

La Regresión Loglineal para modelos Saturados y Jerárquicos aplicada al Avalúo Inmobiliario.

Por:
Ing. Roberto Piol Puppio
CIV 32.290 / SOITAVE 260

Abstract:

Un problema al que se enfrenta cotidianamente el Tasador es el análisis de Variables Cualitativas (No Numéricas). Debido a que las Variables Cualitativas no presentan un comportamiento lineal (ni reducible a lineal); al aplicar las técnicas convencionales de regresión múltiple, pueden generar modelos irreales o simplemente no convergen a ningún valor.

La Regresión Loglineal, es un método estadístico cuyo objetivo consiste en estudiar la "Clasificación" de las Variables Cualitativas. Es esencialmente un Modelo de Regresión Lineal Múltiple entre las Variables Cualitativas y el Logaritmo Neperiano de la Frecuencia de los datos (referenciales), de la forma:

$$\text{Ln}(\text{frecuencia}) = \mu + \lambda^A + \lambda^B + \lambda^C + \lambda^{AxB} + \lambda^{AxC} + \lambda^{BxC} + \lambda^{AxBxC}$$

El Modelo anterior se corresponde a un modelo completo (Saturado) para las variables Cualitativas A, B y C; sin embargo el modelo anterior supone un modelo pesado y complejo. Por un principio elemental de parsimonia, se debe encontrar uno o mas modelos mas simples, que generen un resultado con un grado aceptable de precisión y los definimos como "Modelos Jerárquicos", en el sentido de que si un parámetro es nulo, también los serán aquellos términos de orden inferior.

El Paquete Estadístico SPSS, permite mediante un algoritmo, determinar el Modelo Saturado y el Modelo Jerárquico mas adecuado.

De esta manera obtenemos un Modelo cuya Variable de Respuesta sea el Logaritmo Neperiano de la Frecuencia ("Conteo") de los datos (referenciales) y las Variables de Diseño aquellas variables cualitativas seleccionadas para la determinación de Valor de un Inmueble.

Palabras Claves: regresión, correlación, análisis loglineal, análisis de correspondencia, dicotómica, variable cualitativa, variable categorial, análisis multivariante, tabla de contingencia, modelo jerárquico, modelo saturado.

1.0 Marco Teórico:

1.1 ¿Qué es la Regresión Loglineal?

Los Modelos Loglineal, a diferencia de los Modelos de Regresión Múltiple, permiten un mejor análisis para el caso de variables cualitativas (categóricas) dicotómicas o politómicas¹.

En Análisis Loglineal, se podría entonces definir como un Método cuyo objetivo consiste en el estudio de las relaciones entre las variables cualitativas (o no numéricas).

La Regresión o Análisis Loglineal, es un método estadístico cuyo objetivo consiste en estudiar la "Clasificación"² de las Variables Cualitativas.

La Regresión Loglineal, es esencialmente un Modelo de Regresión Lineal Múltiple entre las Variables Cualitativas y el Logaritmo Neperiano de la Frecuencia de los datos (referenciales), de la forma:

$$\ln(\text{frecuencia}) = \mu + \lambda^A + \lambda^B + \lambda^C + \lambda^{AxB} + \lambda^{AxC} + \lambda^{BxC} + \lambda^{AxBxC}$$

Donde A, B y C; son Variables Cualitativas

El uso fundamental del análisis Loglineal, consiste en determinar la contribución de las diferentes variables categoriales seleccionadas, en la conformación de una "Tabla de Contingencia".

1.2 ¿Qué son las Tablas de Contingencia?

Se define como "Tabla de Contingencia" (Crosstabulation Tables), a una combinación de dos o más tablas de distribución de frecuencia, arregladas de manera que cada celda o casilla de la Tabla resultante represente una única combinación de las "variables cruzadas (crosstabuled)".

¹ Una variable dicotómica (tal como "Con Vista al Mar" = 1 ó "Sin Vista al Mar" = 0) es un ejemplo de una variable cualitativa / categorial. Debido a que la variable cualitativa y dicotómica "Vista al Mar" no es lineal, al combinarlas con otras variables independientes, sean cuantitativas o cualitativas (tales como área de construcción, edad del inmueble, etc.) en un modelo de regresión lineal múltiple, será muy poco probable estimar o predecir correctamente la variable "Precio Unitario" en función a dichas variables mixtas.

² Cuando se utilizan variables cuantitativas (o numéricas) se habla de "Técnicas de Predicción o Estimación"; mientras que cuando se utilizan variables cualitativas o categoriales se hablaría entonces de "Técnicas de Clasificación"

De tal manera que la “Tabla de Contingencia” nos permita examinar las frecuencias observadas que pertenecen a cada una de las combinaciones específicas de dos o más variables.

Por ejemplo:

Tabla de Contingencia del análisis de la relación que existe entre la Preferencia de la población de vivir en casa o apartamento en Dos principales ciudades:

	VIVIR EN CASAS	VIVIR EN APART.	
CARACAS	10	40	50
VALENCIA	30	20	50
	50	50	100

Examinando estas frecuencias, podemos identificar las relaciones entre las variables de la “Tabla de Contingencia” (por ejemplo, La población de Caracas claramente prefiere vivir en apartamentos).

La Regresión Loglineal, nos proporciona una manera más sofisticada de analizar las “Tablas de Contingencia” y determinar las interacciones estadísticas de las variables seleccionadas.

1.3 Variables de Diseño vs. Variables de Respuesta

En las técnicas de Regresión Múltiple, se hablan de “Variables Independientes” y “Variable Dependiente”, definiendo a la Variable Dependiente como aquella que es explicada por la combinación de Variables Independientes.

En la Regresión Loglineal, no se pueden hablar de Variables Dependientes o Independientes, ya que se trata de relacionar únicamente la el Logaritmo Neperiano de la Frecuencia (ocurrencia o conteo de datos) en función de una serie de variables categoriales (cualitativas).

Entonces, hablaríamos de “Variables de Diseño” y “Variable de Respuesta”; siendo las “Variables de Diseño” aquellas variables categoriales (Dicotómicas o politómicas) que seleccionamos para construir nuestra Tabla de Contingencia y la “Variable de Respuesta” la frecuencia o ocurrencia de la data.

1.4 La Bondad de Ajuste

La bondad de ajuste de una Regresión Loglineal, se basa en la “significancia” de la desviación (residuo) entre la Frecuencia Observada de los datos y la Frecuencia Esperada que genera el modelo loglineal.

