

# Modelo de Regresión Logística Multinomial y Análisis de Correspondencias Múltiple:

## Un Estudio de la Siniestrabilidad en el IPP-ULA

**Elizabeth Flores Peña**

Licenciada en Estadística  
Banco Central de Venezuela  
elizabethfp@gmail.com; eliflore@bcv.org.ve

**Surendra Sinha**

M.Sc. y Ph.D. Statistical Genetics  
Profesor Jubilado Activo del Instituto de  
Estadística Aplicada de FACES de la ULA  
sinha@ula.ve; sinha32@yahoo.com

**Luís A. Nava Puentes**

Licenciado en Estadística  
M.Sc. En Estadística Aplicada  
Profesor Agregado de la Facultad de Ciencias Económicas y Sociales.  
Universidad de los Andes.  
navalu@ula.ve

Recibido 07/07/2006 - Revisado 22/03/2007 - Aceptado 30/03/2007

### Resumen

El propósito del trabajo es estudiar la siniestrabilidad del seguro HCM implantado por el Instituto de Previsión del Profesorado de la Universidad de los Andes-IPP, para los años 2002, 2003 y 2004. Posterior al análisis exploratorio se aplicó un Análisis de Correspondencias Múltiple, en el cual se determinó que gran parte de la siniestrabilidad en "Otros Servicios", está relacionada con la póliza Básica y los grupos familiares cuyo titular labora a dedicación exclusiva, especialmente aquellos beneficiarios entre 45 y 64 años de edad, quienes a su vez tramitan el registro por farmacia, laboratorio y órdenes especiales. Por último, se ajustó un modelo de Regresión Logística Multinomial para determinar la probabilidad de ocurrencia de un servicio en función de algunas covariables, resultando el sexo y la edad del titular, factores significativos. El modelo de logits generalizado, requiere la interpretación de los cocientes odds utilizando la matriz de diseño con parametrización de efectos diferenciales. En ocasiones, la magnitud de sus dimensiones representa un serio inconveniente en dicha interpretación, no obstante, se ha aplicado una nueva metodología que prescinde de esta matriz, basada en las denominadas "*Posibilidades Generalizadas*", calculadas a partir de las probabilidades estimadas de la variable respuesta<sup>1</sup>. Entre otros puntos relevantes, el servicio por hospitalización o cirugía en relación al de gastos médicos u otros, es 22% más probable en aquellos casos cuyo titular tiene una edad comprendida entre los 45 y 65 años, comparado con los que se encuentran entre 65 y 84 años.

#### Palabras clave:

Análisis de Correspondencias Múltiple, Efectos Diferenciales, IPP, Logit Generalizado, Posibilidades, Regresión Logística Multinomial, Seguro HCM.

### Abstract

The purpose for the work is to study the siniestrability for the insurance plan HCM implemented by the Instituto de Previsión del Profesorado de la Universidad de los Andes-IPP, from the years 2002, 2003 and 2004. Following to the exploratory analysis a Multiple Analysis of Correspondences has been applied, in which there is demonstrated that a great part of the siniestrability for concept of "Other Services" in relationship to the Basic policy and the familiar groups which holder works to exclusive dedication, specially for those beneficiaries under 45 and 64 years of age, who in turn proceed with the record for drugstore, laboratory and special orders. Finally, a Model of Multinomial Regression Logistic was adjusted to determine the probability of occurrence for a determined service in relationship to certain variables, turning out to be the sex and the age of the holder, significant factors. The Model of Generalized Logits, requires the interpretation of the odds ratios using the matrix of design with parametrization of differential effects. In occasions, the magnitude of its dimensions represents a serious inconvenient in the above mentioned interpretation, nevertheless, there has applied to itself an unpublished methodology, that does without this matrix, based on the called "*Generalized Possibilities*", calculated from the probabilities estimated of the response<sup>1</sup>. Among other relevant points, the service for hospitalization or surgery in relation to that of medical expenses or others, is 22% more probable in those cases which holder has an age included between 45 and 65 years, compared with those who meet between 65 and 84 years.

#### Keywords:

Multiple Analysis of Correspondences, Differential Effects, IPP, Generalized Logit, Possibilities, Logistic Regression Multinomial, Insurance Policy HCM.

1 [Técnica inédita aplicada en el trabajo sin publicación oficial (Sinha S., 2005)]

## 1. Introducción

El IPP ha implantado el programa HCM, cuya finalidad es cubrir los gastos de los asegurados inscritos en él, por concepto de hospitalización, cirugía y maternidad, siempre y cuando tales gastos se originen como consecuencia directa de enfermedad o accidentes ocurridos durante la vigencia de la póliza y de acuerdo con los límites y condiciones establecidas. No obstante, el costo por indemnización de la siniestrabilidad se incrementa sustancialmente en el tiempo, generando un delicado estado financiero que atenta contra los convenios y los programas de salud implementados, pues las propuestas y acciones a corto plazo no han sido suficientes para afrontar el déficit de los últimos años. Esta situación ha propiciado que se emprenda una investigación formal, orientada a determinar elementos que podrían influir sobre la ocurrencia de los siniestros, e incluso para realizar ajustes acerca de la normativa a partir de políticas consistentes.

Con el propósito de establecer un criterio que permita al Instituto cumplir con sus compromisos ante el gremio profesoral sin afectar la calidad de los servicios, se ha propuesto realizar un estudio estadístico exploratorio así como también, construir un modelo de regresión logística y un análisis de correspondencias, ambos dentro del contexto multivariado. Dado que no existen antecedentes, la utilidad del trabajo se reflejará en la toma de decisiones a corto y mediano plazo, constituyendo una visión inicial del comportamiento de la siniestrabilidad del seguro HCM y el punto de partida en la realización de estudios a posteriori, en búsqueda de un modelo basado en la teoría actuarial que cuantifique el riesgo y permita ajustar en general, las primas a cancelar.

## 2. Diseño y Metodología Estadística de la Investigación

La investigación incluye todos los titulares del seguro HCM y su grupo familiar inscrito, afiliados a las diferentes pólizas y vinculados a toda categoría, dedicación o condición, sujetos a la evaluación de la cartera desde enero del 2002 hasta diciembre del 2004, sin contemplar el ajuste de las primas del año 2005.

Inicialmente se elaboró un sistema de información para facilitar la organización y el manejo de la data, diseñado según los fundamentos del modelo relacional bajo la aplicación Microsoft Access® v\_2002. Algunas herramientas tales como gráficos y tablas dinámicas, enriquecen los resultados del análisis exploratorio de la información.

El siguiente paso consiste en aplicar el análisis de correspondencias múltiple (ACM), cuyo objetivo es estudiar la asociación entre las variables de tipo cualitativo a fin de precisar la relación entre las distintas variables, orientándonos posteriormente en la formulación del modelo logístico. Ahora bien, se desea conocer cuáles son los factores que influyen en la solicitud de un servicio y la probabilidad asociada a este evento. Ciertamente, si consideramos como variable respuesta el "Tipo de Servicio", la técnica de regresión logística multinomial resulta apropiada, pues la variable es de naturaleza policotómica y no continúa como sucede en el caso de regresión lineal. Además, se puede agregar al respecto, que no es necesario cumplir con los supuestos de normalidad conjunta multivariada de las variables explicativas como sucede en el caso del análisis discriminante.

## 3. Principales Variables Bajo Estudio

a) **Sexo:** F=femenino; M=masculino.

