

Amenazas ambientales y vulnerabilidad en un contexto de variabilidad climática en Bolivia

Environmental threats and vulnerability in a variability climate context for Bolivia

*Gimmy Nardó Sanjines Tudela**

Resumen

El objetivo del presente trabajo de investigación es mostrar el impacto de las amenazas de la inundación, la sequía y la helada en el bienestar de las diferentes regiones de Bolivia, tomando a las vulnerabilidades como variables de aproximación a los indicadores socioeconómicos de bienestar.

La investigación aporta al estado del conocimiento con un nuevo método para estimar e identificar los impactos en el bienestar socioeconómico producido por el cambio climático, a partir de la combinación del análisis de la variabilidad climática con la inteligencia artificial.

Para lograr el objetivo planteado de identificar las relaciones entre amenazas, vulnerabilidades y el bienestar se construyen diversos Sistemas Expertos (los mismos que pertenecen al área de la Inteligencia Artificial) con base en el algoritmo C4.5. El algoritmo tiene como característica principal que permite ensamblar los árboles de decisión con base en tablas de datos, haciendo de esta manera el trabajo de reconocimiento de patrones mediante testores.

El reconocimiento de patrones permite describir el incremento o decremento del bienestar producido por las variables objeto de estudio.

* Docente de la Carrera de Economía e Ingeniería Ambiental perteneciente a la Universidad Católica Boliviana a nivel de Post Grado y Pregrado. Administrador de "Joint European Latin American Universities Renewable Energies Project" JELARE - Bolivia. Email: g.sanjines97@egresados.uniandes.edu.co

Palabras clave: Inteligencia Artificial, Sistemas Expertos, riesgos ambientales, Path theory, Impacto en el bienestar, impacto socioeconómico, economía del bienestar.

Abstract

The object of the present investigation paper is to show the impact of the threats of flooding, drought and frosting in the welfare of the different regions of Bolivia taking into account the vulnerabilities as variables of approximation to the different indicators socio-economical of welfare.

The investigations is important because it brings to the state of knowledge a new method to calculate and identify the impacts on the socio-economic produced by the climatic change. The method for its validation is applied in a very particular way in Bolivia. This contribution is generated by combining the analysis of the climatic variability –that basically is expressed in threats, and vulnerabilities and the risk- with artificial intelligence.

To obtain the objective manifested we must identify the relations between threats, vulnerabilities and welfare and welfare you must construct different Expert Systems (the same that belong to the area of artificial Intelligence) with the basis in the algorithm C4.5. The algorithm has as main characteristic that allows assemble the tree of decision with base in facts tables, making in this way the work of recognition of patterns.

The recognition of this pattern allows to describe, the increase or decreasing of the welfare produced by the variables object of study acquiring in this way, the objective stated.

Keywords: Artificial Intelligence, Expert Systems, Environmental Risks, Path theory, Impact on Welfare, Socio-economic, Welfare Economics.

Clasificación / Classification JEL: C45, C54, C89, D63, Q28

1. Introducción

El calentamiento del sistema climático es inequívoco, como se desprende ya del aumento observado y demostrado del promedio mundial de temperatura del aire y del océano, de la fusión generalizada de nieves y hielos, y del aumento del promedio mundial del nivel del mar. El calentamiento global, fenómeno de largo plazo, se considera que genera el escenario

propicio para la variabilidad climática, fenómeno de corto plazo. La variabilidad climática es el fenómeno que las personas perciben en su cotidiano vivir, el mismo se expresa en cambios en la temporalidad, intensidad y frecuencia en que ocurren los fenómenos atmosféricos.

Por otro lado, los fenómenos naturales tienen una incidencia social, económica y natural, lo cual los convierte en amenazas. Actualmente, la manifestación de fenómenos hidroclimáticos extremos, como efecto del cambio climático, está agravando las amenazas e incrementado la posibilidad de escenarios de desastre. Estos escenarios de desastres y su potencial ocurrencia se denominan riesgos. El riesgo, funcionalmente hablando, resulta de la composición de las amenazas y vulnerabilidades que existen en una región determinada. Estas amenazas y vulnerabilidades impactan en el bienestar de las sociedades.

Tradicionalmente, en los trabajos de investigación acerca del bienestar se enfoca el análisis en indicadores macroeconómicos, en función generalmente del Producto Interno Bruto, dejando de lado las interacciones directas e indirectas que se dan entre medio ambiente y bienestar. Empero, en los ámbitos económico y, sobre todo, social, en la actualidad también se observa una utilización creciente del concepto de vulnerabilidad, que posee casi las mismas variables incluidas en el Índice de Desarrollo Humano (IDH), el Índice de Pobreza Humana (IPH) y las Necesidades Básicas Insatisfechas (NBI), que en su conjunto son considerados indicadores de bienestar. Esto se realiza con la finalidad de abordar cuestiones diversas desde perspectivas distintas a las que generalmente se utilizan.

Esta nueva perspectiva de análisis, basada en el concepto de vulnerabilidad, traslada la atención principalmente hacia los grupos o entidades expuestas a cambios ambientales. A diferencia del planteamiento tradicional, el análisis de vulnerabilidad considera las diferentes presiones a las que puede verse sometida una comunidad, un municipio, un departamento o un país. En este marco se realiza el trabajo para el caso de Bolivia, con la finalidad de, por un lado, determinar una estructura formal matemática que describa el impacto que generan las amenazas y vulnerabilidades en el bienestar de las personas y, por otro, cuantificar el impacto.

2. Generalidades

Bolivia se encuentra situada en una zona de intensa actividad climática, por lo que cada año es amenazada por ondas tropicales, como tormentas, y disturbios, como heladas y sequías estacionales, que afectan a los asentamientos humanos y las actividades económicas de toda índole.

En los últimos cuarenta años, la mortalidad a causa de inundaciones representa el 45% del total de vidas perdidas como efecto de eventos adversos, correspondiendo el 30% a epidemias, el 16% a deslizamientos, el 8% a terremotos y el 1% a vientos huracanados. Si se describe a la población afectada por eventos climáticos adversos en estos mismos cuarenta años, un 69% del total corresponde a sequías. Los afectados por inundaciones corresponden al 28% y por deslizamientos al 3% (Quiroga *et al.*, 2010).

Las consecuencias de estos eventos se expresan en pérdidas económicas. Según la CEPAL, éstas ascienden a 965,6 millones de dólares a causa de sequías, 400 millones a deslizamientos y 804,6 millones a de inundaciones.

Entre los desastres naturales reportados que se han incrementado en los últimos cinco años están la inundación y la helada: en 2002 se llegó a 353 casos de inundación y en 2006 a 868; los casos de helada subieron de 66 a 121 en el mismo período. Los departamentos que sufrieron más en el período indicado fueron La Paz, con 404 casos de desastres naturales reportados; Beni, con 182; Potosí, con 163; Tarija, con 111; y Oruro, con 110 casos.

En relación con el número de familias damnificadas, éstas se incrementaron entre 2003 y 2006, por efecto de inundaciones, de 38.631 a 45.928; por heladas, de 2.402 a 7.851 familias; por granizadas, de 6.225 a 11.528 familias; por deslizamientos, de 426 a 714 familias. Los departamentos en los que mayor cantidad de familias sufrieron los efectos de los desastres naturales en 2006 fueron La Paz, con 16.849; Potosí, con 12.256; Beni, con 9.511; Santa Cruz, con 8.862 y Oruro, con 8.113 familias (INE, 2010).