Es decir, el modelo será mejor en función de la minimización de la diferencia entre la Frecuencia Observada y la Esperada³.

$$FREC_{OBSERV} - FREC_{ESPERADA} \Rightarrow 0$$

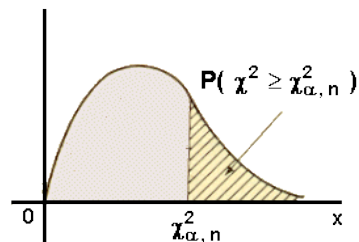
Se evaluará las Significancia (p) o “Bondad de Ajuste” de un Modelo Loglineal particular, mediante: El Test del Chi Cuadrado (χ^2) Tradicional⁴ y Estadístico de Máxima Verosimilitud de Pearson (L^2)⁵ (o Pearson Likelihood Ratio Chi-square como es su denominación en inglés).

³ Concepto análogo a la Teoría de los Residuos que se utiliza en los Métodos de Regresión Múltiple para identificar los Datos Atípicos contenidos en una serie de datos.

⁴

DISTRIBUCIÓN PROBABILIDAD CHI - CUADRADO

$$\sqrt{2 \chi_n^2} \xrightarrow{n \rightarrow \infty} N(\sqrt{2n-1}; 1)$$



⁵ La prueba de Máxima Verosimilitud de Pearson (LRT , L^2 ó G^2 tal como se le conoce en diferentes textos), es una prueba estadística de la bondad de ajuste entre dos modelos. Un modelo relativamente más complejo se compara a un modelo más simple para ver si se comparan significativamente bien para un juego de datos en particular. El LRT sólo es válido si compara a modelos jerárquicamente anidados. Es decir, el modelo más complejo sólo debe diferir del modelo simple por la adición de una o más variables; de tal forma que agregando variables adicionales se obtenga una bondad de ajuste mayor. Sin embargo, se llega a un punto cuando agregar variables adicionales NO mejorará significativamente la bondad de ajuste del modelo más simple. El LRT mantiene un criterio objetivo seleccionando entre los posibles modelos. El LRT se inicia con una comparación de los Chi Cuadrados de la forma:

$$L^2 = 2 * (Ln\chi^2_1 - Ln\chi^2_2)$$

Es decir, se evalúa la diferencia de los logaritmos neperianos de los Chi-Cuadrados, basados en la diferencia de los Grados de Libertad. El LRT sigue aproximadamente al χ^2 del modelo.

De tal manera, que se cumplan los siguientes parámetros:

- 1) Chi Cuadrado (χ^2): Máximo
- 2) Estadístico de Máxima Verosimilitud de Pearson (L^2): Máximo
- 3) Significancia (Sig.): Mínima ($p < 0.0001$)

2.0 El Modelo Loglineal Saturado

2.1 Definición de un Modelo Saturado

El Análisis o Regresión Loglineal, analiza el Logaritmo Neperiano (Ln) de la Frecuencia de cada celda o casilla de una Tabla de Contingencia, por medio de un modelo lineal.

Por lo tanto, el Ln de la frecuencia de cada celda o casilla se puede expresar como la suma de las contribuciones de las diferentes variables que intervienen en la formación del Modelo Loglineal.

Se define como Modelo Saturado (o Completo) a aquel que contiene TODOS los posibles efectos principales y TODAS las posibles combinaciones (efectos de 2do., 3er. o enésimo orden) de la Variables seleccionadas que lo componen.

Debido a que el Modelo Loglineal Saturado, puede reproducir perfectamente la data estudiada, debido a que contiene todas las posibles combinaciones de las variables seleccionadas; se supone en un modelo pesado y complejo, y usualmente no es el modelo más deseable.

Por un principio elemental de parsimonia, se debe encontrar uno o mas modelos mas simples, que generen un resultado con un grado aceptable de precisión y los definimos como "Modelos Jerárquicos" y que serán analizados mas adelante.

2.2 Ejemplo de un Modelo Loglineal Saturado

Supongamos que para un estudio de valoración de apartamentos vacacionales en una ciudad costera, queremos estudiar la relación entre las Variables Catoriales:

X: Vista al Mar

Y: Edificio con Piscina

Z: Edificio con Puestos de Estacionamiento para visitantes

La ecuación para un Modelo Loglineal Saturado, esta dado por:

$$\text{Ln}(\text{frecuencia}) = \mu + \lambda^X + \lambda^Y + \lambda^Z + \lambda^{XxY} + \lambda^{XxZ} + \lambda^{YxZ} + \lambda^{XxYxZ}$$

Donde:

- (*frecuencia*): Representa la Frecuencia de cada celda o cuadrícula de la Tabla de Contingencia producto de la interacción de los Datos (referenciales) y Variables seleccionadas
- μ : Representa el Término Independiente de la Ecuación
- λ^X : Representa el “efecto principal” de la Variable Cuantitativa o Categorical X
- λ^Y : Representa el “efecto principal” de la Variable Cuantitativa o Categorical Y
- λ^Z : Representa el “efecto principal” de la Variable Cuantitativa o Categorical Z
- λ^{XxY} : Representa el “efecto de segundo orden” o combinación de las Variables Categoricals X y Y
- λ^{XxZ} : Representa el “efecto de segundo orden” o combinación de las Variables Categoricals X y Z
- λ^{YxZ} : Representa el “efecto de segundo orden” o combinación de las Variables Categoricals Y y Z
- λ^{XxYxZ} : Representa el “efecto de tercer orden” o combinación de las Variables Categoricals X, Y y Z

2.3 Desventaja del Modelo Saturado

EL Modelo Saturado, aunque evidentemente siempre se cumple, supone un número inmanejables de ecuaciones; por ejemplo para el caso anterior Nueve Ecuaciones. Por lo tanto es un modelo pesado y sumamente complejo.

Por lo tanto, es necesaria la búsqueda de uno o varios modelos mas simples que den cuenta de dichas frecuencias con un grado de precisión aceptable para un nivel dado de confianza⁶.

Cuando se analizan Tablas de Contingencias de Cuarto Orden o mayor, la determinación del mejor modelo de Regresión Loglineal puede resultar altamente dificultoso. Aquí entraría la búsqueda de una Modelo de Correlación más simple.

⁶ Se puede citar la analogía de este concepto al del Análisis Factorial para la Regresión Múltiple. El Análisis Factorial, permite obtener un Modelo de Regresión Lineal Múltiple más simple y con menos variables, mediante la eliminación de “Variables No Significativas”, pero a la vez manteniendo un resultado aceptable a un nivel de confianza predeterminado.