- b) **Tipo de Personal:** Señala si el titular es empleado, profesor o viudo.
- c) **Fecha de Nacimiento:** Variable que permite conocer la edad.
- d) **Grupo Etéreo:** Se consideran 5 grupos de edades: Menor a 25 años, Entre 25 y 44 años, Entre 45 y 64 años, Entre 64 y 84 años, 84 años y más.
- e) **Dedicación del Titular:** Exclusiva, Tiempo Completo, Medio y Convencional.
- f) **Tipo de Póliza:** HCM-Básico (Empleados, Profesores); PCA (Empleados; PCA tipos A, B, C, D, E, F, G y H); Complementario.
- g) **Tipo de Servicio:** Naturaleza médica del siniestro; Hospitalización, Cirugía, Maternidad, Gastos médicos, Otros.
- h) **Tipo de Registro:** Mecanismo a través del cual el beneficiario solicita el desembolso del servicio: Órdenes Especiales (OE), Reembolsos (R), Cartas Avalués (CC), Laboratorio (L), Farmacia (FA), Préstamos HCM (PH)
- i) **Monto clínico:** Monto a cancelar por el beneficiario en la unidad médica donde se atendió el siniestro.
- j) **Monto aceptado:** Monto final que será reconocido al beneficiario por el seguro HCM, según el tipo de póliza a la cual se encuentra inscrito.
- k) **Usuario del HCM:** Si ha incurrido o no en algún tipo de servicio.
- l) **Ingresos:** Tipo de Póliza, Año, Mes del Ingreso, Descripción, Aporte ULA, Aporte CIDIAT, Acuerdo Federativo 1,5%, Aporte Profesorial, Préstamos IPP, Deducible.
- m) **Egresos:** Tipo de Póliza, Mes del Egreso, Siniestro, Reembolso,

Préstamo Profesor, Impuesto Sobre la Renta, Medicina, Fallecimiento.

## 4. Resultados

### 4.1. Análisis Exploratorio Previo

**4.1.1. De los beneficiarios inscritos en el HCM:** A finales del año 2004, 16143 personas se encontraban afiliadas al programa HCM, de las cuales 52,21% pertenecen al sexo femenino y 47,79% al masculino. Cerca del 55,4% no han incurrido en siniestro. Se totalizan 3739 grupos familiares, con un número promedio de 4 integrantes por núcleo. El promedio de edad se sitúa alrededor de los 41 años, siendo 24 la edad más común, tal como se muestra en la Figura 1. La composición porcentual de los asegurados según el grupo etéreo se puede apreciar en la Figura 2. El 75% de los hijos no sobrepasa los 28 años de edad, aunque existen ciertos casos hasta de 77 años inclusive. Existían 3926 titulares de las pólizas, de los cuales, el 75% laboraba en la Universidad a dedicación exclusiva, presentando una edad promedio de 52 años y aproximadamente 95% conformado por profesores, frente a empleados y viudos. Los afiliados con dedicación de medio tiempo sólo representan cerca del 2%, lo cual no significa poco interés por los beneficios del programa de salud, si consideramos que la proporción total del personal de medio tiempo en la Universidad es menor comparada con las demás categorías.

**4.1.2. De las Pólizas del Seguro HCM:** La póliza básica abarca el 74,24% de los asegurados en el HCM, seguida por la cobertura ampliada con 19,85% y por último el complementario tan sólo con 5,91% del total. Los afiliados al plan Básico presentan una edad promedio de 38 años aproximadamente, y se encuentra constituido principalmente por beneficiarios

de los tres primeros grupos, entre los que destacan aquellos menores a los 25 años con un 31% aproximadamente. En cuanto a los afiliados a la PCA, la edad promedio se sitúa alrededor de los 52 años de edad, oscilando la mayoría entre los 45 y 84 años con un estimado del 59%. Finalmente, en el caso del plan Complementario la edad media fue de 39 años de edad, y se tiene 42% de prevelecia en los inscritos con edades comprendidas entre los 25 y 44 años (Ver Tabla 1 y Figura 3).

**4.1.3. Sobre la Siniestrabilidad en el seguro HCM:** Para la fecha en cuestión, 44,65% de los inscritos han incurrido en siniestro desde la implementación del sistema, para un número estimado de cinco servicios promedio por beneficiario durante el período en estudio. De los 3979 grupos familiares, el 79,28% ha incurrido al menos una vez en siniestro, obteniéndose un promedio de 12 siniestros por grupo. Aún cuando ni siquiera la mitad de los beneficiarios ha utilizado el seguro, muchos de los núcleos familiares pueden estar compuestos por incluso una persona, que a su vez, genera gran cantidad de siniestros. La cantidad de siniestros aumentó a una tasa anual promedio de 34,67%. Respecto al tipo de servicio, destaca la demanda por concepto de "Otros Servicios" especialmente durante los años 2002 y 2003, predominando el 40,79% en la incidencia del grupo entre los 45 y 64 años, seguido por un 32,54% correspondiente a los asegurados entre 65 y 84 años de edad. Esto significa, que un alto porcentaje de siniestrabilidad del HCM es causado por las personas entre 45 y 84 años de edad, particularmente en la utilización de "otros servicios". Por el contrario, a partir de los 84 años se registran pocos casos de asistencia médica (5,5%), predominado los casos por hospitalización. Al prescindir de "otros servicios", los gastos médicos presentan

la mayor cantidad de casos, mientras que la maternidad escasamente alcanza el 1%, fenómeno que se mantuvo constante en el tiempo (Consulte la Figura 4).

En cuanto a los tipos de registro, las órdenes especiales y las cartas avales son las más utilizadas con 48,62% y 20,18% respectivamente. En lo referente a los préstamos HCM y la farmacia, la recurrencia es menor, pues su implementación tuvo lugar a mediados del año 2004, cuando fueron separados de las órdenes especiales. A pesar de ello, en este último año, sorprendentemente la farmacia se posiciona en segundo lugar en la demanda, lo que revela que existe una fuerte proporción de solicitudes por este concepto, motivo suficiente para revisar detalladamente el reglamento correspondiente (Ver Figura 5).

Desde otro punto de vista, el 79,13% de la siniestrabilidad estuvo representada por la póliza Básica que cubre a profesores y empleados de la Universidad, en tanto que la incidencia en el Complementario sólo alcanzó el 2,50%. La incidencia de la PCA fue de 18,37%, destacando las coberturas de 15 y 30 millones de bolívares (PCA Tipos A, B, C y D), para los hijos menores y mayores a los 26 años. Los tipos E, F, G y H no presentan casos en algunos años, puesto que los planes de 40 y 50 millones fueron creados recientemente, incluso la clasificación por empleados tuvo lugar en septiembre del 2003. Adicionalmente, se observó una tasa de siniestrabilidad creciente de la póliza básica y la de cobertura ampliada, tal como se aprecia en la Figura 6. La Tabla 2 señala que la mayor incidencia sobre la póliza Básica y la PCA, fue ocasionada por las personas entre 45 y 84 años (grupos etáreos 3 y 4), abarcando alrededor del 75% y 72% respectivamente. En el caso del plan Complementario, destaca

una preponderancia del 58,5% en los siniestros para el grupo de edades entre 25 y 44 años.