Por otro lado, si se describen las vulnerabilidades, éstas se encuentran asociadas estructuralmente a los modelos de desarrollo vigentes, que generan desigualdades sociales, económicas y políticas, marginalidad, inequidad y exclusión social. Las condiciones de vulnerabilidad se relacionan a largo plazo con la migración acelerada, el deterioro ambiental, las condiciones de pobreza y la debilidad institucional.

Cabe mencionar que Bolivia mejoró en la agenda de la gestión del riesgo mediante la aplicación de estrategias de intervención y la formulación de instrumentos normativos que coadyuvan en el fortalecimiento de las instancias de decisión política (CONARADE) y técnica (SINAGER). Sin embargo, las entidades relacionadas con la temática tienen dificultades para desarrollar tareas más amplias e integrales que contemplen todas las etapas del ciclo de la gestión del riesgo, debido, entre otros, a factores de manejo de información, aplicación

de investigaciones, diseño de herramientas e instrumentos y procesos de planificación relacionados con la reducción del riesgo.

Por lo tanto, la problemática general del escenario de riesgo en Bolivia, desde una visión integral que asocia la identificación de las amenazas, la evaluación de las vulnerabilidades y la definición de líneas de acción estratégicas de reducción del riesgo, se relaciona con las condiciones socioeconómicas.

3. Metodología

Para investigar el impacto de las amenazas y vulnerabilidades en el bienestar se plantean los siguientes pasos:

- Acopio de información referente a amenazas, vulnerabilidades y variables que pueden mostrar el bienestar de los 347 municipios de Bolivia.
- Segmentación de las variables en altiplano, valles y llanos, para un mejor análisis.
- Utilización del algoritmo C4.5 para construir los sistemas expertos.
- Análisis de significancia y diagnóstico del sistema experto mediante la matriz de confusión.
- Finalmente se efectúa el análisis e interpretación de los resultados simulando las respuestas mediante el software generado.

4. Aspectos teóricos sobre inteligencia artificial y los modelos C4.5

4.1. Reconocimiento de patrones

Se entiende por reconocimiento de patrones a la ciencia que se integra como un componente del área de la Inteligencia Artificial que se ocupa, en este caso en particular, de describir procesos y/o fenómenos económicos y su relación con el medio ambiente, utilizando dispositivos computacionales. Estos dispositivos computacionales procesan la información que finalmente permite establecer propiedades y/o vínculos entre conjuntos de variables con las cuales se intenta explicar el fenómeno. Los objetos abstractos, materia prima para el reconocimiento de patrones, son n-uplos de un cierto producto cartesiano de

conjuntos de cualquier naturaleza -como la Teoría Clásica de Conjuntos, la Teoría Difusa, la Teoría Rugosa o cualquier teoría de conjuntos que en un futuro se pueda crear.

En general, el reconocimiento de patrones (RP) es una ciencia interdisciplinaria cuyas fuentes integrantes son la matemática y la ingeniería de sistemas. El terreno de aplicación y trabajo del RP es amplio y variado, tanto desde el punto de vista de las investigaciones llamadas fundamentales, teóricas o básicas, como de las aplicaciones en diferentes áreas del conocimiento.

Ingresando al reconocimiento de patrones de manera formal, se describe inicialmente un objeto O , el mismo que es un n -uplo $I(0) = (x_1(0), \dots, x_n(0))$ donde $x_i(0) \in M_i$ con $i = 1, \dots, n$. Sobre M_i no se asume estructura algebraica, topología o lógica alguna.

A partir de un conjunto de descriptores de m objetos, que se denominan testores, se describe cualquiera de los patrones inmersos. Es decir, si se tiene una matriz MA de una muestra de entrenamiento de la matriz $M \subseteq U$, los descriptores son representados por tantas columnas como variables y de tantas filas como registros, y estas filas agrupadas en r clases. Si el conjunto completo de las variables R es tal que las descripciones de objetos de la muestra M (filas de MA) que se encuentran en clases diferentes no se confunden, decimos que R es un testor.

De manera consecuente, se dice que el conjunto $\tau = \{i_1, \dots, i_s\}$ de columnas de la tabla T (y sus respectivos rasgos $(x_{i_1}, \dots, x_{i_s})$) se denomina testor para $(T_0, T_1) = T$, si, después de eliminar de T todas las columnas excepto las de τ , no existe fila alguna en T_0 igual a una de T_1 . (Abadia, 2010).

4.2. El proceso de extracción del conocimiento

El descubrimiento de conocimiento en bases de datos (en inglés *knowledge Discovery from Databases*, KDD) es un proceso iterativo e interactivo. Es iterativo porque la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de calidad. Es interactivo porque el usuario, o más generalmente un experto en el dominio del problema, debe ayudar en la preparación de los datos y la validación del conocimiento extraído.

El proceso de KDD se organiza en cinco fases. En la fase de integración y recopilación de datos se determinan las fuentes de información que pueden ser útiles. A continuación,

se transforman todos los datos en una sistematización formal, frecuentemente mediante un almacén de datos que logre unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias. Estas situaciones se tratan en la fase de selección, limpieza y transformación, en la que se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. La selección incluye tanto una fusión horizontal como otra vertical. Las dos primeras fases se suelen complementar bajo el nombre de “preparación de datos”. En la fase de minería de datos se analiza cuál es la tarea a realizar y se elige el método que se va a utilizar (se define cuál es la finalidad del sistema experto y cuál de éstos se va a utilizar). En la fase de evaluación e interpretación se evalúan los patrones y se analizan por los expertos, y si es necesario se vuelve a las fases anteriores para una nueva iteración. Esto incluye resolver conflictos con el conocimiento que se disponía anteriormente. Finalmente, en la fase de difusión se muestra el trabajo del sistema experto.

4.3. Algoritmo de inducción de Sistemas Expertos por árboles de decisión C4.5

El algoritmo C4.5 permite construir, a partir de un conjunto de datos de entrenamiento, un sistema experto que tiene como estructura básica un árbol de decisión que representa la relación que existe entre la decisión y sus atributos o variables. Para construir el sistema experto se realizan estratificaciones binarias sucesivas en el espacio de las variables explicativas, de forma que para realizar cada partición se escoja la variable que aporta más información en función de una medida denominada entropía, o cantidad de información que posee.

El árbol de decisión se construye bajo las siguientes premisas:

- Cada nodo corresponde a un atributo y cada rama al valor posible de ese atributo. Una hoja del árbol especifica el valor esperado de la decisión de acuerdo con la base de datos utilizada. La explicación de una determinada decisión viene dada por la trayectoria desde la raíz a la hoja representativa de decisión.
- A cada nodo se le asocia aquel atributo más informativo que haya sido considerado en la trayectoria desde la raíz.
- Para medir el nivel informativo de un atributo se emplea el concepto de entropía. Cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será el atributo para la clasificación.

El algoritmo, como principio, utiliza el criterio denominado *gain* (ganancia) para elegir el atributo (variable) con base en el cual se realiza cada partición; estas particiones llegan a constituir el árbol.