3.0 Los Modelos Loglineal Jerárquicos

3.1 Definición

Se define como Modelos Loglineal Jerárquicos, a los diferentes modelos, todos sub-juegos (ecuaciones de menor orden que el Modelo Saturado) provenientes del Modelo Loglineal Saturado, que cumplan las condiciones siguientes:

- Si un parámetro es nulo, también los serán aquellos términos de orden inferior.
- Que exista completa independencia entre las variables seleccionadas⁷

Si estas condiciones se cumplen, se genera un Modelo Loglineal más sencillo, más elegante y con un grado aceptable de precisión

3.2 Relación Jerárquica

Para un Modelo con Tres (3) Variables de Diseño A, B y C, se pueden obtener una gran cantidad de Modelos Jerárquicos de Orden Inferior, como por ejemplo los siguientes:

- Modelo Jerárquico de Tercer Orden (Modelo Saturado):

$$\text{Ln}(frecuencia) = \mu + \lambda^A + \lambda^B + \lambda^C + \lambda^{AxB} + \lambda^{AxC} + \lambda^{BxC} + \lambda^{AxBxC}$$

- Modelos Jerárquico de Segundo Orden

$$\text{Ln}(frecuencia) = \mu + \lambda^A + \lambda^{AxB}$$

$$\text{Ln}(frecuencia) = \mu + \lambda^{AxB} + \lambda^C$$

$$\text{Ln}(frecuencia) = \mu + \lambda^B + \lambda^C + \lambda^{AxC}$$

$$\text{Ln}(frecuencia) = \mu + \lambda^B + \lambda^{BxC} + \lambda^{AxC}$$

- Modelos Jerárquicos de Primer Orden

$$\text{Ln}(frecuencia) = \mu + \lambda^A + \lambda^B + \lambda^C$$

$$\text{Ln}(frecuencia) = \mu + \lambda^B + \lambda^C$$

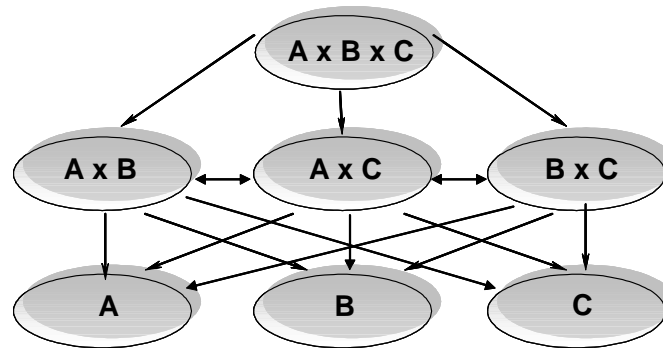
$$\text{Ln}(frecuencia) = \mu + \lambda^C$$

⁷ EL test de “Completa Independencia” implica que en el modelo jerárquico todas las variables sean independientes entre sí. Esto se obtiene al comparar los Chi-Cuadrados de cada variable, con la “Hipótesis Nula” (χ^2 obtenido de la Tabla para su correspondiente Grados de Libertad). Donde se debe cumplir que:

$$\chi^2_i \gg \chi^2_0 \quad y \quad p \leq 0.0001$$

Si lo anterior se cumple, indica que el Modelo Jerárquico independiente es significativamente diferente del Modelo Saturado, sugiriendo que ese Modelo Jerárquico contiene la variable o variables necesarias para obtener una buena correlación o ajuste

- O cualquier combinación válida:



3.3 Obtención del Modelo Jerárquico de Mejor Ajuste

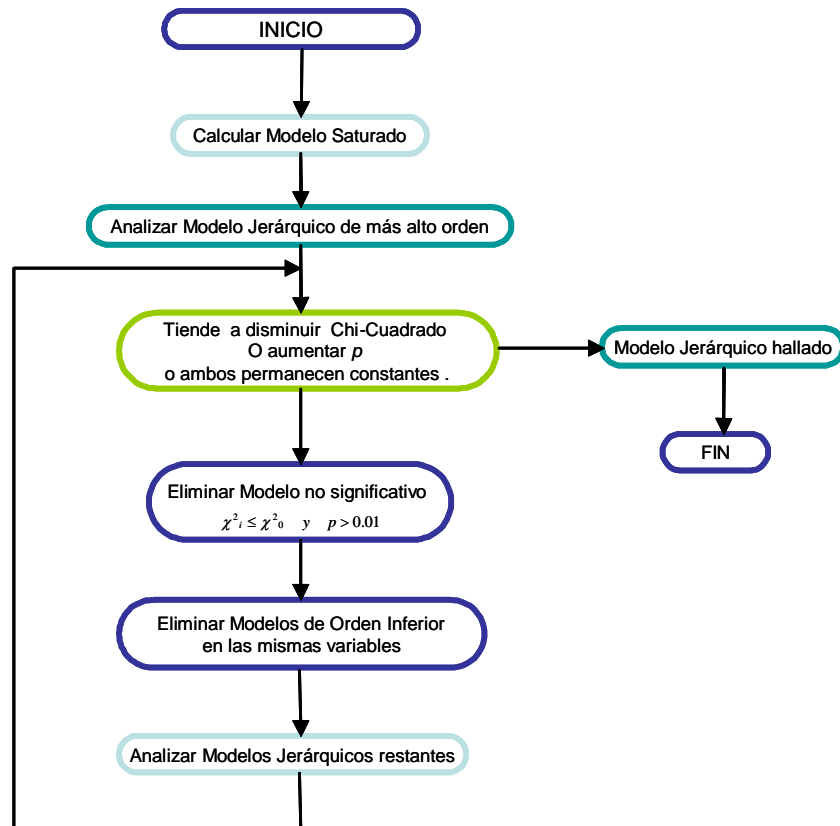
3.3.1 Método General

El Método de búsqueda del Mejor Modelo Jerárquico más utilizado por los paquetes estadísticos dedicados en el conocido como “Retro-eliminación” (Backward Elimination).

Esta metodología combina el uso de los k(ésimo)-órdenes y el test Chi - cuadrado para encontrar un Modelo Jerárquico o varios Modelos Jerárquicos significativos

La lógica del proceso es la siguiente:

- a) Se comienza calculando el Modelo Saturado.
- b) Se analiza el Modelo Jerárquico o los Modelos Jerárquicos de más alto orden
- c) Se elimina el Modelo o Modelos de ese orden que no sean significativos ($\chi^2_i \leq \chi^2_0$ y $p > 0.01$)
- d) Se eliminan los Modelos Jerárquicos de Orden Inferior en las mismas variables
- e) Se analizan los Modelos Jerárquicos restantes
- f) El proceso se continúa hasta el punto en que no puedan seguir eliminándose mas efectos sin sacrificar el poder predictivo del modelo (χ^2 (y/o) p) permanezcan constantes o tiendan a disminuir χ^2 o aumentar p .



3.3.2 Uso del SPSS en la determinación de los Modelos de Regresión Loglineal

El propósito de esta monografía es en absoluto un texto sobre las técnicas de la Regresión Loglineal.

Más bien, es la aplicación de una herramienta estadística novedosa a la resolución de problemas relacionados con la Valuación, en el caso de la utilización de variables no numéricas o cualitativas.