**4.1.4. Desembolso del seguro HCM y plan general de contabilidad:** En general, el monto a reconocer por el seguro será inferior al cotizado por las instituciones médicas-hospitalarias a las que asisten los asegurados. La cantidad en cuestión, se ha incrementado notablemente en los últimos años, y paralelo a ello, las dificultades de la entidad para indemnizar los mismos. En la Figura 7, se puede observar un fenómeno particular en julio, octubre y noviembre, meses en los que la liquidación de los siniestros, experimenta un brusco aumento comparados con el resto del año, especialmente para los años 2002 y 2003, puesto que parte de la liquidación del 2004 aún estaba bajo revisión al momento de la captura de la data. El comportamiento de la póliza Básica es similar al descrito en el desembolso total, lo que revela la influencia que ejerce este plan sobre el importe de los servicios. Los egresos consideran los desembolsos (clasificándolos en siniestros, reembolsos y medicinas), pero también otros compromisos importantes como impuestos sobre la renta y gastos de fallecimiento. La Figura 8, presenta las operaciones del plan de contabilidad del HCM, enfatizando como resultado del ejercicio un déficit importante en la institución durante los últimos años. La revisión de las cuentas se realiza posterior a la fecha de ejecución de las operaciones. El déficit más significativo en las operaciones de la póliza básica tuvo lugar en los años 2001 y 2002 y el mayor auge de endeudamiento de la PCA se observa en los años 2002 y 2003.

## 4.2. ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLE (ACM)

Las variables consideradas son: Sexo, Grupo Etéreo, Dedicación, Tipo de

Póliza, Tipo de Servicio, Tipo de Personal y Registro, cada una de ellas define modalidades  $p_j$ , con  $j=1, \dots, 7$ . Se desea conocer características asociadas a las modalidades de los registros del seguro HCM, razón por la cual se ha utilizado como variable suplementaria y a las demás como variables activas. Se trabajó con la tabla B o de Burt empleando el PROC CORRESP del SAS® v\_9.0. El análisis se desarrolla sobre los planos factoriales definidos por las tres primeras dimensiones. La primera dimensión representa cerca del 10% de la variación y con las dos subsiguientes se reúne el 25,36% de la inercia total.

(Tabla 3).

La mayor variabilidad en orden decreciente se presenta en la dedicación de medio tiempo, la maternidad, los viudos y el plan complementario, puesto que exhiben los valores más altos de la inercia, en consecuencia, son los más influyentes pues presentan menor número de siniestros. La contribución absoluta de las diversas categorías en la formación de las tres dimensiones, indica que el plan complementario y la cobertura adicional (PCA), conjuntamente con hospitalización, cirugía y otros servicios, son las modalidades que más intervienen en la formación del primer eje factorial. Análogamente, la formación del segundo eje depende en gran parte del segundo conjunto etéreo (entre 25 y 44 años), de la póliza complementaria, la maternidad y los empleados. Las modalidades que más influyen en la formación del tercer eje son los menores a 25 años, los servicios de maternidad y los empleados. La Figura 9, indica que los siniestros en los servicios por maternidad y gastos médicos, al igual que aquellos correspondientes a empleados, viudos y del plan complementario, se encuentran alejados del origen, señalando que son modalidades con participación baja. Podemos apreciar relación entre siniestros de individuos entre 45 y 64 años

con titulares a dedicación exclusiva y la póliza básica, farmacia, otros servicios y los registros por órdenes especiales y laboratorio.

También destaca la asociación entre el reembolso y los siniestros de personas menores a 25 años, así como aquellas entre 65 y 84 años, y titulares de grupos familiares que son profesores. Las cartas avals y los préstamos HCM en cambio, se encuentran relacionados con el grupo que supera los 84 años de edad y la póliza de cobertura ampliada PCA, con una proximidad a los servicios por hospitalización.

### 4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL

La bondad del ajuste global no favoreció un modelo que permitiera incorporar todas las variables incluidas en la etapa exploratoria. Se optó por considerar el Tipo de Servicio como variable respuesta, con tres categorías posibles: Hospitalización y Cirugía (HC), Gastos Médicos y Otros (OG), y Maternidad (M). La reagrupación de las modalidades obedece a la similitud en la naturaleza de los siniestros, lo cual evita problemas de correlación. Las covariables son el sexo y el grupo etéreo del titular<sup>2</sup>, cada una con modalidades según su definición, generando un total de 8 subpoblaciones. Dado que la escala de las variables es nominal, se ha ajustado un modelo de logits generalizado, utilizando el PROC CATMOD del SAS® v\_9.0. Sólo una iteración fue necesaria para estimar los parámetros del modelo utilizando el método de máxima verosimilitud. El análisis de varianza para el modelo de efectos principales preliminar se muestra en la Tabla 4. Acerca de la bondad de ajuste y la adecuación del modelo, vemos en principio que para el test del cociente

de verosimilitud, el estadístico  $Q_L=3,17$  con 6 grados de libertad arroja un valor  $p=0,7876$ , lo cual indica que en efecto el modelo es apropiado. De esta manera, se rechaza la hipótesis nula del contraste:

$$H_0: \underline{\beta}=0 \quad \text{Vs} \quad H_1: \underline{\beta} \neq 0 \quad \underline{\beta}=\{\beta_0; \beta_1; \beta_2\}$$

Ello indica que al menos una de las covariables influye sobre la probabilidad de ocurrencia de un tipo de servicio.

Los contrastes individuales están basados en las siguientes hipótesis:

$$H_0: \underline{\beta}_s=0 \quad \text{vs.} \quad H_1: \underline{\beta}_s \neq 0 \quad \text{con } s=0, 1, 2 \quad \text{donde } \underline{\beta} =\{\beta_0; \beta_1; \beta_2\}$$

Para cada covariable, el test basado en el  $\chi^2$  de Wald ( $Q_W$ ), indica que todos los coeficientes son significativos, con un 95% de confiabilidad. Sobre la inclusión de interacciones en el modelo de efectos principales, el coeficiente para la interacción doble resultó no significativo con un  $Q_L=3,14$  asociado a un valor  $p=0,7916$  y un nivel de significación del 5%, correspondiente al test del cociente de verosimilitud con 0 grados de libertad (modelo saturado para dos variables).

Los dos logits generalizados se construyeron a partir de la probabilidad de que acontezca un servicio por hospitalización o cirugía con respecto a uno por gastos médicos u otros, y de la probabilidad de que sea por maternidad en relación a gastos médicos u otros.

$$P [ Y = 1 / X_h ] =$$

$$\check{\delta}_{h1}(x_h) = \frac{e^{\hat{a}_1 + X_h^t \hat{a}_1}}{1 + e^{\hat{a}_1 + X_h^t \hat{a}_1} + e^{\hat{a}_1 + X_h^t \hat{a}_2}}$$

$$P[Y=2/X_h] =$$

$$\check{\delta}_{h2}(x_h) = \frac{e^{\hat{a}_2 + X_h^t \hat{a}_2}}{1 + e^{\hat{a}_1 + X_h^t \hat{a}_1} + e^{\hat{a}_2 + X_h^t \hat{a}_2}}$$

<sup>2</sup> Dado que se trata del grupo etéreo del titular, la variable ha sido decodificada para aquellos menores a los 44 años (1y2=A), entre 45 y 64 (3=B), entre 65 y 84 (4=C), y quienes superan los 84 años de edad (5=D).

$$P[Y=3/X_h] = 1 - P[Y=1/X_h] - P[Y=2/X_h]$$

Las cantidades  $\alpha_j$  y  $\beta_j$  son los parámetros de intercepto y de regresión respectivamente por cada logit, con  $j=1,2$ .