Lo anterior, de manera formal matemática, se expresa de la siguiente manera: sea un conjunto aleatorio de elementos de un conjunto T denominado población y que pertenece a una clase C_j . La probabilidad del mensaje que nos indica la clase a la que pertenece el elemento es:

$$P = \frac{freq(C_j, T)}{|T|} \quad (1)$$

y la información que proporciona dicho mensaje es:

$$I = -\log_2\left(\frac{freq(C_j, T)}{|T|}\right) \text{ bits}, \quad (2)$$

donde: $freq(C_j, T)$ representa el número de casos u observaciones en el conjunto T que pertenecen a la clase C_j y $|T|$ denota el número de casos u observaciones que contiene el conjunto T . Por tanto, tomando la esperanza matemática de la cantidad $-\log_2\left(\frac{freq(C_j, T)}{|T|}\right)$, se tiene la cantidad media de información necesaria para identificar la clase, de entre k clases posibles, a la que pertenece un caso en el conjunto T (la denominada entropía del conjunto T). Esa cantidad media viene dada por la expresión:

$$\text{info}(T) = -\sum_{j=1}^k \frac{freq(C_j, T)}{|T|} \times \log_2\left(\frac{freq(C_j, T)}{|T|}\right) \quad (3)$$

Si se conoce el valor que toma un determinado atributo X para cada elemento del conjunto T , entonces para clasificar cada elemento se requiere una cantidad de información menor que $\text{Info}(T)$; esta cantidad se puede expresar como:

$$\text{info}_X(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times \text{info}(T_i) \quad (4)$$

donde T_i es cada una de las particiones hechas en el conjunto T , de acuerdo con los distintos valores que tome el atributo X , y $|T_i|$ es el número de observaciones que contiene cada una de dichas particiones. La magnitud:

$$\text{gain}(X) = \text{info}(T) - \text{info}_X(T) \quad (5)$$

mide la cantidad de información que se gana dividiendo el conjunto de datos T de acuerdo con el atributo X . Entonces, el criterio de ganancia selecciona para hacer la partición aquel atributo para el cual se maximiza la ganancia de información.

Para evitar los sesgos que favorezcan a los atributos con muchos valores posibles, se plantea el siguiente estimador robusto:

$$split\ info(X) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2 \left(\frac{|T_i|}{|T|} \right) \quad (6)$$

que representa la entropía del conjunto T cuando es dividido de acuerdo con los valores que toma el atributo X . Esta entropía será tanto mayor cuanto más elevado sea el número de dichos valores. De este modo, puede ser utilizado como divisor de $gain(X)$ para corregir los elevados valores que esta magnitud tiene para aquellos atributos que adopten un mayor número de valores posibles. Entonces, el atributo elegido para la partición es aquél para el cual el ratio de ganancia sea mayor, definiéndose esta medida como:

$$gai\ ratio(X) = \frac{gain(X)}{split\ info(X)} \quad (7)$$

El árbol construido por aplicación reiterada del criterio mostrado consta del mínimo número de atributos (variables) que se requieren para la clasificación eficiente (Muguerza, 2006)

5. Análisis de resultados

Para construir el sistema experto mediante el algoritmo C4.5, se utiliza el concepto de entropía de la información. Los datos de entrenamiento son un sistema $S = s_1, s_2, \dots$ de muestras ya clasificadas. Cada muestra s_i es un vector conformado por $s_i = x_1, x_2, \dots$, donde $x_1, x_2, \dots$ representan las cualidades o las características de la muestra. Los datos del entrenamiento se aumentan con un vector $C = c_1, c_2, \dots$, donde $c_1, c_2, \dots$ representan la clase de cada muestra que pertenece a la base de datos.

Cuadro 1
Variables objeto de estudio

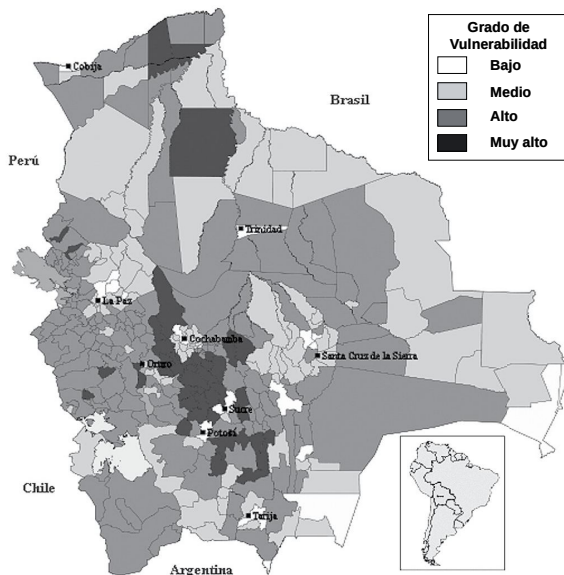
Código	Variable
A1	Número
A2	Departamento
A3	Municipio
A4	Población total del municipio
A5	Población infantil
A6	Población adulto mayor
A7	Tasa de crecimiento
A8	Analfabetos menores de 15 años
A9	Cobertura en educación
A10	Escolaridad menores de 19 años
A11	Terminó 8vo primaria
A12	Tasa de mortalidad infantil de menores de 1 año
A13	Déficit de personal calificado en salud por cada 1000 habitantes
A14	Mujeres embarazadas con cuarto control prenatal
A15	Incidencia de pobreza extrema
A16	Consumo promedio
A17	Razon de dependencia (poblacion entre 14 y 64 años)
A18	Calidad de construcción
A19	Habitabilidad de la vivienda
A20	Sin cañería
A21	Sin baño
A22	Con luz
A23	Sin luz
A24	Superficie (has)
A25	Población (CNPV 2001)
A26	Vulnerabilidad global
A27	Multi amenaza
A28	Riesgo total
A29	Grado de inundación
A30	Grado de sequía
A31	Grado de helada
A32	Riesgo de inundación
A33	Riesgo de sequía
A34	Riesgo de helada

Elaboración propia

C4.5 utiliza el hecho de que cada cualidad de los datos se puede utilizar para tomar una decisión que estructure los datos en subconjuntos más pequeños. Es por esta razón que se examina el aumento de información en la variable seleccionada (diferencia en entropía), lo cual resulta de elegir una cualidad para diferenciar el conjunto de datos. La cualidad con el aumento normalizado más alto de la información es la que se utiliza para tomar la decisión. El algoritmo entonces se repite en las sub-listas más pequeñas.

Se aclara que antes del proceso de datos, para un análisis por segmentos, se toman bases de datos separadas pertenecientes a la zona del altiplano (departamentos de La Paz, Oruro y Potosí), valles (departamentos de Cochabamba, Chuquisaca y Tarija) y llanos (departamentos de Santa Cruz, Beni y Pando).

Gráfico 1: Mapa de grados de vulnerabilidad



Fuente y elaboración: Atlas-OXFAM (2008)

Para mostrar el análisis de la presente investigación, primeramente se muestra el mapa de vulnerabilidades socio-económicas (población, educación, salud, economía, vivienda y servicios básicos), luego la estructura del sistema experto y los resultados estimados, y finalmente las interpretaciones.