Por lo tanto, una vez expuesto en forma muy esquemática la teoría estadística – matemática de los Modelos Loglineales (Saturados y Jerárquicos), se pasará al estudio de un caso práctico utilizando el paquete estadístico SPSS versión 10.0⁸.

El paquete estadístico SPSS, mediante el Módulo **Loglineal**, permite la determinación de Modelos Estadísticos Saturados y Jerárquicos.

Dentro de los “Modelos Jerárquicos”, el SPSS permite o bien seleccionar automáticamente el Modelo Jerárquico de Mejor Ajuste, o permite al usuario escoger o modelar cualquier sub-juego de variables.

⁸ SPSS versión 10.0 es propiedad de SPSS, Inc. Todos los Derechos Reservados.

El desarrollo de esta monografía se basa en la determinación automática del Modelo Jerárquico de mejor ajuste, de acuerdo a un algoritmo propio de SPSS.

El SPSS utiliza para determinar el Modelo Loglineal de mejor ajuste el Método denominado “Retroeliminación” (Backward Elimination), previamente citado.

Es de hacer notar, que el algoritmo de “Retroeliminación” es largo, ocupando unas Quince (15) páginas de texto al imprimir los resultados⁹

Básicamente, se puede resumir el proceso de cálculo en los siguientes pasos generales:

- a) Enterar correctamente los datos a correlacionar en el “Editor de Datos” (Data Editor) del paquete SPSS
- b) Los procedimientos de cálculo en el SPSS se encuentran en el Menú **Analyze**
- c) Dentro del Menú Analyze se encuentra la subrutina o módulo **Loglineal**.
- d) Dentro del módulo Loglineal, se encuentran Tres (3) Submenús:
 - i. **General...**
 - ii. **Logit...**
 - iii. **Model Selection¹⁰...**
- e) Una vez ubicados en la pantalla Model Selection Loglineal Análisis, se procede a:
 - i. Seleccionar la “Variables de Diseño” que van a ser correlacionada.
 - ii. Definir los Rangos mínimos y máximos de la data que conforman cada una de las “Variables de Diseño”¹¹.
 - iii. Definir las Opciones de cálculo y salida de los modelos.
 - iv. E Iniciar el análisis mediante el botón **OK**.¹²
- f) Una vez el SPSS finalizado los cálculos, presentará la salida de los modelos en la pantalla **SPSS Viewer**, en forma de texto.

⁹ Sin embargo, el SPSS permite exportar íntegramente la salida como archivo de texto, permitiendo que el Ingeniero Tasador pueda agregar al Informe de Avalúo extractos de la misma como soporte a las operaciones.

¹⁰ El Submenú **Model Selection...**, es la subrutina o módulo donde se encuentra el algoritmo para la determinación de los Modelos Jerárquicos; sin embargo aquí se determina automáticamente el “Modelo Saturado”, ya que el Método de Retroeliminación arranca con el cálculo del Modelo Saturado, tal como ya se explicó en el texto.

¹¹ Debido a que la Regresión Loglineal es un método enfocado a la clasificación de variables categoriales, usualmente la data viene expresada como números enteros (por ejemplo: 0 y 1 en caso de variables dicotómicas).

¹² Para una mejor explicación e instrucciones detalladas de la operación del programa, refiérase al Manual de Instrucciones del SPSS

- g) De allí el usuario podrá o bien imprimirlo en papel o “exportar” la salida en un archivo de texto para su uso en un procesador de palabras u hoja de cálculo.

3.4 Interpretación de la Salida del SPSS

Quizás el problema más grande para el Ingeniero Tasador es la interpretación de la salida del SPSS.

En primer lugar, hay que tener algún conocimiento básico de lo que se pretende realizar ya que el paquete estadístico genera una gran cantidad de información, la cual hay que saberla identificar y evaluar. En segundo lugar hay que estar claro que el paquete SPSS va a generar “Tablas de Contingencia” de diferentes niveles y hay que tener el conocimiento básico para su rearmado a fin de poder interpretar los resultados del modelo.

El paquete estadístico dedicado SPSS, presenta una salida estándar¹³ la cual generalmente está compuesta por:

- a) Generación del Modelo Loglineal Saturado¹⁴
- b) Estadísticos de Control del Modelo Loglineal Saturado
 - a. Grados de Libertad (*D.F.*)
 - b. Chi Cuadrado (χ^2)
 - c. Estadístico de Máxima Verosimilitud de Pearson (L^2)
 - d. Significancia (Prob.)
 - e. Otros estadísticos de control
- c) Procedimiento de Retroeliminación para el cálculo del Model Loglineal Jerárquico de Mejor Ajuste, donde el software:
 - a. Parte del Modelo Loglineal Saturado
 - b. Analiza todas los posibles los Modelos Jerárquicos partiendo del más alto orden a mas bajo orden
 - c. Indica para cada uno de los “Modelos Jerárquicos” sus correspondientes estadísticos de control
 - d. Sugiere a final como “Modelo Jerárquico de Mejor Ajuste” aquel que tenga
 - i. El “Estadístico de Máxima Verosimilitud de Pearson” (L^2): Máximo
 - ii. La “Significancia” (p): Mínima

¹³ La salida del SPSS puede variar en función de las Opciones de cálculo y salida que determine el propio usuario

¹⁴ Los modelos Loglineal se imprimen como matrices multiniveles en la salida del paquete

- d) Generación de:
 - a. Modelo Loglineal Jerárquico de “Mejor Ajuste”¹⁵
 - b. Estadísticos de Control del Modelo Loglineal Jerárquico
 - i. Grados de Libertad (*D.F.*)
 - ii. Estadístico de Máxima Verosimilitud de Pearson (L^2)
 - iii. Significancia (Prob.)

4.0 La Regresión Loglineal para modelos Saturados y Jerárquicos aplicada al Avalúo Inmobiliario.

4.1 Aplicación de la Regresión Loglineal al Avalúo Inmobiliario

Un problema al que se enfrenta cotidianamente el Tasador es el análisis de Variables Cualitativas o Catoriales (No Numéricas), tales como Vista de apartamento, Calidad del Vecindario, Facilidades en las áreas comunes, Seguridad; la cuales siendo “Características Intangibles”, sin lugar a dudas contribuyen a la formación o destrucción del valor de un inmueble.

Debido a que las Variables Cualitativas o Catoriales no presentan un comportamiento lineal (ni reducible a lineal); al aplicar las técnicas convencionales de regresión múltiple, inclusive Técnicas de Redes Neurales; pueden generar resultados incongruentes o simplemente no converger a ningún valor.