Con  $\hat{a}_1 = \{\hat{a}_1\}$  y  $\hat{a}_2 = \{\hat{a}_2, \hat{a}_3\}$ .  $\{X_{hij}\}$  es el conjunto de covariables para el h-ésimo sexo, el i-ésimo grupo etéreo del titular y el j-ésimo logit. En definitiva, las funciones logit son:

$$\text{logit}_{h1} = \log \frac{\delta_{h1}}{\delta_{h3}} = \hat{a}_1 + X_{h1}^t \hat{a}_{-1} \quad j=1$$

$$\text{logit}_{h2} = \log \frac{\delta_{h2}}{\delta_{h3}} = \hat{a}_2 + X_{h2}^t \hat{a}_{-2} \quad j=2$$

Donde  $\delta_{h1}$  es la Pr{Hospitalización o Cirugía},  $\delta_{h2}$  la Pr{Maternidad}, y  $\delta_{h3}$  la Pr{Gastos Médicos u Otros} para el hij-ésimo grupo:

$h=1,2$  para el sexo

$i=1, \dots, 4$  para el grupo etéreo del titular

$j=1,2$  que identifica a la función logit

Acercados los parámetros estimados y los contrastes individuales, se observó que los coeficientes para el grupo etéreo A (titulares que no superan los 45 años) no son significativos con 95% de confianza. El grupo C (titulares entre 45 y 64 años) resulta no significativo en el primer logit, sin embargo, presenta el mayor efecto de los grupos etéreos, específicamente en el segundo logit que compara la maternidad con respecto a los gastos médicos u otros, aun cuando en este caso su coeficiente es marginalmente significativo. La parametrización de efectos diferenciales, sigue el esquema matricial mostrado en la Figura 10, donde las dos primeras columnas corresponden a los interceptos de ambos logits generalizados, y  $\beta_q^*$  es el

efecto diferencial para la q-ésima covariable por ambos logits,  $q=1, \dots, 8$ . El primer bloque de filas odds se identifican con los parámetros para el primer logit<sub>11</sub>, y las restantes con el segundo. Análogamente, para las columnas. La celda de referencia se identifica con el valor (-1) de acuerdo a la codificación de la desviación media. Así por ejemplo, para el efecto del sexo, la función 1 corresponde al parámetro para aquellas personas del género femenino del logit<sub>h11</sub>, mientras que la función 2 es el parámetro relativo a este mismo grupo pero del segundo logit, logit<sub>h12</sub>. De igual manera para el resto de las variables.

Los cocientes odds del modelo predicho relativos a cada función logit para cada subpoblación, depende del logit que se esté considerando a partir de la matriz de diseño, exponenciando los coeficientes según la estructura por filas. En particular, la matriz de diseño de nuestro estudio tiene 8 columnas y dos más para los parámetros de intercepto de los logits, para un total de 10 columnas, mientras que el número de filas es igual a 16. De modo que incrementar la cantidad de variables, aumenta la complejidad de construir esta matriz, pues los valores dependen de las subpoblaciones a las cuales se haga referencia en el indexado de los logits.

El software presenta esta matriz para el primer logit cuando la estimación de parámetros se realiza por el método de mínimos cuadrados ponderados pero no en el caso de máxima verosimilitud (Tabla 5). No obstante, ha sido posible diseñar una nueva metodología para obtener los cocientes odds y que prescinde de la matriz de diseño, a partir de los valores estimados de probabilidad derivados del método de máxima verosimilitud, propuesta por el Dr. Sinha (2005): La opción PROB de la declaración MODEL

en el procedimiento CATMOD, presenta tales valores predichos conjuntamente con los observados y los residuos de la estimación. Ahora bien, cada uno de los valores predichos corresponde a la probabilidad de ocurrencia de los niveles de respuesta en función a las distintas categorías de las covariables. Si pensamos en calcular un cociente de cada una de esas probabilidades con respecto a las de un nivel de referencia de la respuesta por cada logit, obtendremos entonces una razón que ha sido denominada "Posibilidad Generalizada":

$$E_{hij} = \frac{P^{\hat{b}}(\text{respuesta})}{P^{\hat{b}}(\text{respuesta} = g)} = \frac{\hat{\delta}_{hij}}{\hat{\delta}_{hig}}$$

con g la modalidad de referencia en la respuesta;  
 h=1,2 ; i=1,...,4 ; j=1,2 ; g=3

Cada una de las *Posibilidades Generalizadas* corresponde a la probabilidad de ocurrencia de los niveles de respuesta en relación a una categoría de referencia, para las posibles combinaciones de las covariables. Si deseamos comparar entre dos covariables, mantenemos constante el resto de las categorías de las demás covariables, y nuevamente calculamos cocientes pero entre las posibilidades generalizadas, obteniendo de esta manera las razones odds estimadas ( $\hat{\theta}_{hij}$ ). Considerando los Gastos Médicos u Otros (GO) como nivel base de respuesta, se obtuvieron las *Posibilidades Generalizadas* para comparar con respecto a la primera categoría de respuesta (HC). Así por ejemplo, vemos que la probabilidad asociada a gastos médicos u otros, es más alta comparada con el resto de las modalidades del servicio; para ambos sexos, aquellos siniestros procedentes de titulares mayores a los 84 años de edad ostentan las cifras más elevadas, siendo

igual a 0,8443 en las mujeres y de 0,8138 en los hombres. Análogamente, para ambos géneros, la probabilidad de que ocurra un siniestro por hospitalización o cirugía predomina en los núcleos familiares cuyo titular se encuentra entre los 45 y 64 años de edad; los hombres con 0,2393 y en el caso de las mujeres siendo igual a 0,1917.

En este caso, se estima una

$$E_{121} = \frac{\hat{\delta}_{121}^E}{\hat{\delta}_{123}^E} = \frac{0,1917}{0,7952} = 0,241 \quad \text{para el sexo femenino, mientras que para el masculino}$$

$$E_{221} = \frac{\hat{\delta}_{221}^E}{\hat{\delta}_{223}^E} = \frac{0,2393}{0,7582} = 0,316$$

Entonces, a partir de las  $PG_{hi1}$  supóngase que se desea conocer la posibilidad de ocurrencia del servicio por hospitalización o cirugía en relación al de gastos médicos u otros, comparando particularmente los grupos etáreos B y C se tiene entonces:

$$\hat{\theta}_{1j1} = P_{121}/P_{131} = 0,241/0,198 = 1,2204 \quad \text{para el sexo femenino (j=1) y,}$$

$$\hat{\theta}_{2j1} = P_{221}/P_{231} = 0,316/0,259 = 1,2206 \quad \text{para el sexo masculino (j=2).}$$

Ello indica que un servicio por hospitalización o cirugía en relación al de gastos médicos u otros, es 22% más probable en los casos donde el titular tiene una edad comprendida entre los 45 y 64 años, comparado con los que se encuentran entre 65 y 84 años. Esta posibilidad se incrementa en alrededor del 39% cuando se compara a este segundo grupo con el conjunto de titulares que supera los 84 años de edad. La Tabla 6, muestra las razones de desigualdad para el contraste entre los diferentes grupos etáreos del titular para ambas modalidades del sexo, referidas al primer nivel de la respuesta (HC).