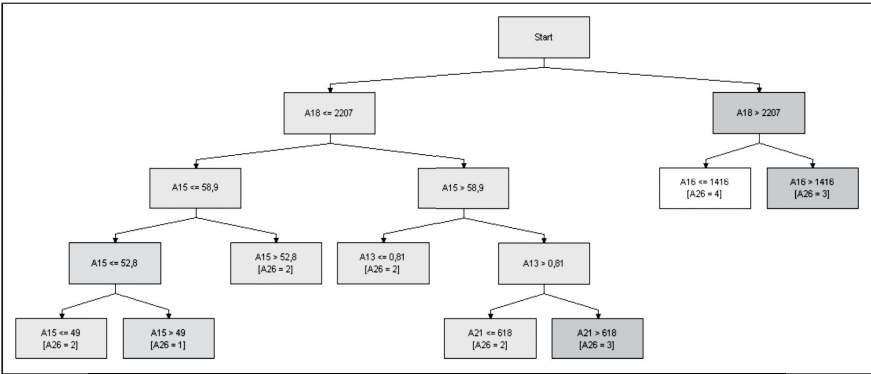
En este orden de ideas, la descripción de la vulnerabilidad económica permite ver que el grado de vulnerabilidad socioeconómica bajo abarca el 7% del total de los municipios de Bolivia, subiendo a un 28.7% los de grado medio. Más de la mitad de los municipios tiene un grado de vulnerabilidad socioeconómica alto (50.8%) y están distribuidos en todos los departamentos, llegando en el caso del departamento de Oruro al 71.4% de sus municipios. Los municipios de grado muy alto, que representan el 13.5%, se ubican en mayor proporción en la zona norte del departamento de Potosí y hacia el oeste del departamento de Cochabamba, representando un 39.5% y 28.9% de los municipios de estos departamentos, respectivamente.

En el contexto del presente documento, los cambios que se consideran afectan al bienestar de la sociedad son dos: i) las vulnerabilidades y ii) las amenazas naturales. Las amenazas

que incurren en un comportamiento atípico o rompen el patrón son aquéllas que generan desastres. Se aclara que la separación que se realiza entre bienestar y vulnerabilidad es artificial, puesto que las vulnerabilidades se pueden incluir en las variables que explican el bienestar y las amenazas naturales (Que se encuentran dentro de la categoría de cambios ambientales) se pueden incluir como el escenario que excitó al sistema de bienestar como variables exógenas. Sin embargo, por su importancia en relación con el concepto de vulnerabilidad, los desastres se han considerado separadamente y en la categoría de cambios en el medio ambiente, donde se incluyen otros cambios que, sin alcanzar la categoría de desastre, afectan a los sistemas humano, económico y ecológico. Todo lo descrito establece por sus características un sistema complejo, el mismo que se analiza a continuación con base en la inteligencia artificial, de manera particular, aplicando los sistemas expertos (SE).

El sistema experto estimado para la zona del Altiplano describe la estructura de la vulnerabilidad¹, según se puede ver en el siguiente gráfico:

Gráfico 2: Sistema experto para vulnerabilidad (zona Altiplano)



Elaboración propia

¹ Vulnerabilidad, según la norma, está referida "... al factor interno del riesgo, de un sujeto, objeto o sistema expuesto a una amenaza, que corresponde a su disposición intrínseca a ser afectado". En consecuencia, la vulnerabilidad es un factor interno que afecta al sistema social (interrelación social y actividades), a los grupos humanos (sujetos) o a la infraestructura (objetos). Está relacionada con la disposición intrínseca de estar expuesto a una amenaza y en consecuencia ser propenso a un riesgo de desastre (Cárdenas, 2008).

La vulnerabilidad es resultado de la interacción de factores físicos, sociales, económicos, culturales y ambientales, que acrecientan o reducen la propensión o predisposición al impacto de las amenazas. Habitualmente, se considera como opuesto a vulnerabilidad la noción de "seguridad" o "capacidad"; es decir, la habilidad para proteger a la comunidad y restablecer los medios de vida.

El diagrama del sistema experto muestra que las variables significativas para explicar la vulnerabilidad, explicitando la codificación, son: el déficit de personal calificado en salud por cada 1000 habitantes, la incidencia de la pobreza extrema, el consumo promedio, la calidad de la construcción y la ausencia de servicio sanitario.

Como se puede observar, en el árbol aparecen únicamente cinco variables de las 33 disponibles, lo que indica que 28 de las variables empleadas no aportan información relevante para construir estructuras lógicas de decisión con respecto a la vulnerabilidad y evaluar el grado de la misma. Se debe entender que los valores enteros de 1 a 4 expresan la menor y mayor incidencia respectivamente, quedando como valores de transición los valores 2 y 3 de la variable vulnerabilidad.

El árbol nos proporciona el menor número de atributos necesarios para alcanzar el objetivo deseado, y se interpretara del modo siguiente:

- ♦ Si el número de casas con una aceptable calidad en su construcción es menor que 2207 y la incidencia de la pobreza se encuentra en el intervalo de $[52.8 \ 58.9]$, la vulnerabilidad es de 2.
- ♦ Si tomamos otra rama del árbol, ésta se puede interpretar de la siguiente manera: cuando la vulnerabilidad es de 3, es decir, próxima a la mayor, la pobreza extrema es mayor a 58.9, el déficit de personal calificado por cada 1000 habitantes es mayor al 0.81 y la cantidad de familias sin servicios sanitarios es menor a 618.

De esta manera se van interpretando cada una de las ramas y hojas del árbol de decisión que estructura el sistema experto. La confianza alcanzada por éste es del orden del 76%, lo que se muestra en la matriz de confusión. Cabe aclarar que la confianza es aceptable por tratarse de datos de corte transversal y porque se evalúan los pronósticos como acierto o error, sin dar cabida a una respuesta aproximada.

Cuadro 2
Matriz de confusión

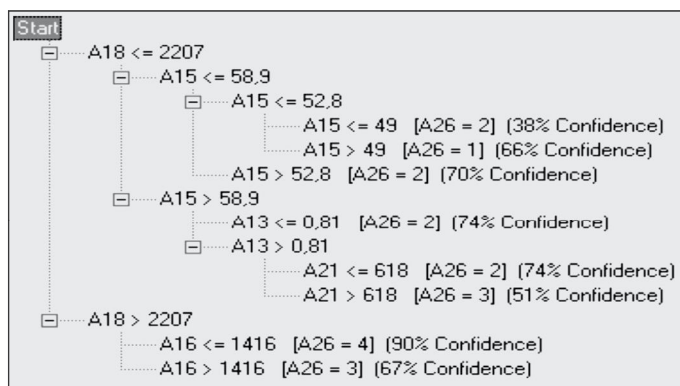
		Actual				Prediction Totals	Prediction Error %
		1	2	3	4		
PREDIC	1	6	0	1	0	7	14,29 %
	2	6	50	6	1	63	20,63 %
	3	1	13	40	13	13	34,43 %
	4	0	0	0	13	13	0,00 %
Actual Totals		13	63	47	21	144	24,31 %
Actual Error %		53,85 %	20,63 %	14,89 %	38,10 %	24,31 %	

Elaboración propia

La evaluación de este árbol de decisión construido con la muestra de entrenamiento (144 municipios) indica que el árbol consta de 8 ramas y comete un total de 35 errores (24,31%). También se muestra en la matriz de confusión el tipo de errores cometidos. Por último, para comprobar la capacidad predictiva del árbol, se clasifican de acuerdo con éste los 144 municipios de la muestra de validación, obteniendo un porcentaje de clasificaciones correctas del 72,2%.