4.2 Valoración mediante Tablas de Contingencias

Debido a que el objetivo último del análisis Loglineal, consiste en determinar la conformación de “Tablas de Contingencia”. La valoración de un inmueble estará en función de la ubicación de las características del mismo dentro de esa “Tabla de Contingencia generada”.

Por lo tanto, aquí se presentará un enfoque diferente a la Metodología Tradicional para la elaboración de un avalúo.

Hasta ahora, se analizaban las características físicas (área, edad, ubicación, etc.) del inmueble a fin de determinar su valor.

En este nuevo enfoque, el “Rango del Valor”¹⁶ del inmueble, va a depender de su posición dentro de la “Tabla de Contingencia” generada por el “Análisis Loglineal”. Debido a que la variable “Precio” (o Precio Unitario) es ahora una solo “Variable de Diseño” mas.

¹⁵ El modelo Loglineal Jerárquico se imprime como una matriz multinivel en la salida del paquete

¹⁶ Es obvio que la Variable “Precio” o “Precio Unitario” es una variable numérica o cuantitativa y por lo tanto debe ser “transformada” en una variable cualitativa o categorial. Para esto ya no se podrá hablar de una Variable Precio (Precio Unitario), sino de una Variable “Rango de Precio”.

Bajo esta nueva perspectiva, se le ha disminuido la “categoría”, que hasta ahora tenía, la Variable “Precio” (o Precio Unitario), y se ha reducido a ser un componente más de los diferentes parámetros que conjuntamente podrán definir la posición de un inmueble dentro de una tabla de contingencia y por lo tanto determinar su “Rango de Valor”.

4.3 Ejemplo de Aplicación

Debido a lo novedoso del método y la complejidad de la interpretación de las entradas / salidas de la data. Se explicará paso por paso este método, bajo el enfoque de un modelo de valoración muy sencillo.

4.3.1 Descripción del Problema

Se trata de la Valuación de apartamentos vacacionales en varias urbanizaciones similares del Barlovento venezolano (en las cercanías de Río Chico. Estado Miranda), tales como Los Canales, Las Mercedes, Lagunamar, etc.

Debido a que este es un simple ejemplo de aplicación, solo se seleccionaron una pequeña muestra de 18 datos referenciales de apartamentos muy similares entre sí.

4.3.2 Variables Catoriales Seleccionadas

Debido a que este ejemplo ilustra la demostración de la Regresión Logística, se utilizarán únicamente las Variables Catoriales o Cualitativas siguientes:

PRECIO DE VENTA: (Categorial)	MENOR DE 25 MM DE Bs.	1
	ENTRE 25 MM Y 40 MM DE Bs.	2
	ENTRE 40 MM Y 55 MM DE Bs.	3
	MAYOR DE 55 MM DE Bs.	4

MERCADO: (Categorial/Dicotómica)	APARTAMENTO USADO	0
	APARTAMENTO NUEVO	1

PISCINA: (Categorial/Dicotómica)	APARTAMENTO SIN PISCINA	0
	APARTAMENTO CON PISCINA	1

VISTA: (Categorial/Dicotómica)	OTRA VISTA	0
	VISTA AL CANAL	1

4.3.3 Referenciales Seleccionados

I.D.	URBANIZACION	VENDEDOR	COMPRADOR	DIRECCIÓN DEL INMUEBLE	AREA M2	ESTAC	PRECIO VENTA	Bs/m2	AÑO	MERC	CIRC	TOMO	No
1	LAGUNA MAR	OLGA TELEGA	SANCHEZ	APTO.11-D	45.00	1	28,000,000.00	622,222.22	1992	USADO	01	03	16
2	LOS CANALES	CARLOS SCHWARZ	ROSEMBERG	AL SOL 1 P/1 N°209	46.00	1	25,000,000.00	543,478.26	1992	USADO	01	04	16
3	LAGUNA MAR	COLMENARES	YASENKA ROJAS	DEL MAR 3 N°13B	39.00	1	17,000,000.00	435,897.44	1992	USADO	01	04	08
4	LOS CANALES	FONDO COMUN,C.A.	ZULAY SALAZAR	SUITES N°10	72.00	1	30,000,000.00	416,666.67	1996	USADO	01	04	44
5	LOS CANALES	REINALDO RONDON	RAQUEL MORALES	DE LA LAGUNA 1 P/1 N°108	85.00	1	35,000,000.00	411,764.71	1990	USADO	01	04	42
6	LOS CANALES	CLAUDIO CILIA	VICTOR GAMERO	3B P/2 N°123B	76.00	1	26,638,500.00	350,506.58	1994	USADO	01	04	13
7	LAS MERCEDES	CONST,C.A.	GINA MAZZOCCHIN	N°133	71.00	1	22,000,000.00	309,859.15	1980	USADO	01	03	47
8	LOS CANALES	FREDDY SCHEMIDT	CAROLINA MAGGI	EDIF.A PB APTO.16-A	80.00	1	24,045,000.00	300,562.50	1981	USADO	01	03	30
9	LAS MERCEDES	INV.ANECAR	NERY EGAÑA	2 P/11 N°1122	130.00	1	25,000,000.00	192,307.69	1975	USADO	01	04	21
10	RIO CHICO	EDUARDO BELLO	JORGE PALLARES	N°515	41.31	0	224,500,000.00	108,932.46	1992	USADO	01	01	04
11	LOS CANALES	EDITUR C.A.	BUCHAZER	4 APTO.46-B	94.00	1	92,026,000.00	979,000.00	1999	NUEVO	01	01	14
12	LOS CANALES	CANALES	ATENA VILLASMIL	LOS CANALES 7 PB APTO.7-B	45.00	1	33,000,000.00	733,333.33	2002	NUEVO	01	03	08
13	BARLOVENTO	CHICO,CA.	EVELIO DAPENA	CARIBE C N°2C	97.00	1	69,840,000.00	720,000.00	1998	NUEVO	01	04	09
14	BARLOVENTO	PROINCA 44, C.A.	JOSE FERNANDEZ	APTO.4	60.00	1	42,950,000.00	715,833.33	1999	NUEVO	01	03	38
15	LOS CANALES	EDITUR,C.A.	RODRIGUEZ	A PH N°3A	206.00	1	139,514,000.00	677,252.43	1999	NUEVO	01	04	15
16	LOS CANALES	CANALES	RODRIGUEZ	DELMAR LOS CANALES 5 P/2	71.00	1	47,782,545.00	672,993.59	2002	NUEVO	01	04	35
17	LAGUNA MAR	C.A.	GERSON CARRERO	19 APTO.219-B	49.00	1	20,438,470.00	417,111.63	1998	NUEVO	01	04	31
18	LOS CANALES	,C.A.	MYRIAN ESTABA	MARAZTEC V3	79.00	1	25,782,705.00	326,363.35	1999	NUEVO	01	04	02

Todos los Referenciales fueron tomados de la Oficina Subalterna de Registro Público de Río Chico y se corresponden a Documentos protocolizados el primer Trimestre del año 2,003.