## 5. Conclusiones

Desde la perspectiva del desembolso anual, se observa un fenómeno particular en julio, octubre y noviembre, meses en los cuales, la liquidación de los siniestros, experimenta un brusco aumento en comparación con el resto del año, especialmente durante los años 2002 y 2003, con una marcada influencia del plan Básico sobre el importe de los servicios. Más del 70% de la incidencia sobre la póliza básica y la de Cobertura Adicional (PCA), recae en el grupo entre 45 y 84 años, mientras que en el caso del plan Complementario destaca una preponderancia del 59% en los siniestros del segundo grupo etéreo, con edades entre 25 y 44 años. Respecto a la PCA, destaca la incidencia de los planes con cobertura de 15 y 30 millones, abarcando cerca del 17% de la siniestrabilidad durante los últimos tres años. Si bien a principios del 2005, fue incorporado un aporte de la Universidad por insuficiencia de salud en la póliza Básica y otro aporte por acuerdo federativo por el HCM docente para la cobertura adicional, el plan de contabilidad enfatiza como resultado del ejercicio un déficit importante en la institución durante los últimos años; el mayor auge de endeudamiento en las operaciones de la póliza básica tuvo lugar en los años 2001 y 2002, y el déficit más significativo de la PCA se observa en los años 2002 y 2003.

Los trámites mas utilizados son las órdenes especiales y las cartas avales, con 48,62% y 20,18% respectivamente. A mediados del 2004 se realizó una clasificación de las órdenes especiales, sin embargo, sorprende el alto porcentaje de solicitudes por concepto de farmacia (19,8%).

A partir del ACM, se determinó

que los siniestros de individuos entre 45 y 64 años guardan algún tipo de relación con los registros por órdenes especiales, titulares de pólizas que trabajan a dedicación exclusiva, el plan Básico, así como también la farmacia, laboratorio y otros servicios. También queda evidenciada la asociación entre los siniestros de personas menores a 25 años, y del grupo entre 65 y 84 años, con los registros por reembolso, los servicios por gastos médicos y profesores titulares. Las cartas avales y los préstamos HCM en cambio, se encuentran relacionados con el grupo que supera los 84 años de edad y la PCA, con inmediatez a los servicios por hospitalización.

Entre los aspectos mas importantes, la investigación ha descartado la hipótesis según la cual, la mayor cantidad de siniestros serian atribuibles a la población más longeva (>84 años), de manera que las acciones deben estar dirigidas en primer término, al estudio particular de beneficiarios entre 45 y 84 años de edad, en particular su recurrencia por "Otros Servicios". Sólo el 5,5% de los servicios médicos están referidos a personas mayores a 84 años, predominado los casos por hospitalización.

Al respecto, si dirigimos nuestro enfoque a los titulares de las pólizas, el modelo de regresión logística multinomial demuestra que un siniestro por Hospitalización o Cirugía en relación al de Gastos Médicos u Otros, es 22% más probable en los casos donde el titular del grupo familiar tiene una edad comprendida entre los 45 y 64 años, comparado con los que se encuentran entre 65 y 84 años. Asimismo, se tiene que un servicio por Hospitalización o Cirugía en relación al de Gastos Médicos u Otros, es 39% más probable cuando el titular tiene entre 65 y 84 años, comparando con aquellos que

superan los 84.

Lo expuesto previamente, se fundamenta en las “*Posibilidades Generalizadas*”, que son el resultado de simples cocientes entre las probabilidades estimadas de la variable respuesta (Tipo de Servicio), sin tener que recurrir a la matriz de diseño con parametrización de efectos diferenciales que correspondería a un modelo de logits generalizado como en nuestro caso, constituyendo un valioso aporte para el desarrollo de este tipo de modelos en el contexto multivariado.

## 6. Recomendaciones

Estudiar nuevas alternativas en el Programa HCM para los beneficiarios del plan Básico y los titulares que laboran a dedicación exclusiva, especialmente para los afiliados entre 45 y 64 años, de acuerdo a los diferentes planes.

Considerar la redefinición del grupo familiar del plan Básico, fijando como criterio un límite de edad en los hijos inferior a los 25 años.

Establecer criterios que comprueben la veracidad en la recurrencia de los servicios por farmacia, laboratorio y otros, así como el servicio por hospitalización en la normativa de la PCA.

Organizar un grupo multidisciplinario que realice estudios oportunos garantizando la universalidad e integridad de la información.

Construir un “Modelo de Seguros por Reembolsos Cancelados”, sustentado en la teoría actuarial y la investigación estadística desarrollada en el presente estudio.

Considerar algunas medidas auxiliares, tales como la elaboración de un baremo oficial de centros médicos-asistenciales y farmacias, instalación de clínicas de salud con diversas especialidades médicas dentro del Instituto, creación de un sitio Web.

## Referencias bibliográficas

Chartered Insurance Institute (1977). **Elementos del Seguro** (2<sup>da</sup> ed.). Madrid: Editorial MAPFRE, S.A.

Díaz, G. (2000). **Análisis Estadístico de Datos Categóricos Aplicado a la Investigación Biomédica**. Bogotá: Departamento de Estadística, Universidad Nacional de Colombia.

Díaz, G. (2002). **Estadística Multivariada: Inferencia y Métodos**. (1<sup>era</sup> ed.). Bogotá: Universidad Nacional de Colombia.

Dobson, A. (1990). **An Introduction to Generalized Linear Models** (1<sup>era</sup> ed.). New South Wales: Chapman and Hall.

Gutiérrez, M. y Torres, C. (agosto, 2001). **Guía a la Quinta Edición del Publication Manual of the American Psychological Association 2001**. Pasado 3 de noviembre, 2005, de la Universidad de Puerto Rico, Facultad de Educación, disponible en: <http://biblioteca.uprrp.edu/>

Hosmer, D. and Lemeshow, S. (2000). **Applied Logistic Regression** (2<sup>da</sup> ed.). United States: John Wiley & Sons, Inc.

Instituto de Previsión al Profesorado Universidad de los Andes (julio, 2002). **Normas del Programa de Hospitalización, Cirugía y Maternidad**. (3<sup>era</sup> Ed.). Mérida: APULA-IPP.

Mark, G. (1976). **Riesgo y Seguro** (2<sup>da</sup> ed.). Madrid: Editorial MAPFRE, S.A.

Navarro, L. (1983). **Aspectos Teóricos y Una Aplicación Práctica del Análisis Factorial de Correspondencias**. Estadística Española: N° 99, España, pp. 33-59.

Peña, D. (2002). **Análisis de Datos Multivariantes** (1<sup>era</sup> ed.). Madrid: Editorial McGraw-Hill.

Rodríguez, G. (2004). **Taller de Definición de Políticas de Salud del Profesorado de la Universidad de los Andes**. Mérida, Asociación de Profesores de la Universidad de los Andes-APULA.