El nivel de confianza para cada una de las ramas, es decir, el grado de seguridad o certeza con la que se realiza cada una de las inferencias realizadas, se muestra en el siguiente gráfico:

Con base en el sistema experto estructurado, se puede afirmar que las variables que poseen una mayor incidencia en la vulnerabilidad y finalmente en el bienestar son el consumo promedio, la pobreza extrema y posteriormente la población o número de habitantes, es decir que en poblaciones mayores a 2207 habitantes o más y con un consumo menor a los 1416 bolivianos se genera una vulnerabilidad de 4, la mayor vulnerabilidad en la escala.

Gráfico 3: Análisis de confianza del sistema experto

Elaboración propia

Esta y otras afirmaciones asociadas a la toma de decisiones se pueden simular mediante el sistema experto, cuyo interfaz de entrada de datos y salida de información se muestra en el siguiente gráfico.

Gráfico 4: Interfaz del sistema experto

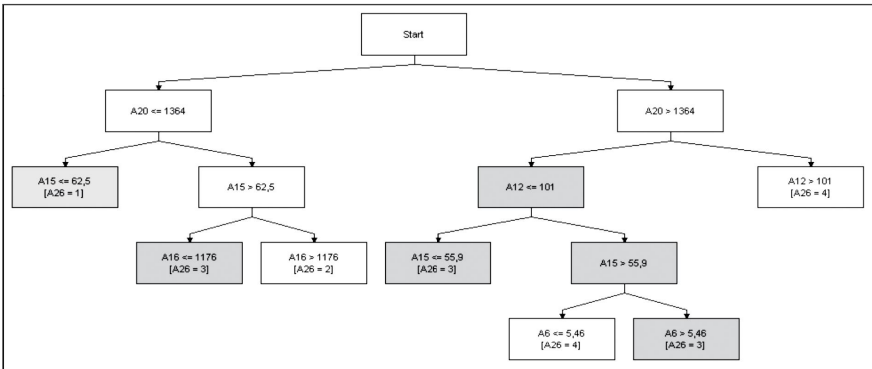
Prediction A26 is 3		Confidence Factor 67%
Field Name	Value	Range
A2	La Paz	
A4	70371	(221 - 793293)
A9	83	(0 - 142)
A12	76	(50 - 170)
A13	0,40	(0 - 4)
A15	61,9	(16 - 97)
A16	1896	(684 - 5220)
A18	6842	(22 - 20217)
A19	3009	(3 - 21299)
A21	13331	(77 - 60599)

Fuente: CorMac Technologies

Mediante el sistema se ingresan, por ejemplo, una reducción del 0.9 al 0.83 en tasas del déficit de personal calificado en salud por cada 1000 habitantes. Entonces se muestra en la parte superior si este cambio es significativo para incrementar o generar un decremento en el valor asociado a la vulnerabilidad. De esta manera se pueden determinar los mejores niveles de las variables socioeconómicas que minimicen la vulnerabilidad y maximicen el bienestar.

A continuación se muestran los sistemas expertos para valles y llanos referentes a la vulnerabilidad. De los anteriores sistemas expertos se puede mencionar que en la zona de los Valles las variables explicativas de la vulnerabilidad son la falta de alcantarillado, la incidencia de la pobreza extrema, la tasa de mortalidad infantil de menores de un año, el consumo *per cápita*, la falta de servicios sanitarios en los hogares y la población adulto mayor.

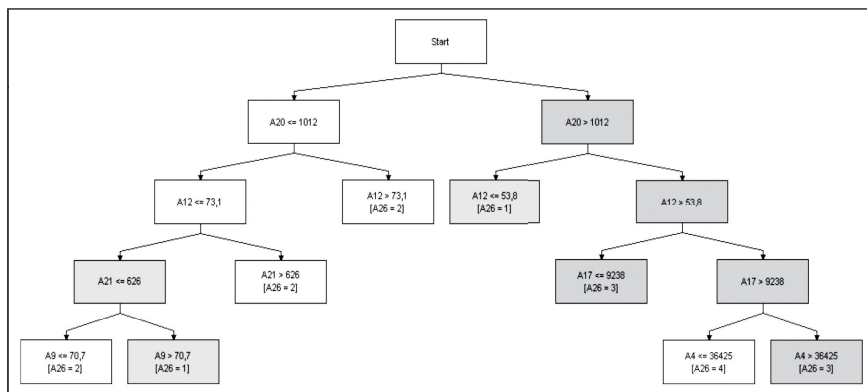
Gráfico 5: Sistema experto para vulnerabilidades en la zona de los valles



Elaboración propia

En la zona de los Llanos las variables significativas son: falta de alcantarillado, tasa de mortalidad infantil de menores de un año, falta de servicio sanitario en el hogar, razón de dependencia de la población entre los 14 y 64 años, cobertura de educación y población del municipio.

Gráfico 6: Sistema experto para vulnerabilidades en la zona de los llanos



Elaboración propia

Cabe entonces destacar que la variable más importante a nivel nacional es el de servicio sanitario en los hogares, siendo particular a los valles y llanos la falta de alcantarillado, la mortalidad infantil, las variables del consumo *per cápita* y la incidencia de la extrema pobreza en la Zona del Altiplano y los Valles. Esto se muestra en el siguiente cuadro.

Cuadro 3
Identificación de vulnerabilidades a nivel nacional

Variablen de vulnerabilidad	Zona del Altiplano	Zona de Los valles	Zona de los llanos
Falta de alcantarillado,		x	x
Tasa de mortalidad infantil menores de un año,		x	x
Falta de servicio sanitario en el hogar,	x	x	x
Razón de dependencia de la población entre los 14 y 64 años,			x
Cobertura de educación			x
Población del municipio.			x
Población adulto mayor.		x	
Consumo per cápita,	x	x	
Incidencia de pobreza extrema,	x	x	
Calidad de la construcción	x		
Déficit de personal calificado en salud por cada 1000 habitantes	x		

Elaboración propia

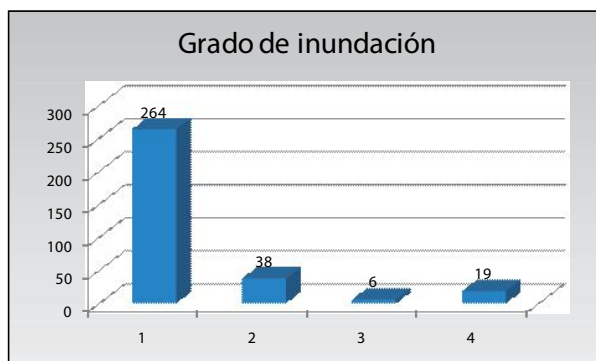
Las variables que agudizan la vulnerabilidad son: en la zona del Altiplano, la calidad de la construcción y el déficit de personal calificado en salud; en la zona de los valles, la población

adulto mayor; y en la zona de los llanos, la cobertura en educación, la razón de dependencia y la población en los municipios. Estas variables y su incidencia en la vulnerabilidad en territorio boliviano deben tomarse en cuenta en cada región, departamento o municipio para la presentación de políticas económicas y públicas con la finalidad de incrementar el bienestar económico y social.