4.3.4 Codificación de la Data

De acuerdo a los criterios utilizados en el Punto 4.3.2 con respecto a las Variables de Diseño seleccionadas, se preparará una matriz de datos codificados¹⁷ a fin de poder enterarlos en el paquete estadístico SPSS.

¹⁷ Obsérvese que se trata únicamente de Variables Categóricas o Cualitativas.

4.3.4.1 Selección de las Variables de Diseño

	<i>Variable de Diseño 1</i>	<i>Variable de Diseño 2</i>	<i>Variable de Diseño 3</i>	<i>Variable de Diseño 4</i>
I.D.	PRECIO VENTA	MERCADO	PISCINA	VISTA
1	28,000,000.00	USADO	SI	CANAL
2	25,000,000.00	USADO	SI	CANAL
3	17,000,000.00	USADO	SI	OTRA
4	30,000,000.00	USADO	SI	CANAL
5	35,000,000.00	USADO	SI	OTRA
6	26,638,500.00	USADO	NO	OTRA
7	22,000,000.00	USADO	NO	OTRA
8	24,045,000.00	USADO	NO	OTRA
9	25,000,000.00	USADO	NO	OTRA
10	24,500,000.00	USADO	NO	OTRA
11	92,026,000.00	NUEVO	SI	CANAL
12	33,000,000.00	NUEVO	SI	CANAL
13	69,840,000.00	NUEVO	SI	CANAL
14	42,950,000.00	NUEVO	SI	CANAL
15	139,514,000.00	NUEVO	SI	OTRA
16	47,782,545.00	NUEVO	SI	OTRA
17	20,438,470.00	NUEVO	NO	OTRA
18	25,782,705.00	NUEVO	NO	OTRA

4.3.4.1 Codificación de las Variables de Diseño

	<i>Variable de Diseño 1</i>	<i>Variable de Diseño 2</i>	<i>Variable de Diseño 3</i>	<i>Variable de Diseño 4</i>
I.D.	PRECIO	MERCADO	PISCINA	VISTA
1	2	0	1	1
2	2	0	1	1
3	1	0	1	0
4	2	0	1	1
5	2	0	1	0
6	2	0	0	0
7	1	0	0	0
8	1	0	0	0
9	2	0	0	0
10	1	0	0	0
11	4	1	1	1
12	2	1	1	1
13	4	1	1	1
14	2	1	1	1
15	4	1	1	0
16	3	1	1	0
17	1	1	0	0
18	2	1	0	0

4.3.5 Extractos de la Salida del Paquete Estadístico SPSS¹⁸

4.3.5.1 Especificación del Modelo Loglineal

* * * * * H I E R A R C H I C A L L O G L I N E A R * * * * *

DATA Information

18 unweighted cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
18 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
PRECIO	4	
MERCADO	2	
PISCINA	2	
VISTA	2	

4.3.5.2 Estadísticos de Control del Modelo Loglineal Saturados¹⁹

* * * * * H I E R A R C H I C A L L O G L I N E A R * * * * *

DESIGN 1 has generating class

PRECIO*MERCADO*PISCINA*VISTA

Note: For saturated models .500 has been added to all observed cells.
This value may be changed by using the CRITERIA = DELTA subcommand.

The Iterative Proportional Fit algorithm converged at iteration 1.
The maximum difference between observed and fitted marginal totals is .000
and the convergence criterion is .250

Goodness-of-fit test statistics

Likelihood ratio chi square =	.00000	DF = 0	P = 1.000
Pearson chi square =	.00000	DF = 0	P = 1.000

¹⁸ El SPSS presenta la salida en la pantalla SPSS Viewer, en forma de texto. Esta salida puede imprimirse o exportarse como un archivo de texto.

¹⁹ El Modelo Loglineal Saturado contiene todos los posibles “Efectos Principales” y todas las posibles combinaciones de “Efectos de Variables Combinadas”, por lo tanto reproduce perfectamente la data ($L^2 = 0.00$, $\chi^2 = 0.00$, $p = 1.00$). No se muestra el Modelo Loglineal Saturado en este extracto de la salida del SPSS

4.3.5.3 Método de Retroeliminación (Backward Elimination): Lista de todos los posibles efectos de 3°, 2° y 1° Orden

```

* * * * * H I E R A R C H I C A L   L O G   L I N E A R   * * * * *
Tests of PARTIAL associations.

Effect Name                                DF   Partial Chisq    Prob   Iter

PRECIO*MERCADO*PISCINA                    3         .044    .9976    3
PRECIO*MERCADO*VISTA                       3         .189    .9794    2
PRECIO*PISCINA*VISTA                       3         .000    1.0000    4
MERCADO*PISCINA*VISTA                      1         .000    1.0000    3
PRECIO*MERCADO                             3         7.869    .0488    4
PRECIO*PISCINA                             3         5.039    .1690    3
MERCADO*PISCINA                             1         .629    .4276    4
PRECIO*VISTA                               3         5.917    .1157    4
MERCADO*VISTA                              1         .777    .3779    4
PISCINA*VISTA                              1         7.530    .0061    2
PRECIO                                      3         8.089    .0442    2
MERCADO                                      1         .223    .6370    2
PISCINA                                      1         .896    .3438    2
VISTA                                       1         .896    .3438    2
-----

```

4.3.5.3 Método de Retroeliminación (Backward Elimination): Selección del Modelo Jerárquico de Mejor Ajuste²⁰

Step 9

The best model has generating class

```

PRECIO*MERCADO
PRECIO*VISTA
PISCINA*VISTA

```

Likelihood ratio chi square = 6.33129 DF = 18 P = .995

20 El paquete SPSS, llega a la determinación del Modelo Loglineal Jerárquico de “Mejor Ajuste”, partiendo del Modelo Saturado y eliminando aquellos efectos de cualquier orden que no contribuyen significativamente al modelo; de manera que el “Estadístico de Máxima Verosimilitud de Pearson” (L^2) sea Máximo y la “Significancia” (P) sea Mínima

* * * * * H I E R A R C H I C A L L O G L I N E A R * * * * *

The final model has generating class

PRECIO*MERCADO
 PRECIO*VISTA
 PISCINA*VISTA

The Iterative Proportional Fit algorithm converged at iteration 0.
 The maximum difference between observed and fitted marginal totals is .000
 and the convergence criterion is .250

Siendo el Modelo Loglineal de Mejor Ajuste el siguiente:

$$\ln(\text{frecuencia}) = \mu + \lambda^{\text{PRECIO* MERCADO}} + \lambda^{\text{PRECIO* VISTA}} + \lambda^{\text{PISCINA *VISTA}}$$

4.3.5.4 Determinación de las Frecuencias Observadas y Esperadas del Modelo Loglineal Jerárquico²¹:

Observed, Expected Frequencies and Residuals.

Factor	Code	OBS count	EXP count	Residual	Std Resid
PRECIO	1				
MERCADO	0				
PISCINA	0				
VISTA	0	3.0	2.5	.45	.28
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	1.0	1.5	-.45	-.38
VISTA	1	.0	.0	.00	.00
MERCADO	1				
PISCINA	0				
VISTA	0	1.0	.6	.36	.46
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	.0	.4	-.36	-.60
VISTA	1	.0	.0	.00	.00
PRECIO	2				
MERCADO	0				
PISCINA	0				
VISTA	0	2.0	1.7	.30	.23
VISTA	1	.0	.0	.00	.00

²¹ El modelo Loglineal Jerárquico presenta la “Tabla de Contingencia” como una matriz multinivel en la salida del paquete. Ahora se procederá a “armar” esta salida en forma matricial, para una mas fácil interpretación del mismo. Se utilizarán las “Frecuencias Esperadas” como coeficientes de la “Tabla de Contingencia”.

PISCINA	1				
VISTA	0	1.0	1.0	.03	.03
VISTA	1	3.0	3.3	-.33	-.18
MERCADO	1				
PISCINA	0				
VISTA	0	1.0	.8	.15	.16
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	.0	.5	-.48	-.70
VISTA	1	2.0	1.7	.33	.26
PRECIO	3				
MERCADO	0				
PISCINA	0				
VISTA	0	.0	.0	.00	.00
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	.0	.0	.00	.00
VISTA	1	.0	.0	.00	.00
MERCADO	1				
PISCINA	0				
VISTA	0	.0	.6	-.64	-.80
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	1.0	.4	.64	1.06
VISTA	1	.0	.0	.00	.00
PRECIO	4				
MERCADO	0				
PISCINA	0				
VISTA	0	.0	.0	.00	.00
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	.0	.0	.00	.00
VISTA	1	.0	.0	.00	.00
MERCADO	1				
PISCINA	0				
VISTA	0	.0	.6	-.64	-.80
VISTA	1	.0	.0	.00	.00
PISCINA	1				
VISTA	0	1.0	.4	.64	1.06
VISTA	1	2.0	2.0	.00	.00

Goodness-of-fit test statistics

Likelihood ratio chi square = 6.33129 DF = 18 P = .995
 Pearson chi square = 4.96161 DF = 18 P = .999

4.3.6 Ploteo de la Tabla de Contingencia²²

FRECUENCIAS ESPERADAS

PRECIO	NUevo	PISCINA	VISTA	FRECUENCIA OBSERV.	FRECUENCIA ESPERADA	DIFERENCIA (Residuo)
1	0	0	0	3	2.5	0.5
1	0	0	1	0	0.0	0.0
1	0	1	0	1	1.5	-0.5
1	0	1	1	0	0.0	0.0
1	1	0	0	1	0.6	0.4
1	1	0	1	0	0.0	0.0
1	1	1	0	0	0.4	-0.4
1	1	1	1	0	0.0	0.0
2	0	0	0	2	1.7	0.3
2	0	0	1	0	0.0	0.0
2	0	1	0	1	1.0	0.0
2	0	1	1	3	3.3	-0.3
2	1	0	0	1	0.8	0.2
2	1	0	1	0	0.0	0.0
2	1	1	0	0	0.5	-0.5
2	1	1	1	2	1.7	0.3
3	0	0	0	0	0.0	0.0
3	0	0	1	0	0.0	0.0
3	0	1	0	0	0.0	0.0
3	0	1	1	0	0.0	0.0
3	1	0	0	0	0.6	-0.6
3	1	0	1	0	0.0	0.0
3	1	1	0	1	0.4	0.6
3	1	1	1	0	0.0	0.0
4	0	0	0	0	0.0	0.0
4	0	0	1	0	0.0	0.0
4	0	1	0	0	0.0	0.0
4	0	1	1	0	0.0	0.0
4	1	0	0	0	0.0	0.0
4	1	0	1	0	0.0	0.0
4	1	1	0	1	0.4	0.6
4	1	1	1	2	2.0	0.0

²² Se utilizará la Hoja de Cálculo MS-Excel para plotear la Tabla de Contingencia.

4.3.7 Otra forma de presentar la Tabla de Continencia de una manera más fácil de comprender:

MENOR DE 25 MM DE Bs.=1

	Con Vista al Mar=1		Sin Vista al Mar=0	
	Con Piscina=1	Sin Piscina=0	Con Piscina=1	Sin Piscina=0
NUEVO=1	0.0	0.0	0.4	0.6
USADO=0	0.0	0.0	1.5	2.5

ENTRE 25 Y 40 MM DE Bs.=2

	Con Vista al Mar=1		Sin Vista al Mar=0	
	Con Piscina=1	Sin Piscina=0	Con Piscina=1	Sin Piscina=0
NUEVO=1	1.7	0.0	0.5	0.8
USADO=0	0.0	0.0	0.0	0.0

ENTRE 40 Y 55 MM DE Bs.=3

	Con Vista al Mar=1		Sin Vista al Mar=0	
	Con Piscina=1	Sin Piscina=0	Con Piscina=1	Sin Piscina=0
NUEVO=1	0.0	0.0	0.4	0.6
USADO=0	0.0	0.0	0.0	0.0

MAYOR DE 55 MM DE Bs.=4

	Con Vista al Mar=1		Sin Vista al Mar=0	
	Con Piscina=1	Sin Piscina=0	Con Piscina=1	Sin Piscina=0
NUEVO=1	2.0	0.0	0.4	0.0
USADO=0	0.0	0.0	0.0	0.0

4.3.8 Ejemplos de Aplicación:

4.3.8.1 Ejemplo Número 1:

Sea un apartamento vacacional en la urbanización Los Canales con las siguientes características:

- Apartamento Nuevo
- Vista al Canal
- Edificio con Piscina

Solución:

Se ubica en la “Tabla de Contingencia” todas las filas que incluyan las siguientes características:

- NUEVO = 1
- VISTA = 1
- PISCINA = 1

Esto conformaría el siguiente Subjuego:

PRECIO	NUEVO	PISCINA	VISTA	FRECUENCIA ESPERADA
1	1	1	1	0.0
2	1	1	1	1.7
3	1	1	1	0.0
4	1	1	1	2.0

Análisis de la “Tabla de Contingencia” para NUEVO =1,•VISTA =1 y•PISCINA =1:

Como se observa en la Tabla de Contingencia, existen Cuatro (4) posibles Rangos de Precios (PRECIO) para determinar el valor del apartamento a valuar. Pero al observar la **Cuarta Fila** se advierte que la FRECUENCIA ESPERADA es máxima.

Por lo tanto el Apartamento está ubicado en esa fila dentro de la Tabla de Contingencia y su rango de precios es PRECIO = 4. Indicando esto que El Valor de dicho inmueble, está en el Rango de **Apartamentos mayores de 55 Millones de Bolívares.**

4.3.8.2 Ejemplo Número 2:

Sea un apartamento vacacional en la urbanización Los Canales con las siguientes características:

- Apartamento Usado
- Vista al Canal
- Edificio con Piscina

Solución:

Se ubica en la “Tabla de Contingencia” todas las filas que incluyan las siguientes características:

- NUEVO = 0
- VISTA = 1
- PISCINA = 1

Esto conformaría el siguiente Subjuego:

PRECIO	NUEVO	PISCINA	VISTA	FRECUENCIA ESPERADA
1	0	1	1	0.0
2	0	1	1	3.3
3	0	1	1	0.0
4	0	1	1	0.0

Análisis de la “Tabla de Contingencia” para NUEVO =0,•VISTA =1 y•PISCINA =1:

Como se observa en la Tabla de Contingencia, existen Cuatro (4) posibles Rangos de Precios (PRECIO) para determinar el valor del apartamento a valuar. Pero al observar la **Segunda Fila** se advierte que la FRECUENCIA ESPERADA es máxima.

Por lo tanto el Apartamento está ubicado en esa fila dentro de la Tabla de Contingencia y su rango de precios es PRECIO = 2. Indicando esto que El Valor de dicho inmueble, está en el Rango de **Apartamentos entre los 25 y 40 Millones de Bolívares.**

4.3.8.3 Ejemplo Número 3:

Sea un apartamento vacacional en la urbanización Los Canales con las siguientes características:

- Apartamento Usado
- Vista al Canal
- Edificio sin Piscina

Solución:

Se ubica en la “Tabla de Contingencia” todas las filas que incluyan las siguientes características:

- NUEVO = 0
- VISTA = 1
- PISCINA = 0

Esto conformaría el siguiente Subjuego:

PRECIO	NUEVO	PISCINA	VISTA	FRECUENCIA ESPERADA
1	0	0	1	0.0
2	0	0	1	0.0
3	0	0	1	0.0
4	0	0	1	0.0

Análisis de la “Tabla de Contingencia” para $NUEVO = 0$, $VISTA = 1$ y $PISCINA = 0$:

Como se observa en la Tabla de Contingencia, existen Cuatro (4) posibles Rangos de Precios (PRECIO) para determinar el valor del apartamento a valorar. Pero al observar la Tabla de Contingencia, se advierte que la FRECUENCIA ESPERADA es 0.0 para todas las posibles combinaciones de datos.

La conclusión lógica para este caso es que no fueron suministrados suficientes datos al modelo para generar cualquier tipo de combinación entre las Variables de Diseño, para la determinación del valor de este tipo de apartamento en particular²³.

Por lo tanto, no se puede determinar el Valor de este apartamento utilizando el presente Modelo Loglineal Jerárquico²⁴

²³ No olvidar que la data está compuesta de solo 18 referenciales, ya que se trata de un modelo didáctico.

²⁴ Una analogía a este tipo de problemas es el caso de Modelos de Regresión Múltiple, que explican perfectamente un inmueble en particular; pero la estimación falla al aplicar el mismo Modelo de Regresión a otro inmueble de características físicas muy diferentes al primero.

4.0 Conclusiones

- a) Las técnicas de Regresión Loglineal Jerárquica, explican mucho mejor el comportamiento de fenómenos estadísticos que las técnicas de Regresión Múltiple o Redes Neuronales, en el caso de series de datos Catoriales.
- b) En tareas de Clasificación, los Modelos Loglineal Jerárquicos generan resultados mucho más exactos que los Modelos de Regresión Logística Múltiple, pese a ser ambos métodos destinados a series de datos Catoriales.
- c) Para series de datos heterogéneos (Catoriales y Cuantitativas mezcladas), las Redes Neuronales Artificiales y las Técnicas de Regresión Múltiple, superan ampliamente a los Modelos Loglineal Jerárquicos.
- d) El uso de los Modelos Loglineales Jerárquicos, es una poderosa herramienta para el “avalúo masivo” de bienes muebles e inmuebles, ya que permite clasificar los bienes dentro de Tablas de Contingencias preestablecidas.
- e) Aunque, el paquete estadístico SPSS es muy explícito en su manejo, funcionamiento e interpretación de los resultados. Es engoroso adaptar los Modelos Loglineales Jerárquicos como herramienta sencilla de análisis para ser usados por tasadores sin conocimientos básicos sobre la materia.



Ing. Roberto Píol Puppío
CIV 32.290 / SOITAVE 260
www.joinme.net/rpiol
E-Mail: rpiol@yahoo.com
Noviembre 2,003

BIBLIOGRAFIA

BERRIDGE D. (1994) "Assessing the goodness of fit of regression Models for ordinal categorical data", 9th International Workshop on Statistical Modelling, Exeter University. Londres.

CAMERON T. y QUIGGIN J. (1994) "Estimation using contingent valuation data from a "dichotomous choice with follow up" questionnaire". Journal of Environmental Economics and Management. Forthcoming. New York

TABACHNICK B.G. y FIDELL L. S. (1996). "Using multivariate statistics", 3rd ed. Harper Collins, New York.

GEORGE D y MALLERY P. (2000) "Spss for windows step by step". Allyn & Bacon. Massachusetts.

JOBSON J.D. (1992) "Applied multivariate data análisis. Vollume II". Springer Verlag. New York.

LOZARES C., LOPEZ P. Y BORRAS V. (1998) "La complementariedad del log-lineal y del análisis de correspondencias en la elaboración y el análisis de tipologías". Papers de la Universitat Autònoma de Barcelona. Número 55 pp. 79-93. Barcelona.

PIOL R. (1989-2002) "Métodos estadísticos aplicados a la valuación de bienes inmuebles". SOITAVE. Caracas.

PIOL R. (2002) "Redes neurales aplicadas al avalúo inmobiliario". Revista SOITAVE. Número 54 Septiembre 2,002 pp. 42-49 Caracas.

VARIOS AUTORES (1999) "Tests of independence using multiway contingency tables in spss". The University of Texas. Austin

VARIOS AUTORES (1995) "Log-lineal análisis of frequency tables" Statsoft, Inc. Electronic Textbook. Acceso <http://www.statsoft.com/textbook/stloglin.html>

WHITELEY, P. (1983) "The analysis of contingency tables". N. Schofield & P. Whitely (editors). Londres.