Salcedo, C. (2002). **Estimación de la Ocurrencia de Incidencias en Declaración de Pólizas de Importación**. Recuperado en Agosto del 2005, de <http://sisbib.unmsm.edu.pe/bibvirtual/tesis>

Salvador, M. (2003). **Modelos de Regresión con Respuesta Cualitativa: Regresión Logística**. [Lección en línea], 5campus.com, Estadística < <http://www.5campus.com/leccion/logis>> [mayo del 2005]

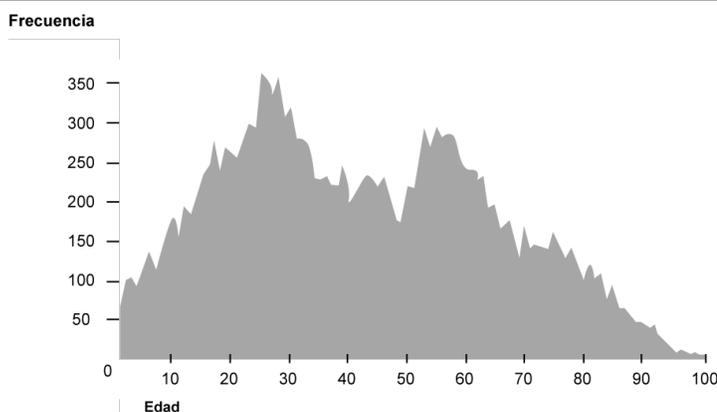
SAS Institute Inc. (2004). **SAS/STAT®9.1 User's Guide**. United States, Cary, NC: SAS Institute Inc.

Stokes, M. Davis C. and Koch, G. (2000). **Categorical Data Analysis Using the SAS System**. United States, Cary NC: SAS Institute, Inc.

Zambrano, A. (2003). **El Sistema SAS**. Mérida, Escuela de Estadística, Universidad de los Andes, pp .

Figura 1

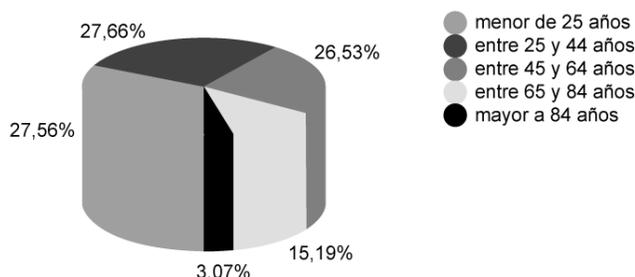
**Distribución de la edad de los beneficiarios inscritos en el HCM**



Fuente: Dirección de Informática del IPP

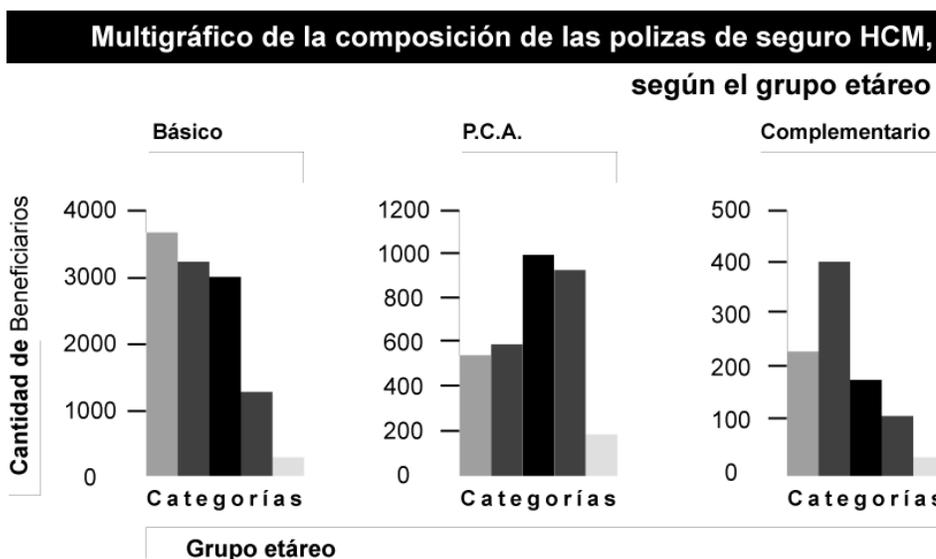
Figura 2

**Porcentaje de beneficiarios del HCM según el grupo etáreo**



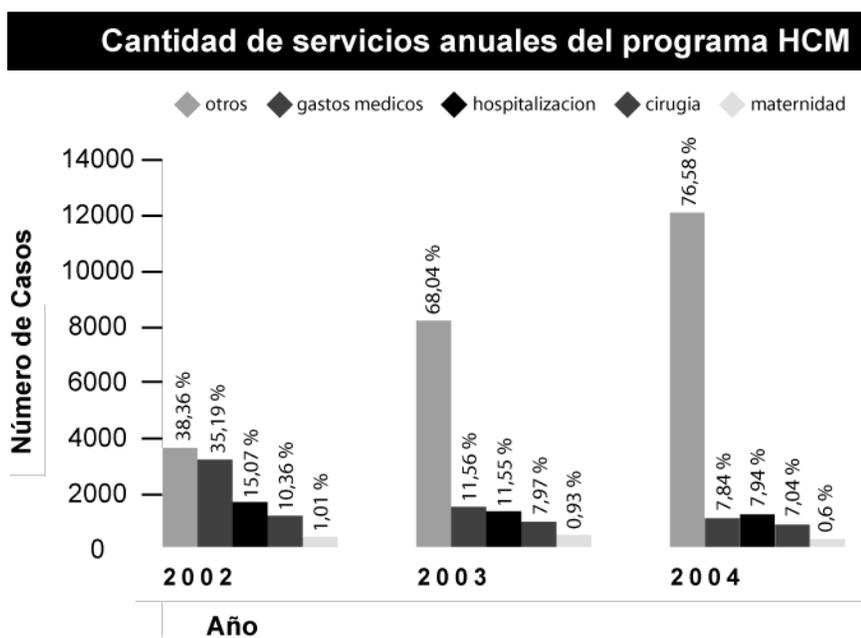
Fuente: Dirección de Informática del IPP

Figura 3



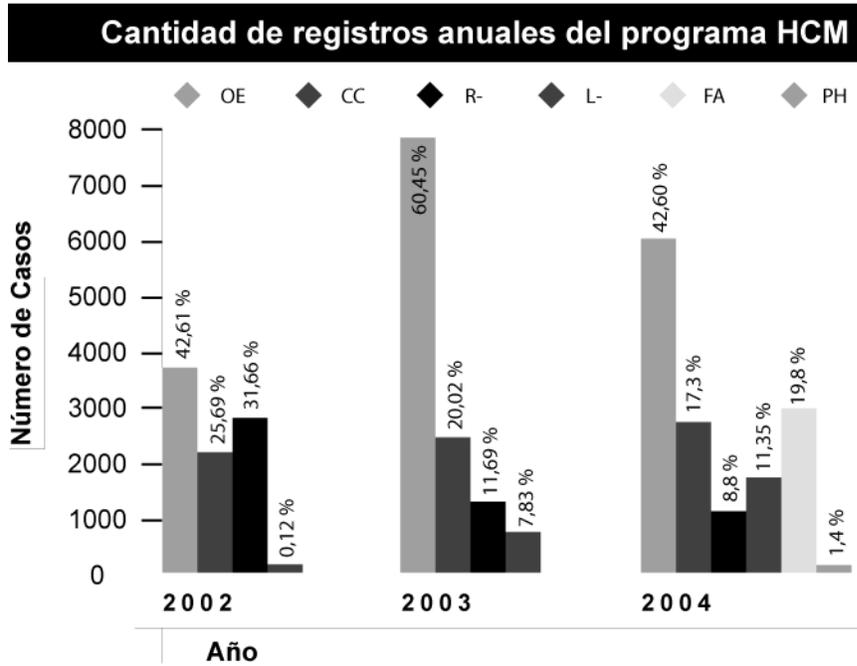
Fuente: Dirección de Informática del IPP

Figura 4



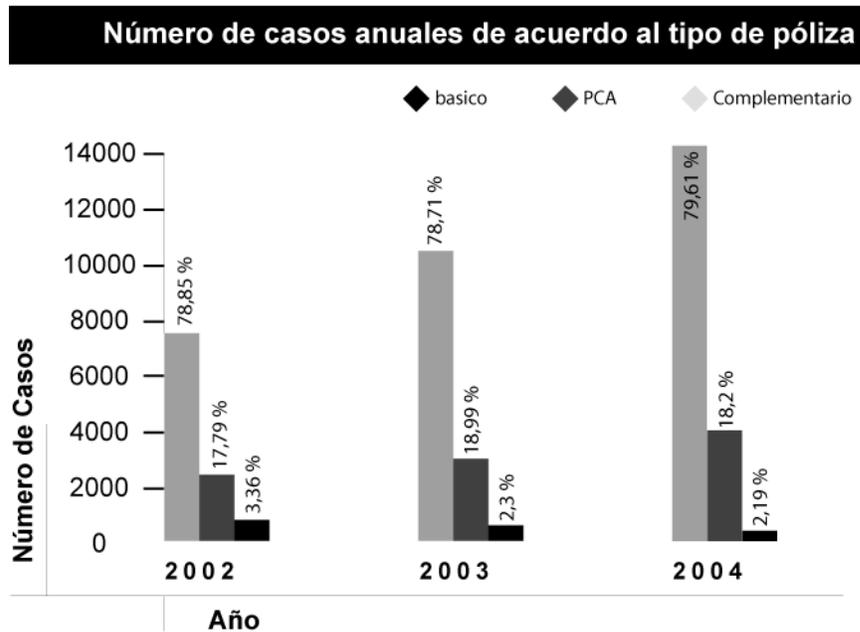
Fuente: Dirección de Informática del IPP

Figura 5



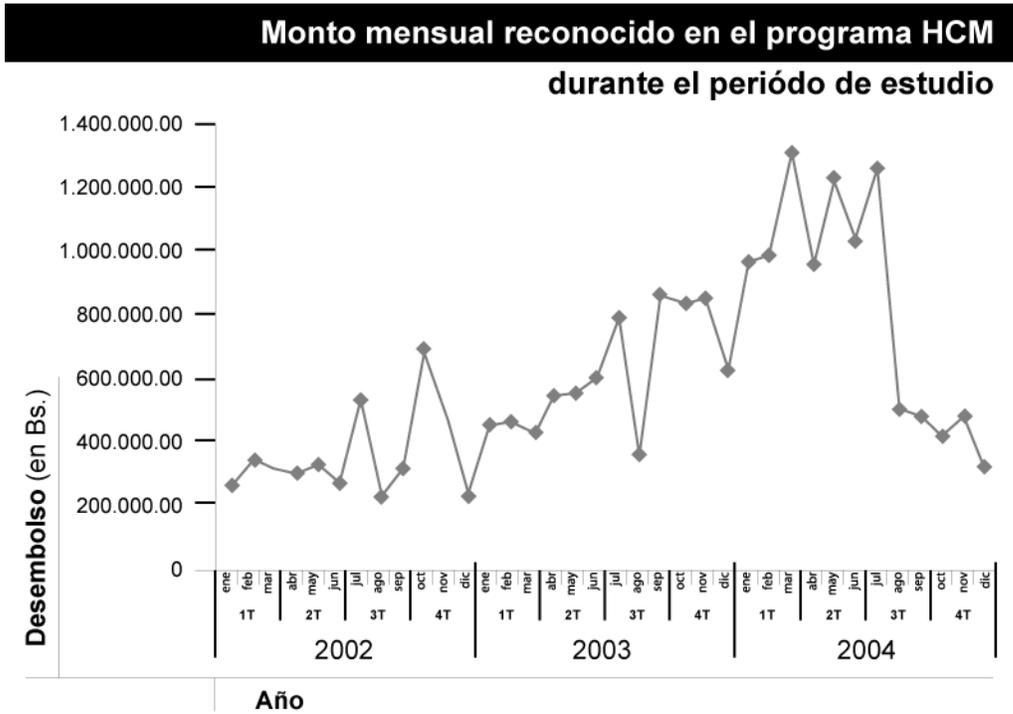
Fuente: Dirección de Informática del IPP

Figura 6



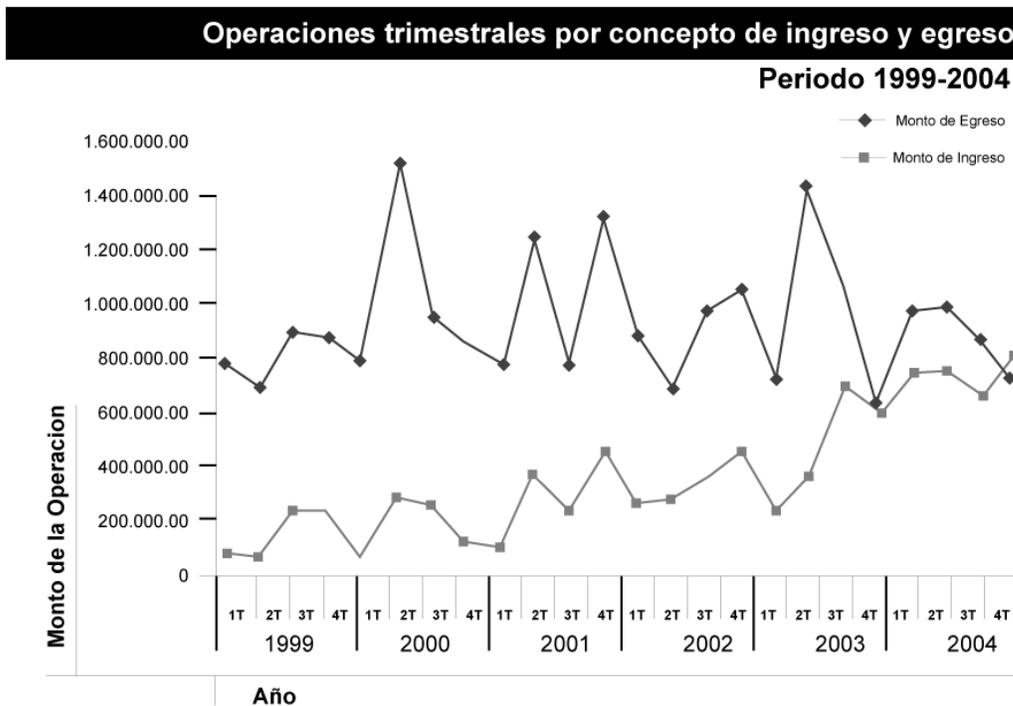
Fuente: Dirección de Informática del IPP

Figura 7



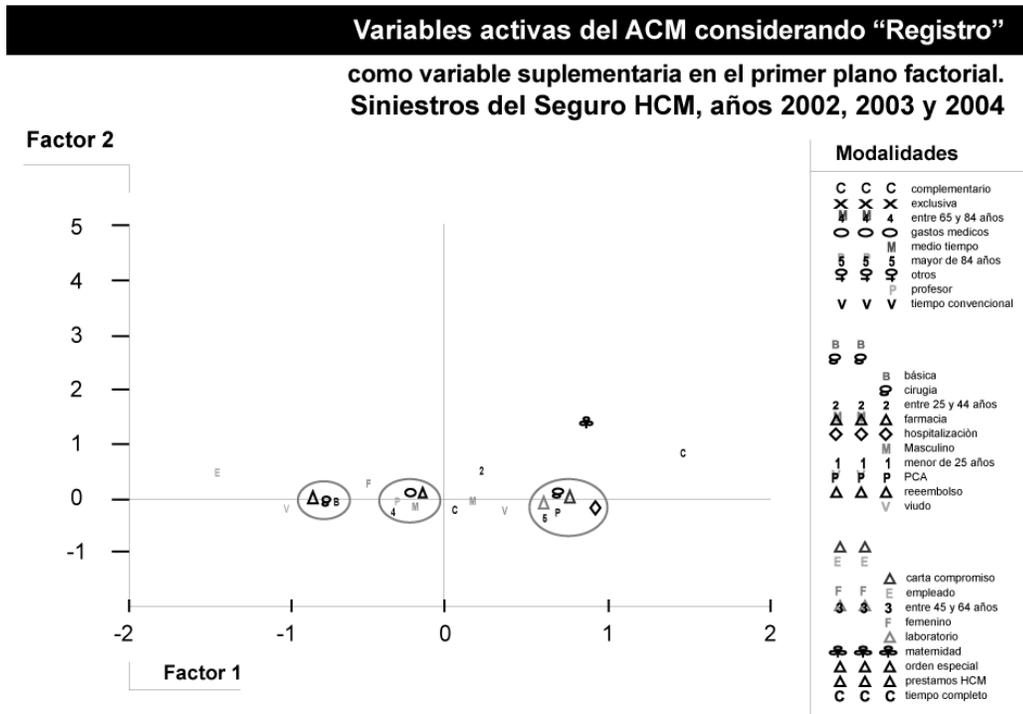
Fuente: Departamento de Contabilidad del IPP

Figura 8



Fuente: Departamento de Contabilidad del IPP

Figura 9



Fuente: Instituto de prevision social al Profesorado de la ULA - IPP

Figura 10

**Esquema de la matriz de diseño y parametrización de efectos diferenciales**

$$\begin{pmatrix} \text{logit}_{111} \\ \text{logit}_{121} \\ M \\ \text{logit}_{242} \end{pmatrix}_{(16 \times 1)} = \begin{pmatrix} 1 & 0 & A & 0 \\ 0 & 1 & 0 & M \\ M & M & 0 & M \\ 0 & 1 & A & 1 \end{pmatrix}_{(16 \times 10)} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta^*_{11} \\ M \\ \beta^*_8 \end{pmatrix}_{(10 \times 1)}$$

Fuente: Elaboración propia

**Tabla 1**

**Composición porcentual de cada una de las pólizas del seguro HCM  
en relación a los grupos de edad de los beneficiarios inscritos  
finales del 2004**

Tipo de Póliza	Grupo Etéreo					Total (%)
	Menor a 25 años	Entre 25 y 44 años	Entre 45 y 64 años	Entre 65 y 84 años	Mayor a 84 años	
BASICO	30,76%	29,24%	26,14%	11,84%	2,01%	100
PCA	16,70%	17,48%	30,34%	28,62%	6,87%	100
COMPLEMENTARIO	23,79%	41,93%	18,55%	12,16%	3,56%	100

Fuente: Elaboración propia

**Tabla 2**

**Porcentaje de siniestros de los grupos etéreos  
por cada tipo de póliza**

Tipo de Póliza	Siniestros de los Grupos Etéreos					Total (%)
	Menor a 25 años	Entre 25 y 44 años	Entre 45 y 64 años	Entre 65 y 84 años	Mayor a 84 años	
BASICO	9,66%	10,42%	43,47%	31,84%	4,60%	100
PCA	8,08%	10,66%	33,23%	38,66%	9,37%	100
COMPLEMENTARIO	14,60%	58,50%	11,68%	10,43%	4,80%	100

Fuente: Elaboración propia

**Tabla 3**

**Inercia y Descomposición Chi-Cuadrado**

Valor Singular	Inercia Principal	Chi - Cuadrado	Porcentaje	Porcentaje Acumulado	
0.50813	0.25819	62423	9.68	9.68	*****
0.47291	0.22364	54069	8.39	18.07	*****
0.44100	0.19448	47019	7.29	25.36	*****
0.42724	0.18253	44130	6.84	32.21	*****
0.42322	0.17911	43303	6.72	38.92	*****
0.42041	0.17675	42732	6.63	45.55	*****
0.40991	0.16803	40623	6.30	51.85	*****
0.40893	0.16723	40430	6.27	58.12	*****
0.40478	0.16385	39614	6.14	64.27	*****
0.40178	0.16143	39028	6.05	70.32	*****
0.39301	0.15446	37343	5.79	76.11	*****
0.38244	0.14626	35360	5.48	81.60	*****
0.38042	0.14472	34989	5.43	87.03	*****
0.36800	0.13543	32742	5.08	92.10	*****
0.33080	0.10943	26456	4.10	96.21	*****
0.31802	0.10114	24452	3.79	100.00	*****
<b>Total</b>	<b>2.66667</b>	<b>644.713</b>	<b>100.00</b>		

Grados de libertad = 441

Fuente: Elaboración propia

Tabla 4

## Análisis de varianza para el modelo de efectos principales

Fuente	DF	Chi-Cuadrado	Pr > Chisq
Término Independiente	2	8382.43	<. 0001
Sexo	2	177.88	<. 0001
etario2	6	80.28	<. 0001
Ratio de probabilidad	6	3.17	0.7876

Fuente: Elaboración propia

Tabla 5

## Matriz de diseño para el modelo de efectos diferenciales

Muestra	en el primer logit				
	1	2	3	4	5
1	1	1	1	1	0
2	1	1	0	0	0
3	1	1	0	0	1
4	1	1	-1	-1	-1
5	1	-1	1	0	0
6	1	-1	0	1	0
7	1	-1	0	0	1
8	1	-1	-1	-1	-1

Fuente: Elaboración propia

Tabla 6

## Cálculo de los cocientes odds estimados en la comparación de los grupos etáreos del titular para ambos sexos

Grupos	Sexo	$\hat{OR}_{hi1}$
		Logit 1
A y B	Femenino	$P_{111}/P_{121}=0,196/0,241=0,8110$
	Masculino	$P_{211}/P_{221}=0,256/0,316=0,8111$
B y C	Femenino	$P_{121}/P_{131}=0,241/0,198=1,2204$
	Masculino	$P_{221}/P_{231}=0,316/0,259=1,2206$
C y D	Femenino	$P_{131}/P_{141}=0,198/0,173=1,1415$
	Masculino	$P_{231}/P_{241}=0,259/0,227=1,1411$
A y C	Femenino	$P_{111}/P_{231}=0,196/0,198=1,9898$
	Masculino	$P_{211}/P_{141}=0,256/0,259=0,9900$
B y D	Femenino	$P_{121}/P_{141}=0,241/0,173=1,3931$
	Masculino	$P_{221}/P_{241}=0,316/0,227=1,3929$
A y D	Femenino	$P_{111}/P_{141}=0,196/0,173=1,1299$
	Masculino	$P_{211}/P_{241}=0,256/0,227=1,1297$

Fuente: Elaboración propia