Se construyen a continuación sistemas expertos para cada una de las amenazas climáticas² expresadas como el grado de inundación, grado de sequía y grado de helada, tomando en cuenta su incidencia en el consumo, la educación, la tasa de mortalidad y la habitabilidad de las viviendas, para las diferentes regiones de Bolivia.

En el siguiente gráfico se muestra que aproximadamente 264 municipios se encuentran en un grado de inundación 1, lo que expresa que el país tiene un bajo grado de inundación; empero, diecinueve municipios poseen un alto grado de amenaza, los mismos que se encuentran principalmente en el departamento del Beni.

Gráfico 7: Número de municipios por grado de amenaza de inundación



Elaboración propia

² Amenaza es definida como "... el factor externo del riesgo representado por la potencial ocurrencia de un suceso de origen natural o generado por la actividad humana que puede manifestarse en un lugar específico, con una intensidad y duración determinadas" La amenaza, un factor del riesgo, compromete la seguridad de las personas y su medio (asentamientos humanos, infraestructura y unidades productivas).

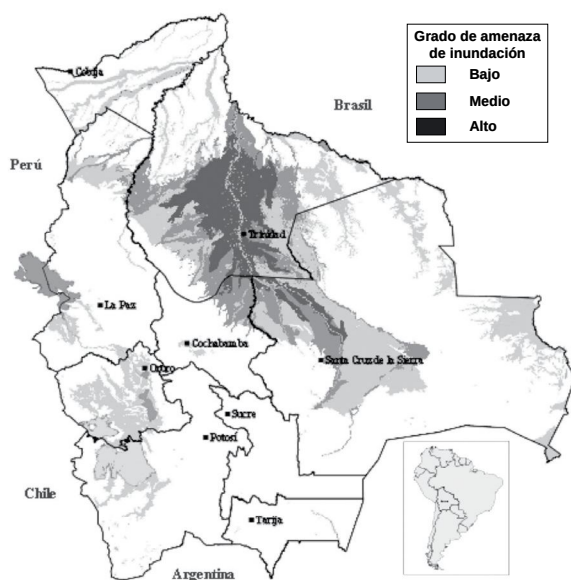
La amenaza es un factor físico externo (a la sociedad, a la comunidad, a la familia y sus interacciones sociales), y su ocurrencia es potencialmente peligrosa. Tiene dos fuentes principales: natural (fenómeno natural) o antrópica (actividad humana); si se considera que la ocurrencia de algunos fenómenos potencialmente peligrosos combina la actividad humana y los fenómenos naturales, debe incluirse la fuente socio-natural (Cárdenas, 2008).

Las amenazas naturales son generadas por las manifestaciones periódicas y circunstanciales de la naturaleza; las antropogénicas están asociadas a las acciones humanas; y en las siconaturales confluyen las prácticas humanas con el ambiente natural.

En el siguiente gráfico se aprecia que en las zonas del Altiplano y de los valles se tiene un grado bajo de amenaza de inundación, y que la zona del Oriente es la más afectada por este fenómeno.

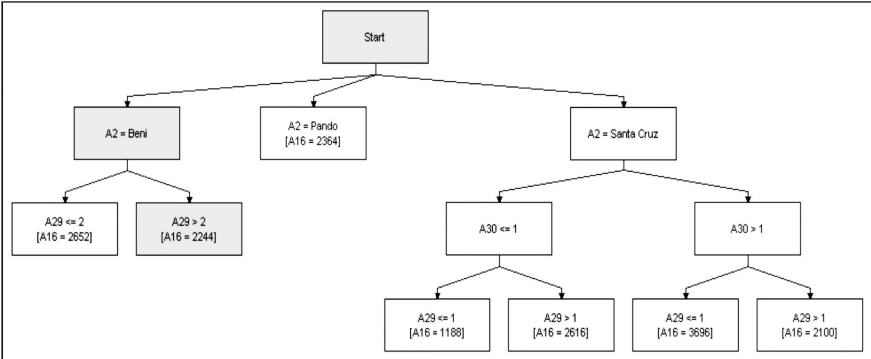
En lo que respecta a la zona de los llanos, se demuestra que existe una mayor incidencia de las inundaciones en los patrones de consumo, de manera particular en los departamentos de Santa Cruz y Beni. Por otro lado, en el departamento de Santa Cruz las sequías tienen un componente importante en las amenazas identificadas.

Gráfico 8: Mapa de amenaza de inundación



Fuente y elaboración: Atlas-OXFAM 2008

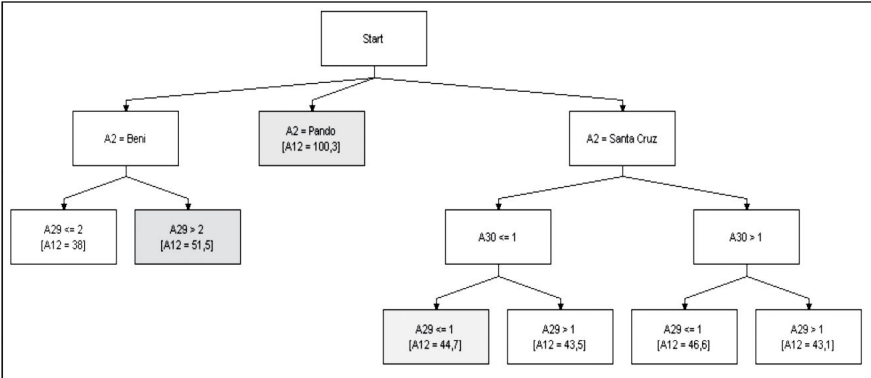
Gráfico 9: Sistema experto de amenazas y consumo (zona de los llanos)



Elaboración propia

Con respecto a la mortalidad en la zona de los llanos, se muestra que a niveles de 1 y 2 se tiene un nivel de 38, y a grados 3 y 4 en amenaza de inundación se tiene una mortalidad promedio de 51,5 por cada mil habitantes. Por otro lado, en Santa Cruz se tiene un rango en mortalidad de entre 43.1 y 46.6 en zonas que son afectadas por sequías o inundaciones.

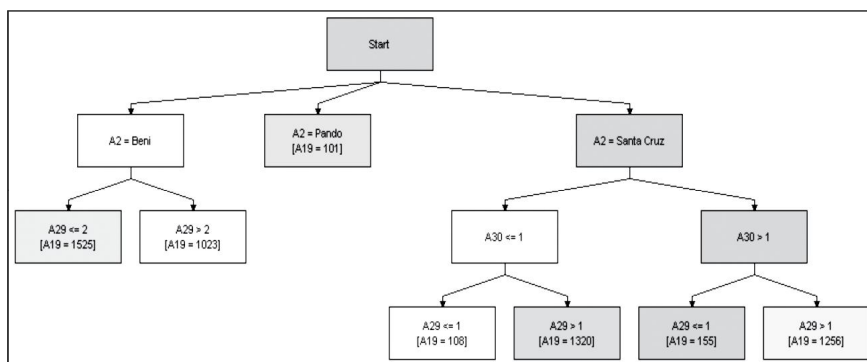
Gráfico 10: Sistema experto de amenazas y tasa de mortalidad (zona de los llanos)



Elaboración propia

Con respecto a la habitabilidad de las viviendas, se muestra que en Beni, a mayor grado de amenaza se reduce la habitabilidad; en cambio, en Santa Cruz se muestra una relación inversa, debido a que el sistema experto toma en cuenta la variable sequía. Esto último se debe a que la ocurrencia de eventos adversos, como las inundaciones, provoca que las personas pierdan sus casas y tiendan a hacinarse.

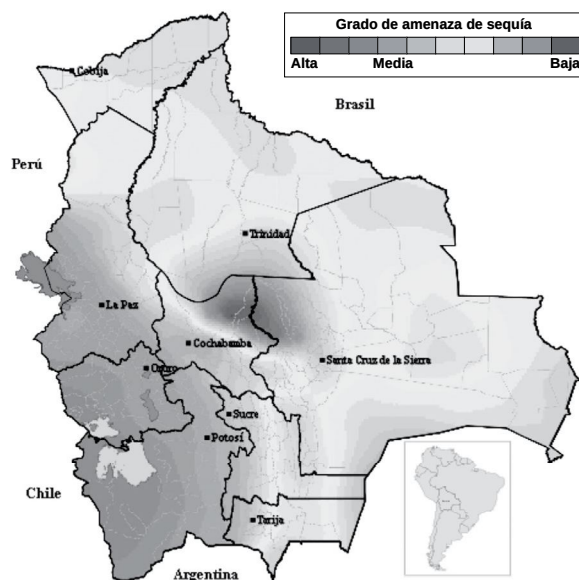
Gráfico 11: Amenazas y habitabilidad de la vivienda



Elaboración propia

En lo que respecta a la sequía, la región más afectada es la zona del Altiplano y parte de la zona de los valles, con una frecuencia expresada en 204 municipios con un grado alto de amenaza, como se muestra en el siguiente gráfico.

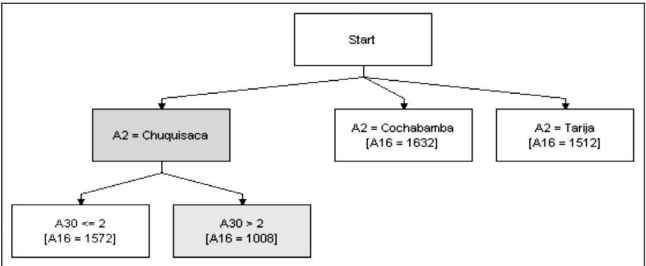
Gráfico 12: Mapa de grado de sequía



Fuente y elaboración: Atlas-OXFAM 2008

Siempre analizando con base en el sistema experto, en la zona de los valles se muestra una relación determinante entre las sequías y el consumo. A nivel departamental se aprecia que Chuquisaca posee las amenazas por inundaciones en relación inversa con el consumo promedio, mientras que en los departamentos de Cochabamba y Tarija no se reconocen patrones. El decremento de consumo *per cápita* de los municipios pertenecientes al departamento de Chuquisaca es de 564 bolivianos por unidad de amenaza de inundación.

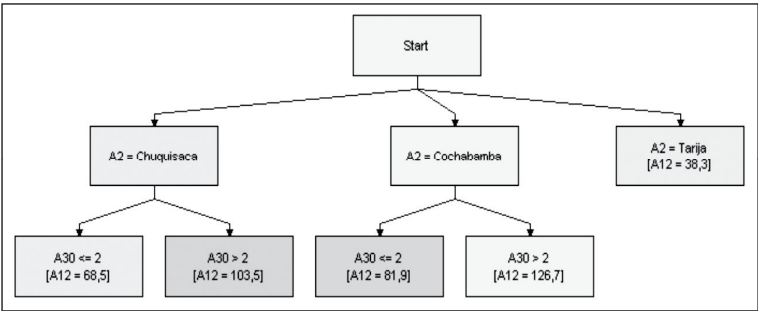
Gráfico 13: Sistema experto para consumo



Elaboración propia

La mortalidad infantil muestra un patrón directamente proporcional en Chuquisaca y en Cochabamba; empero, se tiene una mayor sensibilidad a la variación de amenaza en Cochabamba, la misma que es de 35 con respecto a 45 fallecidos promedio en Chuquisaca.

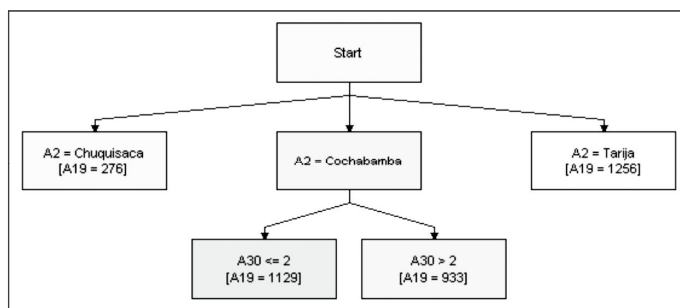
Gráfico 14: Sistema experto para tasa de mortalidad



Elaboración propia

Con respecto a la habitabilidad en la zona de los valles, se identifica un patrón importante perteneciente a Cochabamba. Éste muestra que, a una mayor amenaza de sequía, la habitabilidad en los municipios disminuye.

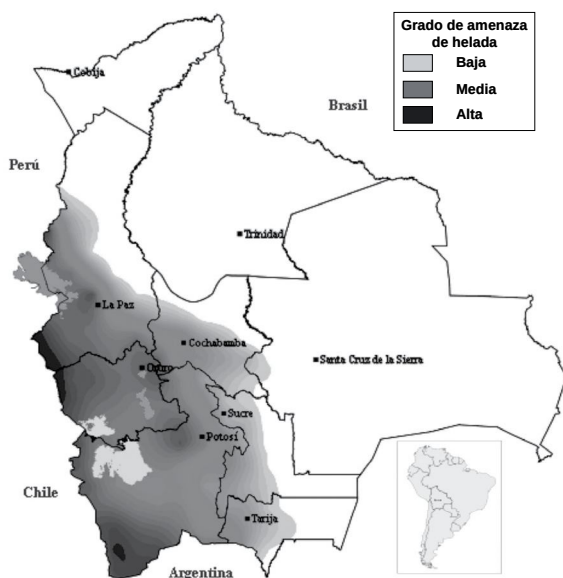
Gráfico 15: Sistema experto para habitabilidad de la vivienda



Elaboración propia

Las heladas, como se puede apreciar en el mapa, afectan en mayor medida a la zona del Altiplano, seguida de la zona de los valles (65 y 35%, respectivamente), lo que muestra un mayor número de municipios afectados en la zona del Altiplano.

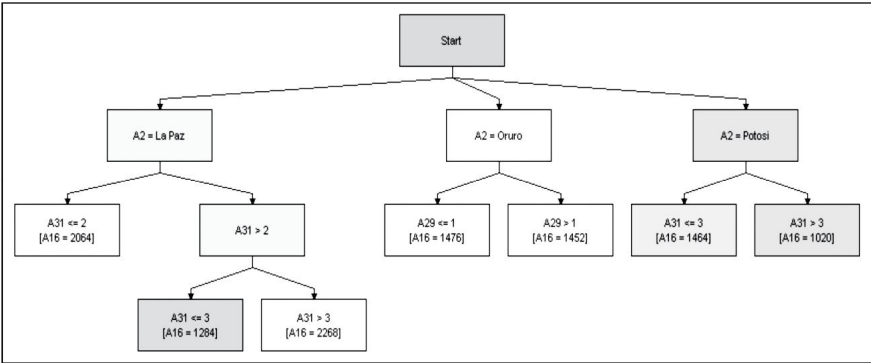
Gráfico 16: Mapa de grado de sequía



Fuente y elaboración: Atlas-OXFAM 2008

En la zona del Altiplano, con base en la información arrojada en el proceso de datos, se puede interpretar que la helada es una variable determinante para explicar el consumo de las personas. En el departamento de La Paz, si el grado de amenaza es menor o igual que 2, el consumo tiene un promedio de 2064 bolivianos, y si el grado de amenaza se incrementa, el consumo es superior. Este mismo patrón se repite en el departamento de Potosí. Se puede apreciar también que el mayor riesgo de helada se registra en Potosí, y que ligado a esto se encuentra un promedio de consumo inferior a los otros dos departamentos. Es decir que en el departamento de Potosí el consumo se ve determinado por las heladas, porque éstas afectan en alto grado la productividad agrícola de la región.

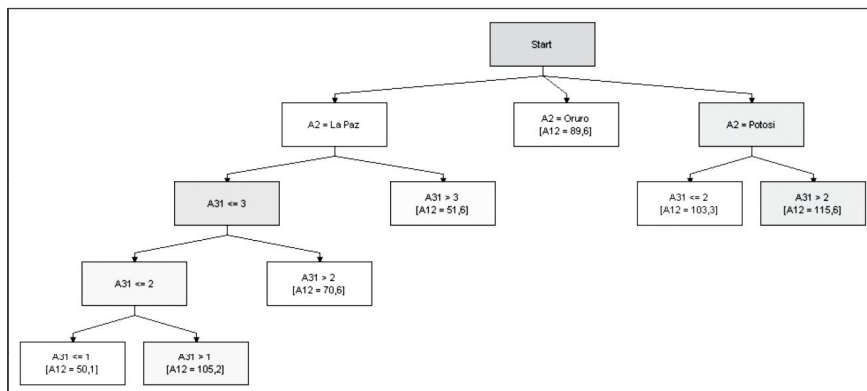
Gráfico 17: Sistema experto para cuantificar el impacto del grado de las amenazas en el consumo de las personas (zona del Altiplano)



Elaboración propia

Con respecto a la variable de mortalidad, ésta registra un máximo en el departamento de La Paz, el mismo que corresponde a un grado de amenaza 2. La variable mortalidad registra un valor máximo de 105, a lo cual se suma una variabilidad mayor tomando en cuenta un mínimo de 50. En Oruro no se registra un patrón, y en Potosí, a un mayor nivel de amenaza por helada, se incrementa el promedio de 103 a 115 fallecidos por cada mil habitantes. Estos resultados se logran cuando tiene lugar un evento adverso.

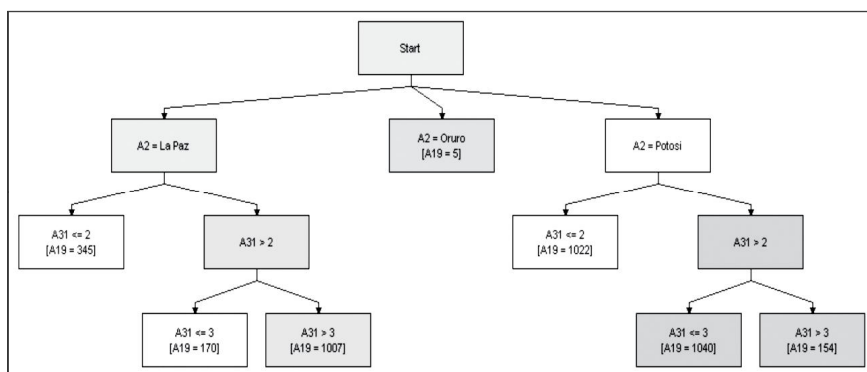
Gráfico 18: Sistema experto para cuantificar el impacto del grado de las amenazas en la mortalidad (zona del Altiplano)



Elaboración propia

La habitabilidad se define como la relación porcentual de las viviendas cuya calidad de habitabilidad es baja respecto del total de viviendas. La calidad de habitabilidad baja de las viviendas, dados los resultados obtenidos, es determinada por tres factores: hacinamiento por dormitorio (más de tres habitantes por cada dormitorio con que cuenta la vivienda), hacinamiento por habitación (más de seis personas por habitación) y carencia de una dependencia exclusiva para cocinar.

Gráfico 19: Sistema experto para cuantificar el impacto del grado de las amenazas en la habitabilidad (zona del Altiplano)



Elaboración propia

Según el sistema experto, el promedio de habitabilidad muestra un mínimo cuando se tiene un grado de amenaza 3, se incrementa a 1007 si la amenaza es 4, y se mantiene en 345 si la amenaza toma los valores de 1 y 2. Por otro lado, en Potosí, a mayor amenaza se reduce el número de viviendas con calidad de habitabilidad baja.

6. Conclusiones

Como conclusión inicial se debe mencionar que el presente trabajo muestra información que coadyuva al fortalecimiento de las capacidades de preparación y respuesta de las instituciones nacionales, regionales y locales de Bolivia mediante la gestión de metodologías de recopilación, sistematización, procesamiento, análisis de información y apoyo a la toma de decisiones relacionada a las amenazas prevalentes y las vulnerabilidades crecientes con respecto al logro de incremento del bienestar. Muestra que existen relaciones significativas entre las amenazas y algunas variables del bienestar, por lo que permite cuantificar en qué rangos se encuentra la variable de bienestar en función de los indicadores de amenazas o vulnerabilidades.

Con base en las anteriores conclusiones se puede afirmar que:

- Se asiste a la toma de decisiones y la planificación en el contexto político-administrativo, físico-natural, sociocultural, institucional y económico-productivo de Bolivia.
- El enfoque teórico constituye el pilar fundamental de la elaboración del documento en el cual se realiza un análisis la realidad.
- Los criterios metodológicos, técnicos y operativos que orientan la modelación y el análisis del escenario de riesgo y bienestar ayudan al desarrollo de sistemas expertos. Éstos nos permiten simular y analizar de mejor manera las posibles soluciones y/o decisiones que se puedan proponer.

Con respecto a la gestión del riesgo, se concluye que el presente trabajo apoya al proceso de adopción de políticas, estrategias y prácticas orientadas a reducir el riesgo o minimizar sus efectos. Esto implica intervenciones en los procesos de planificación para el desarrollo y la implementación de estrategias orientadas a reducir las causas que generan condiciones de vulnerabilidad en las unidades sociales y sus medios de vida.

Se demuestra básicamente que existe una relación entre vulnerabilidades y amenazas ambientales, y bienestar, lo que se describe al interior del documento.

Desde la perspectiva técnica se concluye que:

- El tipo de sistemas expertos utilizados para el modelado supera a las técnicas tradicionales, en el sentido de que tiene un carácter estrictamente no paramétrico, lo que es importante teniendo en cuenta que la información que se maneja en la construcción realizada pertenece a variables cualitativas y cuantitativas.
- Los sistemas expertos planteados difieren de las técnicas tradicionales en que no se requiere de la intervención de un experto humano para la inferencia de las reglas clasificadoras. