,r}$ no se rechaza la hipótesis nula

La diferencia de la desviación a veces se conoce como desviación parcial, que es una prueba de cociente de verosimilitud:

$$\frac{L(\hat{\beta}_1)}{L(\hat{\beta})}$$

El estadístico para la prueba de cociente de verosimilitud es igual a -2 multiplicado por el logaritmo del cociente de verosimilitud, es decir:

$$\chi^2 = -2 \ln \frac{L(\hat{\beta}_1)}{L(\hat{\beta})}$$

Coeficiente de determinación:

Así como el coeficiente de verosimilitud es un estadístico que permite identificar el grado de ajuste del modelo, Cox y Snell en 1989 introdujeron el concepto de coeficiente de determinación para modelos lineales generalizados. Pretende de igual forma cuantificar mediante un valor comprendido entre 0 y 1 la bondad del ajuste; siendo mayor la variación explicada por el modelo mientras más alto es el valor de esta medida:

$$R^2 = 1 - \left(\frac{L(\hat{\beta}_1)}{L(\hat{\beta})} \right)^{2/n}$$

En 1991 Nagelkerke propone:

$$\bar{R}^2 = \frac{R^2}{\max(R^2)} \quad \text{donde} \max(R^2) = 1 - L(\hat{\beta})^{2/n}$$

***Pruebas individuales de parámetros**:*

Pruebas para coeficientes individuales del modelo.

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Se puede utilizar el método de diferencia de la desviación, pero adicionalmente se puede utilizar el estadístico de Wald, que se basa en la teoría de los estimadores de máxima verosimilitud como sigue:

Sea G la matriz de p x p de las segundas derivadas parciales de la función logaritmo de verosimilitud, esto es,

$$G_{ij} = \frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j}, \qquad i, j = 0,1,\ldots,k$$

G se llama matriz hessiana. Si los elementos de la matriz hessiana se evalúan en los estimadores de máxima verosimilitud $\beta = \hat{\beta}$, la matriz de covarianza para muestra grande, de los coeficientes de regresión es,

$$Var(\hat{\beta}) = \hat{\sum} = -G(\hat{\beta})^{-1}$$

las raíces cuadradas de los elementos diagonales de esta matriz son los errores estándar de muestras grandes de los coeficientes de regresión, por lo que el estadístico de prueba de la hipótesis nula es:

$$Z_0 = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

la distribución de referencia para este estadístico es la distribución normal estándar.

1.2.3 SELECCIÓN DE VARIABLES INDEPENDIENTES

La manera mas acertada de seleccionar variables es identificar las relaciones teóricas de los problemas a solucionar, sin embargo, estas relaciones no garantizan en todos los casos un ajuste adecuado del modelo estadístico. Para la selección de variables explicativas que mejor describan el comportamiento de la variable dependiente se han desarrollado diferentes métodos:

Método Backward:
Este método incluye todas las variables explicativas disponibles en el modelo, y paso a paso va eliminando las variables que no sean significativas en la estimación del modelo.

Método Forward:
Este método inicia sin incluir ninguna variable explicativa y paso a paso va incluyendo en orden de importancia las variables que mayor correlación tengan con la variable dependiente.

Metodo Stepwise:
Este método es la combinación de los dos anteriores, comienza como el método forward, sin embargo, la inclusión una variable puede resultar en exclusión de otra variable que ya estaba seleccionada pero que con la nueva inclusión resulta redundante.

Estos métodos utilizan el estadístico de Wald y la puntuación eficiente de Rao, para validar la significancia estadística de las variables explicativas.

Estadístico de Wald:

$$H_0 : \beta_i = 0$$

Si no existe evidencia estadística para rechazar la hipótesis nula, significa que la información que se perderá al eliminar la variable $_i X$ en el siguiente paso no es significativa. *Criterio de decisión:* si p_valor < α se rechazara H_0 con un nivel de significancia de α.

***Puntuación eficiente de Rao*:**

$$H_0 : \beta_i = 0$$

Si no existe evidencia estadística para rechazar la hipótesis nula, significa que si la variable X_i fuera seleccionada en el siguiente paso la información que aportaría no

sería significativa. *Criterio de decisión:* si p_valor < α se rechazara H_0 con un nivel de significancia de α.

1.2.4 SUPUESTOS DEL MODELO

Colinealidad:
Caso en el que dos variables independientes se encuentran altamente correlacionadas, causando que la estimación de los parámetros de dichas variables no sean significativos. Para la validación de este supuesto se utilizan el coeficiente de correlación de Pearson o el de Spearman según la naturaleza de los datos.

Linealidad:
Este supuesto hace referencia a la linealidad o monotonía entre las variables independientes continuas y la probabilidad de ocurrencia del incumplimiento, en este caso.

Para realizar dicha validación es necesario graficar la relación existente entre

$$\left(\log\left(\frac{\pi_i}{1-\pi_i} \right), x_i \right) \text{ donde i = 1,..., p}$$

p es el número de posibles valores que toma la variable x, si existe una relación directa o indirecta ente la probabilidad de ocurrencia del evento y la variable, decimos que se cumple el supuesto de linealidad.

1.2.5 PRUEBAS DE DISCRIMINACION

Existe una serie de pruebas que evalúan desde diferentes ópticas la eficiencia de clasificación de los modelos. Entre las principales se encuentran:

Prueba Kolmogorov-Smirnov:

Esta prueba mide el poder discriminante o de separabilidad de la variable de estudio, calculando la máxima diferencia entre las dos distribuciones de las

categorías de la variable criterio y comparándola con un valor crítico determinado a partir del grado de confianza escogido.

Las hipótesis planteadas en esta prueba son:
H_0 : Las distribuciones de frecuencia acumulada son similares.
H_1 : Las distribuciones de frecuencia acumulada son diferentes.

El estadístico de contraste se calcula como:

$$\underset{\text{Sobre todas las clases i}}{\text{Max}} = \left| (\hat{f}_{i\,B} - \hat{f}_{i\,M}) \right|$$

Donde:

$\hat{f}_{i\,B}$ = Frecuencia acumulada de la categoría [Bueno] para la clase i de la variable en estudio.

$\hat{f}_{i\,M}$ = Frecuencia acumulada de la categoría [Malo] para la clase i de la variable en estudio.

La regla de decisión consiste en rechazar H_0 a un nivel de significación α, si el valor calculado del estadístico de contraste es mayor al valor crítico.

Curva de Características Operativas (por su sigla en ingles Curva ROC):

La curva ROC es un método que se utiliza para evaluar la eficiencia del modelo al clasificar correctamente las observaciones de la categoría referencia de la variable criterio, en este caso [Enfermo] (sensibilidad), mientras clasifica incorrectamente las observaciones de la otra categoría, en este caso [Sano] (1- especificidad).

Cuadro No.1 Tabla Clasificación de Afiliados Pronostico vs. Observado

Pronostico	**Casos Observados**		**Total**
	Enfermos	**Sanos**	
Enfermos	X_b	$X_{b/m}$	$X_b + X_{b/m}$
Sanos	$X_{m/b}$	X_m	$X_m + X_{m/b}$
Total	$X_b + X_{m/b}$	$X_{b/m} + X_m$	

Sensibilidad del modelo:
Probabilidad de que el modelo clasifique a las personas enfermas como enfermas.

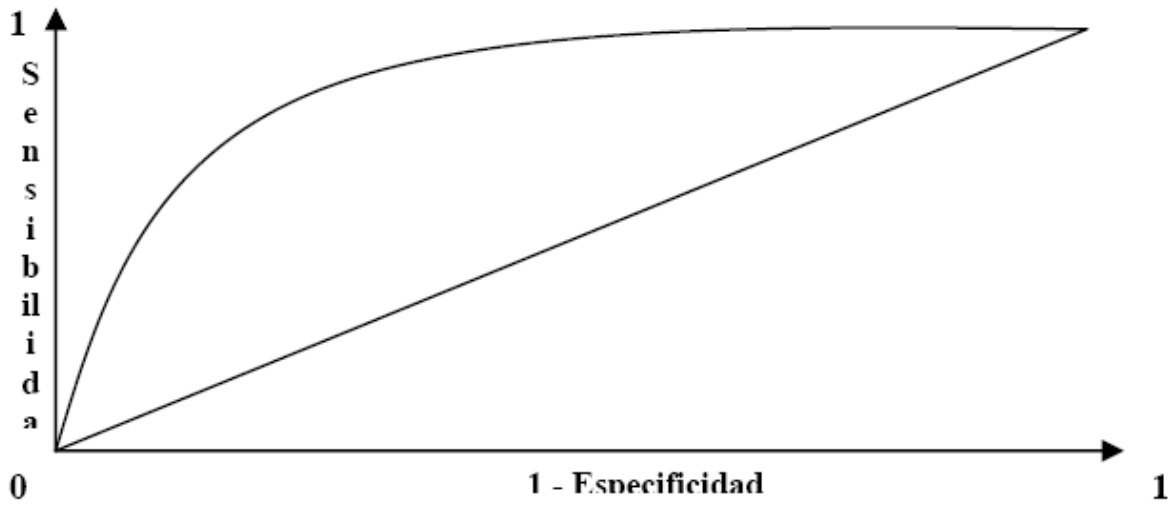
$$S = \frac{X_b}{X_b + X_{m/b}}$$

Especificidad del modelo:
Probabilidad de que el modelo clasifique a las personas sanas como sanas.

$$E = \frac{X_m}{X_m + X_{b/m}}$$

Entonces, en las curvas se representa la sensibilidad en función de los buenos mal clasificados para distintos puntos de corte.

Figura No. 1 Distribución de Afiliados Enfermos y Sanos



El parámetro para evaluar la eficiencia del modelo es el área bajo la curva: Si toma valores mayores a 0,5 el ajuste será mejor que si toma valores inferiores a 0,5.

En general, existen más métodos que ayudan a identificar la eficiencia del modelo al momento de clasificar a los individuos, el coeficiente de Gini, la curva CAP, el Accuracy Ratio conocido como AR, entre otros. Sin embargo, todos estos métodos están basados en la identificación de los casos bien y mal clasificados.

1.3 INSUFICIENCIA RENAL CRONICA IRC

Los riñones presentan una gran cantidad de funciones. Entre ellas destacan las siguientes:

- Filtrar la sangre consiguiendo la eliminación de los productos tóxicos y de desecho. Así, elimina las sustancias nocivas producidas por el propio cuerpo (urea, creatinina, etc.) como algunos fármacos una vez metabolizados (es decir, que han dejado realizar su función o se han transformado en sustancias nocivas).
- Mantenimiento de la presión arterial mediante la eliminación de agua y la secreción de hormonas.
- Retener nutrientes (proteínas, glucosa y vitaminas) y hormonas.

Si la función renal se va haciendo más lenta y el riñón se lesiona gradualmente, se desencadena la incapacidad de éste para realizar su trabajo. Este fenómeno se llama ***insuficiencia renal crónica*** porque el problema se desencadena y desarrolla lentamente, pudiendo llevar al riñón a que deje de funcionar. Cuando ambos riñones fallan, el cuerpo comienza a retener líquido y sustancias nocivas. Entonces la presión sanguínea sube, aparecen edemas, el organismo no produce suficientes glóbulos rojos (comienza a producirse anemia), etc. Cuando esto sucede, es necesario recurrir a tratamientos que sustituyan el trabajo de los riñones.

Causas

La insuficiencia renal crónica puede ser producida por una gran variedad de causas:

- Infecciones
- Medicamentos
- Lesiones
- Enfermedades renales: inflamación de la unidad funcional del riñón (glomerulonefritis) y nefropatías en general
- Diabetes
- Hipertensión
- Aterosclerosis

Sea cual sea la causa, el problema es que el riñón deja de realizar su función de filtrado y puede llegar a instaurarse de forma eventual o permanente un proceso caracterizado por el incremento y acumulación de sustancias tóxicas en la sangre, especialmente urea, denominado uremia.

Tratamiento

- **Hemodiálisis**: El objetivo de este procedimiento es sustituir la acción limpiadora y filtradora del riñón. Extrae del cuerpo la sal, exceso de líquido y desechos tóxicos. Ayudando a mantener en la persona un control de la presión arterial y de la composición del organismo. La sangre pasa por un dializador, es decir un filtro de características especiales capaz de limpiar la sangre.
- **Diálisis peritoneal:** Se trata de otro procedimiento para reemplazar la función del riñón. En este tipo de diálisis se aprovecha el propio revestimiento del interior del abdomen (membrana peritoneal) para limpiar la sangre. En este proceso una solución purificadora, llamada dializante, se introduce en el abdomen mediante un dispositivo especial, consiguiendo que los productos de desecho y sustancias nocivas pasen desde los pequeños vasos presentes en la membrana peritoneal al dializado. Después de varias horas se drena el abdomen (se saca el líquido introducido en el abdomen) y a continuación se repite el proceso.
- **Trasplante de riñón:** Se trata de una cirugía mayor, a la que el paciente debe de acompañar con la toma de medicamentos por el resto de su vida para impedir un rechazo del órgano.

2. PLANTEAMIENTO DEL PROBLEMA

¿Es posible calcular el riesgo de sufrir Insuficiencia Renal Crónica según la declaración del estado inicial de salud en el momento de la afiliación en el municipio de Pasto?

3. JUSTIFICACIÓN

Más de 20 millones de personas en Estados Unidos (uno de cada nueve adultos) padecen de insuficiencia renal crónica, y la mayoría de ellos ni siquiera lo saben. Otras más de 20 millones de personas tienen riesgo elevado de sufrir esta enfermedad. En el departamento de Nariño se han identificado 273 afiliados a Emssanar que padecen la enfermedad y de estos 90 son residentes del municipio de Pasto, estos pacientes demandaron 3,438 millones de pesos en el año 2007 que representa un 3.65% de la UPC. Adicionalmente los estudio han demostrado que el costo social de padecer la enfermedad reduce en promedio 10 años en la expectativa de vida de los que la padecen[3]. Tener riesgo elevado significa una probabilidad por encima del promedio de desarrollarla. La detección temprana y un tratamiento focalizado ayudan a evitarla. Por estas razones es fundamental avanzar en procesos técnicos de identificación temprana ya sean clínicos o como en este a través de la aplicación de encuestas que permitan medir el nivel de riesgo de las personas.

[3] OMS-2004

4. OBJETIVO GENERAL

Estimar la probabilidad riesgo de sufrir Insuficiencia Renal Crónica, de los nuevos afiliados de Emssanar EPS-S con base a su declaración de inicial de estado de salud.

5. OBJETIVOS ESPECÍFICOS

- Determinar los factores de riesgo de la Insuficiencia Renal Crónica y validar el instrumento utilizado por Emssanar para la declaración de los estados iniciales de salud.
- Identificar de las personas que han sido encuestadas y que han demandado atención en salud cuales han sido diagnosticadas con IRC.
- Establecer un modelo estadístico multivariado que permita calcular el nivel de riesgo de los individuos en el momento de la afiliación.
- Desarrollar un programa que permita aplicar el modelo de forma automatizada en el momento de la afiliación.

6. METODOLOGÍA

6.1 Diseño y Técnicas de Recolección de la información

Desde Junio del 2007, Emssanar EPS-S implemento el proceso de la aplicación de la encuesta denominada Ficha de Identificación de Estados de Salud **FIES,** este instrumento fue elaborado por un grupo de profesionales de la salud y su objetivo principal es, a través de preguntas muy sencillas tratar de identificar el estado inicial de salud de la población que se afilia a la EPS, este instrumento se ha implementado en todos los municipios y departamentos donde la organización hace presencia sin embargo el alcance de este proyecto cubrirá solo el Municipio de San Juan de Pasto, Departamento de Nariño.

Esta ficha es aplicada a todo nuevo afiliado a la organización, para lo cual el personal responsable de la afiliación antes de carnetizarlo indaga a através del instrumento información referente a sus hábitos de vida, posibles enfermedades y antecedentes familiares. Las preguntas realizadas a demás de los datos personales como nombre, identificación, Sexo, Edad son:

- ¿En los últimos 6 meses ha sido examinado por algún médico o enfermera?
- ¿Ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años?
- ¿Actualmente fuma cigarrillo?
- ¿Alguna vez un médico o enfermera le ha dicho que ha tenido o que tiene presión alta?
- ¿Actualmente está en tratamiento con medicamentos para controlar su presión arterial?
- ¿Realiza ejercicio físico mínimo 15 minutos por lo menos 3 veces por semana?
- ¿Alguna vez un médico o enfermera le ha dicho que pesa más o menos de lo que debería?
- ¿Alguna vez un médico le ha dicho que tiene colesterol alto?
- ¿Actualmente está haciendo alguna dieta o tratamiento para controlar su colesterol?
- ¿Alguna vez un médico le ha dicho que padece de diabetes o azúcar alta en la sangre?
- ¿Actualmente hace alguna dieta o tratamiento para controlar la diabetes?

- ¿Consume licor por lo menos una vez a la semana hasta “prenderse” o emborracharse?
- ¿Está usted en embarazo actualmente?
- ¿Ha presentado sangrados, amenaza de aborto, parto prematuro o hipertensión arterial?
- ¿Asiste a control prenatal?
- ¿En embarazos anteriores presentó diabetes, hipertensión arterial o infecciones?
- ¿Se ha tomado la citología vaginal durante los últimos doce meses?
- ¿Alguna vez le han diagnosticado insuficiencia renal crónica o aguda?
- ¿Alguna vez le han diagnosticado cáncer?
- ¿Alguna vez le han diagnosticado enfermedades del corazón?
- ¿Consume o le han formulado stocrin, combivir, zerit, kaletra?
- Algún médico le ha dicho que tiene que realizarse alguna de las siguientes cirugías:
- Cirugía del sistema nervioso (médula, cerebro, columna vertebral)
- Cirugías por enfermedades del corazón
- Prótesis o reemplazo articular
- ¿Alguno de sus padres sufre o sufrió de diabetes, hipertensión o cáncer?

Es muy importante aclarar que las respuestas a estas preguntas son digitadas directamente en el sistema “AFILIACIONES” de propiedad de la entidad, es decir no se necesita adelantar procesos de consolidación ya que dicha herramienta se encuentra en línea y alimenta una base de datos centralizada a nivel nacional.

Desde el mes de Enero a Marzo de 2008, en el Departamento de Nariño se han diligenciado 13498 fichas, de las cuales 8829 corresponden a afiliados del municipio de Pasto.

Para el desarrollo del proyecto se procedió a identificar a través de cruces de las bases de datos de atenciones empresariales, los afiliados que en el periodo de tiempo señalado habían demandado servicios de consulta externa en cualquiera de las instituciones contratadas.

De estos cruces se obtienen los siguientes resultados: 531 afiliados demandaron por lo menos una vez servicios de Consulta Médica General, de los cuales 12 fueron diagnosticados con Insuficiencia Renal Crónica. Es necesario aclarar que entre los afiliados que demandaron servicios y que no fueron diagnosticados con la enfermedad pueden existir enfermos que por no haber presentado la sintomatología clínica no fueron detectados por el médico en el momento de la consulta.

6.2 Validación de la información

Una vez identificados los casos de los afiliados que resultaron enfermos con IRC, se procedió a contrastar la información suministrada en la encuesta con la existente en los registros individuales de prestación de servicios de salud (RIPS), obteniendo los siguientes datos:

Cuadro No. 2 Comparación Encuesta – RIPS

Pregunta	Encuesta		RIPS	
	Si	No	Si	No
¿En los últimos 6 meses ha sido examinado por algún médico o enfermera?	6	6	10	2
¿Ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años?	4	8	10	2
¿Actualmente fuma cigarrillo?	2	10	8	4
¿Alguna vez un médico o enfermera le ha dicho que ha tenido o que tiene presión alta?	3	9	7	5
¿Actualmente está en tratamiento con medicamentos para controlar su presión arterial?	2	10	4	8
¿Realiza ejercicio físico mínimo 15 minutos por lo menos 3 veces por semana?	0	12	0	12
¿Alguna vez un médico o enfermera le ha dicho que pesa más o menos de lo que debería?	2	10	4	8
¿Alguna vez un médico le ha dicho que tiene colesterol alto?	4	8	2	10
¿Actualmente está haciendo alguna dieta o tratamiento para controlar su colesterol?	0	12	1	11
¿Alguna vez un médico le ha dicho que padece de diabetes o azúcar alta en la sangre?	2	10	7	5
¿Actualmente hace alguna dieta o tratamiento para controlar la diabetes?	0	12	2	10
¿Consume licor por lo menos una vez a la semana hasta “prenderse” o emborracharse?	1	11	3	9
¿Está usted en embarazo actualmente?	0	12	0	12
¿Ha presentado sangrados, amenaza de aborto, parto prematuro o hipertensión arterial?	0	12	0	12
¿Asiste a control prenatal?	0	12	0	12

¿En embarazos anteriores presentó diabetes, hipertensión arterial o infecciones?	0	12	0	12
¿Se ha tomado la citología vaginal durante los últimos doce meses?	NA	NA	NA	NA
¿Alguna vez le han diagnosticado insuficiencia renal crónica o aguda?	4	8	6	6
¿Alguna vez le han diagnosticado cáncer?	0	12	0	12
¿Alguna vez le han diagnosticado enfermedades del corazón?	0	12	1	11
¿Consume o le han formulado stocrin, combivir, zerit, kaletra?	0	12	0	12
Algún médico le ha dicho que tiene que realizarse alguna de las siguientes cirugías:				
Cirugía del sistema nervioso (médula, cerebro, columna vertebral)	0	12	0	12
Cirugías por enfermedades del corazón	0	12	0	12
Prótesis o reemplazo articular	0	12	0	12
¿Alguno de sus padres sufre o sufrió de diabetes, hipertensión o cáncer?	6	6	8	4

De los datos obtenidos podemos concluir que existen grandes diferencias entre las repuestas dadas por las personas en el momento de la encuesta y las evidencias encontradas en la RIPS de las mismas, esta situación se puede estar presentando por varias razones entre ellas: el temor de las personas encuestadas de no ser afiliadas al declarar su verdadero estado de salud, mala interpretación de la preguntas por el bajo nivel socioeconómico del encuestado o mal registro por parte del encuestador.

Con estos resultados, queda en evidencia que el mecanismo utilizado por la organización no tiene ninguna valides para los fines propuestos ya que de ninguna manera se podría establecer un perfil de riesgo con información dudosa, situación que debe ser revisada y replanteada buscando otros mecanismos que garanticen que aunque posiblemente resulten mas costosos brinden una alta confiabilidad.

Teniendo en cuenta esta situación, y reflexionando sobre los objetivos del presente estudio se propone para los fines del mismo realizar el proceso de validación a los 531 registros de la población objeto (enfermos y sanos), contrastándolos con la información reportada en los RIPS por las entidades prestadoras de tal manera que se garantice que los datos del estudio provienen de una fuente de mayor confiabilidad y que el modelo cumpla con los requerimientos necesarios para ser considerado como válido.

Este proceso de validación consistió en realizar 23 cruces de cada una de las variables con las bases de datos de Consultas, Urgencias, Hospitalizaciones y

Procedimientos, una vez se realiza esta validación se corrigieron los datos en la base datos del estudio.

6.3 Construcción del Modelo de Regresión Logística

Partiendo de los datos obtenidos, nos centramos en la estimación, las pruebas y la interpretación de los coeficientes en un modelo de regresión logística para esto se realizó la siguiente codificación de las variables definidas en la encuesta:

Cuadro No. 3 Codificación de las variables en estudio

Variable	**Descripción**	**Código / valores**	**Nombre**
1	Sexo del encuestado	0= Hombre 1= Mujer	SEXO
2	Grupo Étnico	0= Otro 1= Afrocolombiano/indigena	GRUPOETNIC
3	Zona de Residencia del Encuestado	0= Urbana 1= Rural	ZONA
4	El Encuestado presenta alguna Discapacidad (Mental, Física o Sensorial)	0= No 1= Sí	DI2CAPAC
5	Grupo de edad del encuestado	1 = (< 30 años) 0= (>=30 años)	Edad2
6	Saber Leer y Escribir	0= No 1= Sí	LeeyEscribe
7	¿En los últimos 6 meses ha sido examinado por algún médico o enfermera?	0= No 1= Sí	P1ExamenMedico
8	¿Ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años?	0= No 1= Sí	P2HaFumado
9	¿Actualmente fuma cigarrillo?	0= No 1= Sí	P3FumaHoy
10	¿Alguna vez un médico o enfermera le ha dicho que ha tenido o que tiene presión alta?	0= No 1= Sí	P4PresionAlta
11	¿Actualmente está en tratamiento con medicamentos para controlar su presión arterial?	0= No 1= Sí	P5ControlPresion
12	¿Realiza ejercicio físico mínimo 15 minutos por lo menos 3 veces por semana?	0= No 1= Sì	P6EjercicioFisico
13	¿Alguna vez un médico o enfermera le ha dicho que pesa más o menos de lo que debería?	0= No 1= Sí	P7ControlPeso

14	¿Alguna vez un médico le ha dicho que tiene colesterol alto?	0= No 1= Sí	P8Colesterol
15	¿Actualmente está haciendo alguna dieta o tratamiento para controlar su colesterol?	0= No 1= Sí	P9ControlColesterol
16	¿Alguna vez un médico le ha dicho que padece de diabetes o azúcar alta en la sangre?	0= No 1= Sí	P10Diabetes
17	¿Actualmente hace alguna dieta o tratamiento para controlar la diabetes?	0= No 1= Sí	P11ControlDiabetes
18	¿Consume licor por lo menos una vez a la semana hasta “prenderse” o emborracharse?	0= No 1= Sí	P12Licor
19	¿Está usted en embarazo actualmente?	0= No 1= Sí	P13Embarazada
20	¿Ha presentado sangrados, amenaza de aborto, parto prematuro o hipertensión arterial?	0= No 1= Sí	P13ANovedadE
21	¿Asiste a control prenatal?	0= No 1= Sí	P13BControlPrenatal
22	¿En embarazos anteriores presentó diabetes, hipertensión arterial o infecciones?	0= No 1= Sí	P14EmbarazoAntes
23	¿Se ha tomado la citología vaginal durante los últimos doce meses?	0= No 1= Sí	P15Citologia
24	¿Alguna vez le han diagnosticado insuficiencia renal crónica o aguda?	0= No 1= Sí	P16IRC
25	¿Alguna vez le han diagnosticado cáncer?	0= No 1= Sí	P17Cancer
26	¿Alguna vez le han diagnosticado enfermedades del corazón?	0= No 1= Sí	P18Corazon
27	¿Consume o le han formulado stocrin, combivir, zerit, kaletra?	0= No 1= Sí	P19ControlSida
28	Algún médico le ha dicho que tiene que realizarse alguna de las siguientes cirugías:		
29	Cirugía del sistema nervioso (médula, cerebro, columna vertebral)	0= No 1= Sí	P20ACSNervioso
30	Cirugías por enfermedades del corazón	0= No 1= Sí	P20BCECorazon
31	Prótesis o reemplazo articular	0= No 1= Sí	P20CReemplazo
32	¿Alguno de sus padres sufre o sufrió de diabetes, hipertensión o cáncer?	0= No 1= Sí	P21Antecedentes
33	Diagnosticados confirmados IRC	0= No 1= Sí	IRC

Como se puede observar hay muchas variables independientes que podrían ser incluidos en el modelo. Por lo tanto, se define una estrategia que permita seleccionar aquellas variables que dan lugar a un "mejor" modelo en el contexto del problema.

Se dice que el éxito de modelar un complejo conjunto de datos es parte ciencia, parte de los métodos estadísticos, y ***una parte la experiencia y el sentido común***. El objetivo es proponer un modelo que describa de mejor manera posible los casos en estudio dentro de las limitaciones de los datos disponibles.

6.3.1 Selección de las Variables

El proceso de reducir al mínimo el número de variables busca principalmente que el modelo resultante sea muy estable y permita generalizar con base a los resultados obtenidos.

En epidemiología se sugiere incluir las variables intuitivamente con base a la evidencia clínica relevante en el modelo, independientemente que sean significativas estadísticamente. La justificación de este enfoque es proporcionar lo más completo control de los efectos confundidos como sea posible dentro del conjunto de datos. Esto se basa en el hecho de que es posible que variables individuales que aparentemente no son significativas individualmente, son relevantes cuando se toman en conjunto.

Como primera etapa en búsqueda de ajustar un modelo de regresión logística se realizaron las pruebas a los supuestos a cada una de las variables independientes, basados en que la regresión logística no se basa en supuestos distribucionales como lo hace el análisis discriminante, si es importante considerar que la solución puede ser más estable si los predictores tienen una distribución normal multivariante. Adicionalmente, al igual que con otras formas de regresión, la multicolinealidad entre los predictores puede llevar a estimaciones sesgadas y a errores típicos inflados.

Con los criterios anteriormente descritos y una vez realizado los análisis pertinentes se excluyeron del estudio las siguientes variables:

Cuadro No. 4 Variables excluidas por no cumplir los supuestos del modelo

Variable	Descripción	Código / valores	Nombre	Causa de Exclusión
13	¿Alguna vez un médico o enfermera le ha dicho que pesa más o menos de lo que debería?	0= No 1= Sí	P7ControlPeso	Las 531 respuestas son iguales a 0
15	¿Actualmente está haciendo alguna dieta o tratamiento para controlar su colesterol?	0= No 1= Sí	P9ControlColesterol	Presenta combinación lineal con la variable P8Colesterol

27	¿Consume o le han formulado stocrin, combivir, zerit, kaletra?	0= No 1= Sí	P19ControlSida	Las 531 respuestas son iguales a 0
28	Algún médico le ha dicho que tiene que realizarse alguna de las siguientes cirugías:			
29	Cirugía del sistema nervioso (médula, cerebro, columna vertebral)	0= No 1= Sí	P20ACSNervioso	Las 531 respuestas son iguales a 0
30	Cirugías por enfermedades del corazón	0= No 1= Sí	P20BCECorazon	Las 531 respuestas son iguales a 0
31	Prótesis o reemplazo articular	0= No 1= Sí	P20CReemplazo	Las 531 respuestas son iguales a 0

El siguiente paso es buscar un modelo de regresión logística que ajuste la variable dependiente con las variables independientes. Para realizar esta tarea se utilizaron las versiones académicas de los programas StatGraphics 5.1 y SPSS 15.01.

6.3.2 Primer modelo obtenido

Tomando como Variable dependiente IRC y como variables independientes SEXO, GRUPOETNIC, zona, DI2CAPAC, Edad2, LeeyEscribe, P1ExamenMedico, P2HaFumado, P3FumaHoy, P4PresionAlta, P5ControlPresion, P6EjercicioFisico, P8Colesterol, P10Diabetes, P11ControlDiabetes, P12Licor, P13Embarazada, P13ANovedadE, P13BControlPrenatal, P14EmbarazoAntes, P15Citologia, P16IRC, P17Cancer, P18Corazon, P21Antecedentes. Utilizando el método de mínimos cuadrados ponderados generalizados iterados se obtiene el valor estimado del predictor lineal $\eta = x_i\beta$ y el valor esperado del modelo de regresión logística

$$\hat{y} = \hat{\pi} = \frac{e^{(\hat{\eta})}}{1 + e^{(\hat{\eta})}}$$

$$= \frac{e^{(x'\beta)}}{1 + e^{(x'\beta)}}$$

$$= \frac{1}{1 + e^{(-x'\beta)}}$$

De esta manera se obtiene:

Cuadro No. 5 Modelo de Regresión Estimado

Parámetro	Estimado	Error Estándar
CONSTANTE	-12,5762	2,55482
SEXO	2,6227	1,24515
GRUPOETNIC	2,67342	6,07005
zona	2,09216	0,929367
DI2CAPAC	-4,56661	6,96041
Edad2	-2,08129	1,67052
LeeyEscribe	-0,625059	1,30872
P1ExamenMedico	-0,313302	0,995409
P2HaFumado	6,83433	2,95305
P3FumaHoy	-0,081466	2,65276
P4PresionAlta	10,6097	3,13793
P5ControlPresion	-3,63648	2,86268
P6EjercicioFisico	2,37079	1,95183
P8Colesterol	-1,47073	11,4069
P10Diabetes	5,58515	2,38238
P11ControlDiabete	5,59336	5,28983
P12Licor	2,88693	1,26581
P13Embarazada	3,92842	12,34
P13ANovedadE	-0,986815	12,9178
P13BControlPrenat	1,81785	20,097
P14EmbarazoAntes	-2,02926	2,4188

P15Citologia	-2,57764	2,13459
P16IRC	2,64726	9,61115
P17Cancer	-0,585049	22,4567
P18Corazon	-5,53522	22,3642
P21Antecedentes	1,28383	0,881018

Los datos obtenidos muestran el resultado de ajustar un modelo de regresión logística para describir la relación entre IRC y las 25 variables independientes. La ecuación de este modelo es:

$$IRC = \frac{e^t}{1+e^t}$$

donde

t = -12,5762 + 2,6227*SEXO + 2,67342***GRUPOETNIC** + 2,09216***zona** - 4,56661***DI2CAPAC** - 2,08129***Edad2** - 0,625059***LeeyEscribe** - 0,313302***P1ExamenMedico** + 6,83433***P2HaFumado** - 0,081466***P3FumaHoy** + 10,6097***P4PresionAlta** - 3,63648***P5ControlPresion** + 2,37079***P6EjercicioFisico** - 1,47073***P8Colesterol** + 5,58515***P10Diabetes** + 5,59336***P11ControlDiabetes** + 2,88693***P12Licor** + 3,92842***P13Embarazada** - 0,986815***P13ANovedadE** + 1,81785***P13BControlPrenatal** - 2,02926***P14EmbarazoAntes** - 2,57764***P15Citologia** + 2,64726***P16IRC** - 0,585049***P17Cancer** - 5,53522***P18Corazon** + 1,28383***P21Antecedentes**

Cuadro No. 6 ANAVA

Análisis de Varianza			
Fuente	**Desviación**	**G.I**	**P-Valor**
Modelo	92,79	25	0,0000
Residuos	21,88	505	1,0000
Total	114.68	530	

Dado que el p-valor para el modelo en el cuadro del Análisis de Varianza es inferior a 0.01, se puede concluir que hay una relación estadísticamente significativa entre las variables al 99% de nivel de confianza. Además, el p-valor para los residuos es mayor o igual a 0.10, indicando que el modelo no es significativamente peor que el mejor modelo posible para estos datos al 90% de nivel de confianza o superior.

El coeficiente de determinación (Cox y Snell) para modelos lineales generalizados que permite cuantificar la bondad del ajuste identificando la variación explicada por el modelo dado por :

$$R^2 = 1 - \left(\frac{L(\hat{\beta}_1)}{L(\hat{\beta})} \right)^{2/n}$$

Es igual a 80.92%, este valor nos puede suponer que el modelo planteado explica o ajusta de manera adecuada la variabilidad del mismo. Sin embargo el porcentaje ajustado (Nagelkerke) que es más adecuado para comparar modelos con diferentes números de variables independientes es 35,57%, esta situación nos lleva a tratar de mejorar el modelo identificando variables que no son significativas buscando la simplificación del modelo.

Cuadro No.7 Test de Proporción de Probabilidad

Factores	Chi-Cuadrado	G.l.	P-Valor
SEXO	-0,0582153	1	**1,0000**
GRUPOETNIC	5,96063	1	0,0146
Zona	0,051023	1	**0,8213**
DI2CAPAC	714,237	1	0,0000
Edad2	-4,74832	1	**1,0000**
LeeyEscribe	30,7707	1	0,0000
P1ExamenMedico	714,237	1	0,0000
P2HaFumado	7,75101	1	0,0054
P3FumaHoy	0,00611889	1	**0,9377**
P4PresionAlta	10,4456	1	0,0012
P5ControlPresion	1,28652	1	**0,2567**
P6EjercicioFisico	0,374929	1	**0,5403**
P8Colesterol	0,0169399	1	**0,8964**
P10Diabetes	4,82003	1	0,0281
P11ControlDiabetes	0,124943	1	**0,7237**
P12Licor	0,266267	1	**0,6058**
P13Embarazada	0,0299986	1	**0,8625**
P13ANovedadE	0,00460904	1	**0,9459**
P13BControlPrenatal	0,00381694	1	**0,9507**
P14EmbarazoAntes	15,3526	1	0,0001
P15Citologia	-5,31808	1	**1,0000**
P16IRC	0,0334199	1	**0,8549**
P17Cancer	-0,00042756	1	**1,0000**
P18Corazon	0,00151846	1	**0,9689**
P21Antecedentes	-2,612	1	**1,0000**

Observe en el cuadro No. 7 que existen variables con p-valores más altos para los test de proporción de probabilidad (marcadas con azul). Dado que p-valor es mayor o igual a 0.10, estos términos no son estadísticamente significativos al 90%

de nivel de confianza o superior. Por consiguiente, y teniendo en cuenta que 17 de las 25 variables presentan esta situación y de 12 casos que contienen el evento de interés (enfermos) seis o el 50% se clasificaron incorrectamente (ver cuadro No. 8), se buscará ajustar el modelo utilizando las técnicas propias de la regresión logística.

Cuadro No. 8 Eventos mal clasificados por el modelo

Fila	Observado	Pronosticado	
355	1	0,45	*
428	1	0,09	*
450	1	**1,00**	
452	1	0,45	*
457	1	**0,99**	
488	1	**1,00**	
514	1	0,14	*
517	1	0,42	*
520	1	**0,94**	
524	1	**1,00**	
528	1	0,35	*
531	1	**1,00**	

6.3.3 Ajuste del modelo

Basados en que los datos no se obtuvieron de ensayos clínicos sino de una encuesta y fueron validados con los registros individuales de prestación de servicios de salud RIPS, es procedente aplicar un método que nos permita depurar el modelo propuesto mediante la eliminación de las variables independientes que no le aporte significativamente a la variabilidad del mismo.

Para la selección de las variables independientes que mejor describan el comportamiento de la variable dependiente IRC se aplicó en primera instancia el método Paso hacia atrás dado que ya en las etapas anteriores definimos un modelo preliminar con todas las variables independientes, luego paso a paso se eliminaron las variables no significativas en la estimación del modelo, sin embargo por este método no se encontró una solución definitiva.

Dada esta situación se procedió a trabajar con el Método de selección por pasos hacia adelante que contrasta la entrada basándose en la significación del estadístico de puntuación y contrasta la eliminación basándose en la probabilidad

de un estadístico de la razón de verosimilitud que se basa en estimaciones condicionales de los parámetros.

Para la selección de las variables paso a paso se definieron como criterios de entrada y de salida:

P-para-introducir: 0,05
P-para-quitar: 0,05

Variables en la ecuación

- Paso 0:
    - 0 variables en el modelo.
    - 530 g.l. para el error.
    - Porcentaje de desviación explicada = 0,00%
    - Porcentaje Ajustado = 0,00%

- Paso 1:
    - **Se añade variable:** P4PresionAlta con P-para-introducir = 0,0000000245
    - 1 variable en el modelo.
    - 529 g.l. para el error.
    - Porcentaje de desviación explicada = 27,12%
    - Porcentaje Ajustado = 23,63%

- Paso 2:
    - **Se añade variable**: P2HaFumado con P-para-introducir = 0,00001369
    - 2 variables en el modelo.
    - 528 g.l. para el error.
    - Porcentaje de desviación explicada = 43,61%
    - Porcentaje Ajustado = 38,38%

- Paso 3:
    - **Se añade variable**: P10Diabetes con P-para-introducir = 0,000002755
    - 3 variables en el modelo.
    - 527 g.l. para el error.
    - Porcentaje de desviación explicada = 62,77%
    - Porcentaje Ajustado = 55,80%

- Paso 4:
    - **Se añade variable:** zona con P-para-introducir = 0,0140488
    - 4 variables en el modelo. 526 g.l. para el error.
    - Porcentaje de desviación explicada = 68,03%
    - Porcentaje Ajustado = 59,31%

- Paso 5:
    - **Se añade variable**: P12Licor con P-para-introducir = 0,0491369
    - 5 variables en el modelo. 525 g.l. para el error.
    - Porcentaje de desviación explicada = 71,41%
    - Porcentaje Ajustado = 60,94%

De esta manera se llega al Modelo final seleccionado:

Cuadro No. 9 Modelo de Regresión Ajustado

Variables en el modelo	Estimado	Error estándar
CONSTANTE	-4,19351	1540,78
zona	0,242939	1,16432
P2HaFumado	3,57053	1540,78
P4PresionAlta	3,76937	1540,78
P10Diabetes	2,14678	1089,28
P12Licor	0,235327	1,1714

$$IRC = \frac{e^t}{1+e^t}$$

donde:

t = -4,19351 + 0,242939*zona + 3,57053*P2HaFumado + 3,76937*P4PresionAlta + 2,14678* P10Diabetes + 0,235327* P12Licor

6.3.4 Medidas globales de la bondad de ajuste

Análisis de Varianza.

Cuadro No. 10 Análisis de Varianza

Análisis de Varianza			
Fuente	**Desviación**	**G.l**	**P-Valor**
Modelo	81,8927	5	0,0000
Residuos	32,7906	525	1,0000
Total	**114,683**	**530**	

Como podemos observar en cuadro No. 10, el p-valor para el modelo en la tabla del Análisis de la Varianza es inferior a 0.01, lo que nos permite concluir que hay una relación estadísticamente significativa entre las variables a un 99% de nivel de confianza. Además, el p-valor para los residuos es mayor o igual a 0.10, indicando que el modelo no es significativamente peor que el mejor modelo posible para estos datos al 90% de nivel de confianza o superior.

El R cuadrado de Cox y Snell y R cuadrado de Nagelkerke

El R cuadrado de Cox y Snell explicado por el modelo es igual a 71,41% menor que el obtenido en el primer modelo (85,92%), sin embargo el R cuadrado de Nagelkerke es 60,94% superior al 35.57% del primer modelo, si tenemos en cuenta estos valores podemos decir que el modelo ajustado explica el 60.94% de la variabilidad dándonos un buen indicio sobre la calidad del ajuste..

Test de Bondad de ajuste Chi-cuadrado

Para determinar si las funciones logísticas ajustan adecuadamente los datos observados se aplica la Prueba de Bondad de Ajuste Chi Cuadrado

Cuadro No. 11 Bondad de Ajuste

Clase	Logit	n	VERDADERO		FALSO	
	Intervalo		Observado	Esperado	Observado	Esperado
1	menor que -41,9351	252	0	252	252	
2	-41,9351 a -39,5057	90	0	90	90	
3	-39,5057 a -6,22981	92	0	0,173005	92	91,827
4	-6,22981 o mayor	96	11	10,827	85	85,173
Total		530	11	519		

Chi-cuadrado = 0,176447 con 2 g.l. p-valor = 0,915556

Dado que el p-valor es superior a 0.10, no hay razón para rechazar la adecuación del modelo ajustado para un nivel de confianza del 90% o superior.

Prueba de Hosmer-Lemeshow

Cuadro No 12 Tabla de contingencias para la prueba de Hosmer y Lemeshow

		IRC = 0		IRC = 1		Total
		Observado	Esperado	Observado	Esperado	
	1	252	252,000	0	,000	252
	2	90	90,000	0	,000	90

	3	92	91,827	0	,173	92
	4	42	42,304	1	,696	43
	5	43	42,869	11	11,131	54

Cuadro No. 13 Prueba de Hosmer-Lemeshow

	Chi-cuadrado	gl	Sig.
Modelo Ajustado	,311	3	,958

Para el modelo ajustado, esta se prueba de la hipótesis nula que el modelo se ajusta adecuadamente los datos.

Dado que la significancia de la prueba es mayor que 0,05 no se puede rechazar la hipótesis nula de que no hay diferencia significativa entre los valores observados y los que predice el modelo decir el modelo se ajusta adecuadamente a los datos

Tabla de clasificación

Cuadro No. 14 Tabla de Clasificación Observado – Pronosticado

Observado		Predicho		
		IRC		Porcentaje correcto
		0	1	0
IRC	0	519	0	100
	1	4	8	66.66
Porcentaje global				98,5

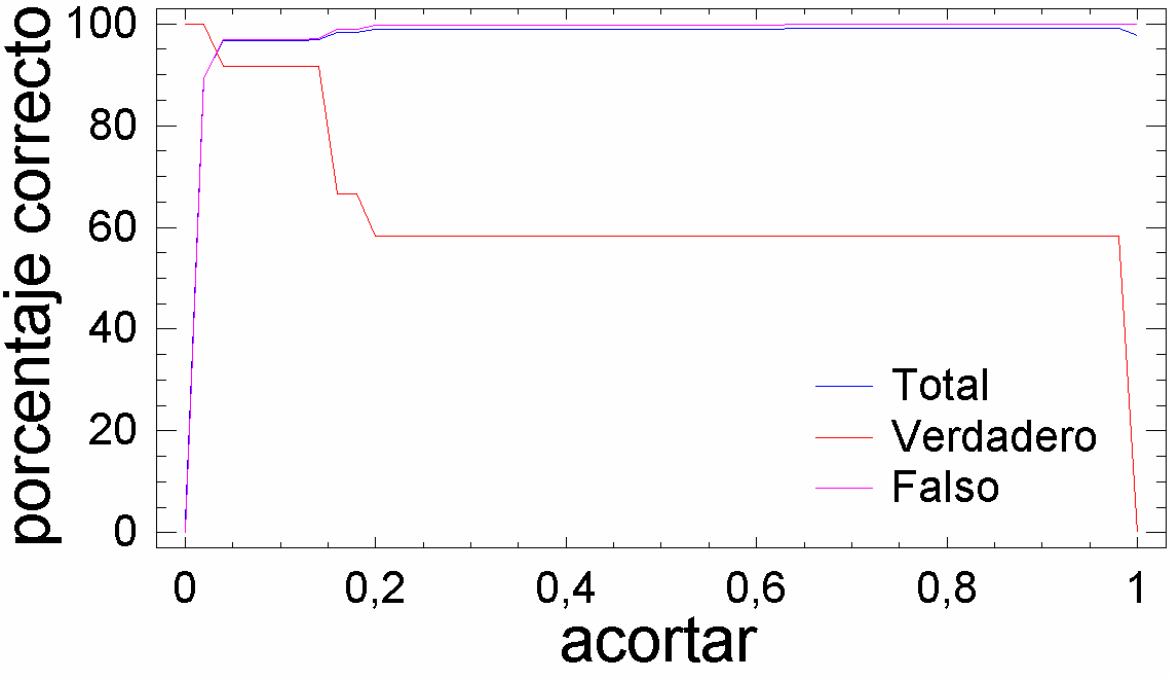
La ecuación del modelo ya diseñado nos proporciona una probabilidad que nos permite predecir a partir de ella para cada sujeto un valor de Y (Y predicho), tal que si $P(Y=1|X) \leq 0.5$ entonces Y pred= 0, y si $P(Y=1|X) > 0.5$ entonces Y pred=1.

Cuadro No. 15 Capacidad de Predicción del Modelo

Cohorte	VERDADERO	FALSO	Total
0,00	100,00	0,00	2,26
0,05	91,67	96,92	96,80
0,10	91,67	96,92	96,80
0,15	66,67	99,04	98,31
0,20	58,33	99,81	98,87
0,25	58,33	99,81	98,87
0,30	58,33	99,81	98,87
0,35	58,33	99,81	98,87

0,40	58,33	99,81	98,87
0,45	58,33	99,81	98,87
0,50	66,66	100,00	98,50
0,55	66,66	100,00	98,50
0,60	66,66	100,00	98,50
0,65	66,66	100,00	99,06
0,70	66,66	100,00	99,06
0,75	66,66	100,00	99,06
0,80	66,66	100,00	99,06
0,85	66,66	100,00	99,06
0,90	66,66	100,00	99,06
0,95	66,66	100,00	99,06
1,00	0,00	100,00	97,74

Figura No. 2 Capacidad de Predicción del Modelo



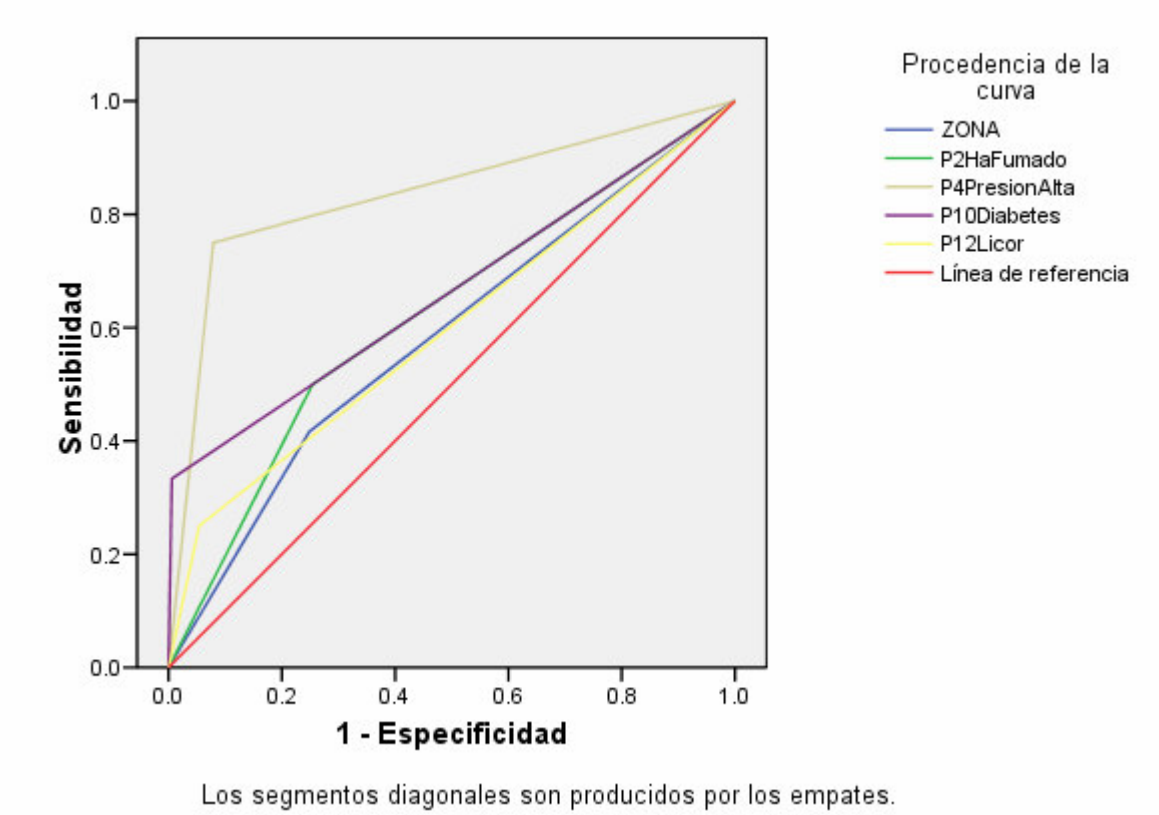
En el cuadro No. 15 se observa la capacidad de predicción del modelo ajustado. Si el valor predicho es más grande que el punto de cohorte, la respuesta se predice como VERDADERA. Si el valor predicho es inferior a o igual al punto de cohorte, la repuesta se predice para ser FALSO. La tabla muestra el porcentaje de datos observados predichos correctamente a diferentes puntos de cohorte. En nuestro caso con el puntoi de cohorte utilizado 0,5, el 66,66% de todas las respuestas VERDADERAS se predijeron correctamente, mientras 100,0% de todas las respuestas FALSAS se predijeron correctamente, para un total de 98,50%.

Los valores predichos de Y se contrastan con los valores reales de Y (Y observados), obteniendo una tabla de 2x2 de la que es posible determinar la tasa global de clasificaciones correctas para nuestro caso el 98,5% de los casos se clasifican correctamente.

Cuadro No. 16 Listado por casos Mal Clasificados

Caso	Observado IRC	Pronosticado	Grupo Pronosticado
428	1	,020	0
514	1	,140	0
517	1	,140	0
520	1	,140	0

Figura No. 3 Área bajo la curva ROC



En la curva ROC se enfrenta en un sistema de ejes la sensibilidad (en el eje y), al complementario de la especificidad (en el eje x). Se pretende determinar las correspondientes tablas de clasificación de puntos de corte de P(Y=1|X) crecientes

(0.1, 0.2, … , 0.8, 0.9, 1), y determinar a partir de ellas las correspondientes sensibilidades y especificidades.

Cuadro No. 17 Área bajo la curva

Variables resultado de contraste	Área	Error típ.(a)	Sig. asintótica(b)	Intervalo de confianza asintótico al 95%	
				Límite superior	Límite inferior
ZONA	,584	,088	,319	,412	,756
P2HaFumado	,623	,087	,145	,453	,793
P4PresionAlta	,836	,074	,000	,691	,980
P10Diabetes	,664	,096	,052	,475	,853
P12Licor	,598	,093	,245	,415	,781

La evaluación de la capacidad predictiva del modelo se realiza comparando la forma de las curvas y el área bajo las mismas; en este contexto y bajo los siguientes criterios:

- las mejores curvas serán aquellas con área más próxima a la unidad.
- Un área de 0.5 implica ausencia de discriminación
- Entre 0.7 y 0.79 es una discriminación aceptable
- Entre 0.8 y 0.89 es excelente;
- 0.9 ó superior es una discriminación excepcional.

Si realizamos el análisis por cada una de las variables que componen el modelo se puede observar que P12Licor, Zona no aportan significativamente en la capacidad de predicción del modelo. Las variables P2HaFumado y P10Diabetes se acercan a una discriminación aceptable mientras que la variable P4PresiónAlta es la que mayor poder de discriminación presenta.

Matriz de correlación para los coeficientes estimados

Cuadro No.18 Matriz de Correlación

	CONSTANTE	Zona	P2HaFumado	P4PresionAlta	P10Diabetes	P12Licor
CONSTANTE	1	-0,0014	-1	-1	-0,707	-0,0012
zona	-0,0014	1	0,0008	0,0008	0,0008	0,0705
P2HaFumado	-1	0,0008	1	**1**	**0,707**	0,0007
P4PresionAlta	-1	0,0008	**1**	1	**0,707**	0,0011
P10Diabetes	-0,707	0,0008	0,707	0,707	1	0,0006
P12Licor	-0,0012	0,0705	0,0007	0,0011	0,0006	1

En el cuadro No. 18 se muestran las correlaciones estimadas entre los coeficientes en el modelo ajustado. Estas correlaciones nos muestran la presencia de serias multicorrelaciones, es decir, correlación entre las variables pronosticadas En este caso, hay 2 correlaciones con valores absolutos superiores a 0.5.

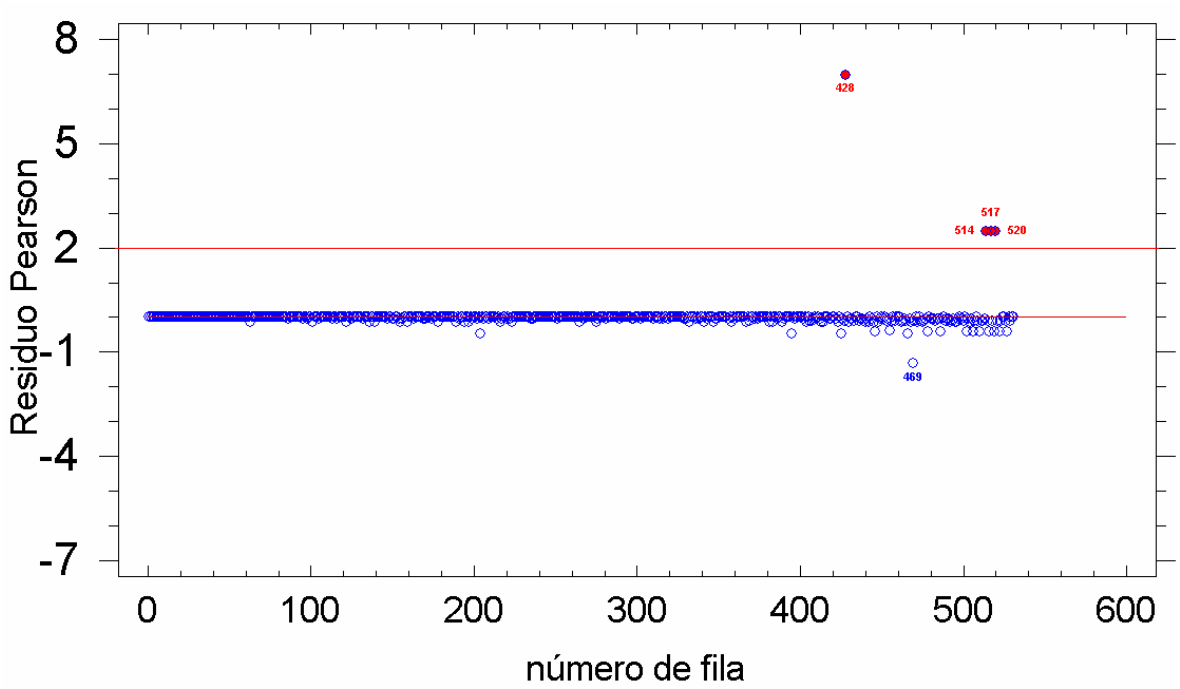
Residuos atípicos

Cuadro No 19. Residuos de Person mayores a 2

Fila	Y	Predicho Y	Residuo	Pearson Residuo	Desviación Residuo
428	1	0,0203019	0,979698	6,95	2,79
514	1	0,140394	0,859606	2,47	1,98
517	1	0,140394	0,859606	2,47	1,98
520	1	0,140394	0,859606	2,47	1,98

La tabla de residuos atípicos lista todas las observaciones que tienen resido Pearson o desviación de residuo superiores a 2.0 en valor absoluto. Estos residuos estandarizados miden cuántas desviaciones estándar de cada valor observado de IRC proceden del modelo ajustado. En este caso, hay 4 residuos estandarizados superiores a 2.0, y uno superior a 3.0. Estos casos una vez evaluados son personas que desarrollaron la enfermedad sin presentar como antecedentes los factores de riesgos definidos en el modelo es decir no sufrieron de Presión alta, No fuman ni han fumado, no padecen de diabetes y no consumen bebidas alcohólicas, por esta razón no es procedente eliminarlas del estudio dado que son casos confirmados y reales.

Figura No. 4 Residuos de Person



7. INTERPRETACION DEL MODELO PARA EL CASO DE ESTUDIO

En general podemos asegurar que el modelo presenta una adecuada bondad de ajuste dada por los estadísticos analizados anteriormente, sin embargo tiene una regular capacidad de discriminación.

Para resolver la pregunta si un determinado sujeto tiene un riesgo elevado de padecer Insuficiencia Renal crónica, utilizando una encuesta en el momento de la afiliación se deberían realizar las siguientes preguntas claves:

Cuadro No. 20 Preguntas claves para determinar la Probabilidad de IRC

No	**Pregunta**	**Código / valores**	**Variable**
1	Zona de Residencia del Encuestado	0= Urbana 1= Rural	ZONA
2	¿Ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años?	0= No 1= Sí	P2HaFumado
3	¿Alguna vez un médico o enfermera le ha dicho que ha tenido o que tiene presión alta?	0= No 1= Sí	P4PresionAlta
5	¿Alguna vez un médico le ha dicho que padece de diabetes o azúcar alta en la sangre?	0= No 1= Sí	P10Diabetes
5	¿Consume licor por lo menos una vez a la semana hasta "prenderse" o emborracharse?	0= No 1= Sí	P12Licor

Hay que tener siempre presente que el sesgo en las respuestas dadas por la persona encuestada, según los resultados obtenidos en los cruces con las bases de datos de atenciones es alta por esta razón se recomienda que la encuesta se diligencie por un profesional de la salud luego de mínimo una consulta médica general, de tal manera que la información registrada en la encuesta permita aplicando el modelo ajustado determinar el Nivel de riesgo de la persona de sufrir Insuficiencia Renal Crónica y de esta manera se puedan priorizar las actividades concernientes a la prevención de la enfermedad.

Para interpretar los resultados obtenidos es necesario indicar que los niveles ZONA= Rural, P2HaFumado = Si, P4PresionAlta = Sí, P10Diabetes = Sí y P12Licor = Sí representan los niveles de referencia usados para las variables cualitativas.

Cuadro No.21 Variables en el Modelo

Variables	B	Exp(B)
ZONA	0,242939	1,27
P2HaFumado	3,57053	35,54
P4PresionAlta	3,76937	43,35
P10Diabetes	2,14678	8,56
P12Licor	0,235327	1,27
Constante	-4,19351	0,02

Comenzaremos con la interpretación de los coeficientes de aquellas variables cualitativas que no están involucradas en la interacción como es lo apropiado, el valor 0,243 ($e^{0,243} = 1,.27$) asociado con la variable zona de residencia, indica que el residir en la zona urbana incrementa las posibilidades de padecer de Insuficiencia renal Crónica en un 27%.

Por otra parte las personas que han fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años, ($e^{3.570} = 35.51$) tienen casi 36 veces más posibilidades de desarrollar la enfermedad que las personas que no fuman

Las personas que alguna vez han sido diagnosticadas como hipertensas (presión alta), tienen 43 veces más posibilidades de desarrollar la enfermedad que una persona que no lo es ($e^{3,77} = 43,38$).

Las personas que alguna vez han sido diagnosticadas como Diabéticas (azúcar en la sangre), tienen 8 veces más posibilidades de desarrollar Insuficiencia renal crónica que una persona no diabética ($e^{2,14} = 8,56$).

Las personas que consumen licor por lo menos una vez a la semana hasta "prenderse" o emborracharse, incrementan en un 127% la probabilidad de desarrollar Insuficiencia renal crónica. ($e^{0,2353} = 1,27$)

Las dos causas más comunes de insuficiencia renal documentadas[4] en artículos de investigación clínica son la **Diabetes** (Azúcar sanguínea elevada de larga duración que es el resultado de diabetes que daña las neuronas), **Presión arterial elevada** (Presión arterial severa y de larga duración que daña los vasos capilares en los riñones) sin descartar otros factores de riesgo como Genética (antecedentes familiares), Raza: Se ven afectados más afroamericanos que caucásicos, Lupus, Uso a largo plazo de analgésicos que contengan aspirina o NSAID en altas dosis, Insuficiencia hepática, ictericia, Insuficiencia respiratoria,

[4] American Foundation for Urologic

VIH, Cáncer, Cirugía reciente a corazón abierto, Cirugía reciente en un aneurisma aórtico abdominal.

Los resultados obtenidos en el presente estudio corrobora los datos anteriores, para la población afiliada a Emssanar entre el 1 de enero al 31 de Marzo de 2008 en el municipio de Pasto, la importancia del antecedente de sufrir de Presión Arterial elevada o de Diabetes así como el hábito de fumar acompañado del consumo de bebidas alcohólicas se destacan como condicionantes vinculados con la enfermedad en estudio. La zona de residencia es sin lugar a dudas un hallazgo que se debe tener en cuenta en estudios posteriores dado que esta variable no se encuentra documentada como factor de riesgo.

Ante la necesidad planteada de desarrollar estrategias preventivas, estos resultados pueden constituir un aporte en relación a la identificación de un grupo de riesgo específico en el municipio de Pasto: Nuevos afiliados que ingresen con antecedentes de Hipertensión arterial, diabetes, fumadores y consumidores de bebidas alcohólicas que residan en la zona urbana deben ser segmentados de tal maneras que sean priorizados en los programa de Promoción y Prevención de la organización.

7.1 Cálculo de probabilidades de casos con el modelo ajustado

Teniendo en cuenta que por lo general un individuo puede presentar varios factores de riesgo, a continuación se muestra el cálculo de las diferentes posibilidades bajo el modelo ajustado.

Cuadro No.22 Casos Posibles

ZONA	P2Ha Fumado	P4Presion Alta	P10 Diabetes	P12 Licor	$\alpha + \beta_1 x_1 + \ldots + \beta_k x_k$	P(y=1)
1	1	1	1	1	5,771436	99,69%
1	1	1	1	0	5,536109	99,61%
1	1	1	0	1	3,624656	97,40%
1	1	1	0	0	3,389329	96,74%
1	1	1	1	1	5,771436	99,69%
1	1	1	1	0	5,536109	99,61%
1	1	0	0	1	-0,144714	46,39%
1	1	0	0	0	-0,380041	40,61%
1	1	0	1	1	2,002066	88,10%
1	1	0	1	0	1,766739	85,41%
1	1	0	0	1	-0,144714	46,39%
1	1	0	0	0	-0,380041	40,61%

ZONA	P2Ha Fumado	P4Presion Alta	P10 Diabetes	P12 Licor	$\alpha + \beta_1 x_1 + \ldots + \beta_k x_k$	P(y=1)
1	0	1	1	1	2,200906	90,03%
1	0	1	1	0	1,965579	87,71%
1	0	1	0	1	0,054126	51,35%
1	0	1	0	0	-0,181201	45,48%
1	0	1	1	1	2,200906	90,03%
1	0	1	1	0	1,965579	87,71%
1	0	0	0	1	-3,715244	2,38%
1	0	0	0	0	-3,950571	1,89%
1	0	0	1	1	-1,568464	17,24%
1	0	0	1	0	-1,803791	14,14%
1	0	0	0	1	-3,715244	2,38%
1	0	0	0	0	-3,950571	1,89%
0	1	1	1	1	5,528497	99,60%
0	1	1	1	0	5,29317	99,50%
0	1	1	0	1	3,381717	96,71%
0	1	1	0	0	3,14639	95,88%
0	1	1	1	1	5,528497	99,60%
0	1	1	1	0	5,29317	99,50%
0	1	0	0	1	-0,387653	40,43%
0	1	0	0	0	-0,62298	34,91%
0	1	0	1	1	1,759127	85,31%
0	1	0	1	0	1,5238	82,11%
0	1	0	0	1	-0,387653	40,43%
0	1	0	0	0	-0,62298	34,91%
0	0	1	1	1	1,957967	87,63%
0	0	1	1	0	1,72264	84,85%
0	0	1	0	1	-0,188813	45,29%
0	0	1	0	0	-0,42414	39,55%
0	0	1	1	1	1,957967	87,63%
0	0	1	1	0	1,72264	84,85%
0	0	0	0	1	-3,958183	1,87%
0	0	0	0	0	-4,19351	1,49%
0	0	0	1	1	-1,811403	14,05%
0	0	0	1	0	-2,04673	11,44%
0	0	0	0	1	-3,958183	1,87%
0	0	0	0	0	-4,19351	1,49%
0= Urbana 1= Rural				0= No 1= Sí		

Como se puede observar un individuo que reside en la zona rural de Pasto, que ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años, que sufren de presión alta, manifiesta ser diabético y consume licor por lo menos una vez a la

semana hasta “prenderse” o emborracharse tiene una probabilidad del 99,69% de sufrir insuficiencia renal crónica (Caso extremo con todos los factores de riesgo).

Por otra parte, en el caso de una persona que no manifiesta ningún factor de riesgo, si reside en la zona Urbana tiene una probabilidad de 1,49%, si reside en la zona rural tiene una probabilidad de 1,89% de enfermar.

En el caso de personas que sufren de presión alta y manifiestan ser diabéticos que residen en la zona urbana tienen una probabilidad del 84,85% y si residen en la zona rural del 87.71% de enfermar.

Una persona con solo presentar hipertensión arterial que reside en la zona urbana tiene una probabilidad del 39,55%, y si reside en la zona rural el 45,48% de enfermar.

Una persona que consume licor por lo menos una vez a la semana hasta “prenderse” o emborracharse y ha fumado por lo menos 100 cigarrillos (5 cajetillas) en los últimos 5 años tiene una probabilidad que reside en la zona urbana tiene una probabilidad del 40,43%, si reside en la zona rural del 46,46%.

8. CONCLUSIONES Y RECOMENDACIONES

En particular, el modelo logístico resultó adecuado tanto por la posibilidad de valorar simultáneamente covariables categóricas, cuanto por proporcionar interpretación epidemiológica a sus coeficientes dada su aproximación al riesgo relativo.

De este modo el recurso analítico empleado permitió no sólo hallar las asociaciones más relevantes entre las variables considerados y la Insuficiencia Renal Crónica, sino también lograr una mejor comprensión acerca de las relaciones existentes entre ellos.

Se logró demostrar que el método utilizado por Emssanar EPS-S para la aplicación de la encuesta no es el adecuado ya que no logra captar el verdadero estado de salud del afiliado en el momento de la afiliación, esta situación se puede corregir si quien registra los datos del encuestado es un profesional de la salud basado en una consulta medica general.

El funcionario encargado de aplicar la encuesta debería iniciar con un proceso de sensibilización al afiliado de tal manera que se le informe el objetivo del instrumento para de esta manera evitar los sesgos basados en el temor de ser rechazado por su condición de salud.

El modelo ajustado solo debe ser calculado únicamente con datos de afiliados residentes en el municipio de Pasto, para otros municipios se deberá buscar otros modelos ajustados a los perfiles epidemiológicos y demográficos de la zona.

Se excluyeron del modelo variables como sexo, edad, grupo étnico, ejercicio físico por no ser significativas, aunque son reconocidas en estudios como factores de riesgo, dado que la fuente primaria de los datos utilizados no proviene de evidencia médica sino de una encuesta.

9. BIBLIOGRAFÍA

Montgomery, Douglas C. P (2000). Introduction to linear regression analysis. John Wiley & Sons.

Wiley – Interscience Publication, Second Edition. Applied Logistic Regression. David Hosmer & Satnley Lemeshow.

Ediciones Díaz de Santos S.A, Excursión a la Regresión Logística en Ciencias de la Salud. Luis Carlos Silva Ayçaguer

Soriano Cabrera S. Definición y clasificación de los estadios de la enfermedad renal crónica. Prevalencia. Claves para el diagnóstico precoz. Factores de riesgo de enfermedad renal crónica. Nefrología 2004; 24 (Supl 6).

Ed. Centro de Estudios Ramón Areces, Métodos multivariantes en bioestadística Víctor Abraira Santos, Alberto Pérez de Vargas Luque, Madrid 1996

Fuentes Adicionales:

American Foundation for Urologic Disease
http://www.afud.org

Canadian Diabetes Association
http://www.diabetes.ca/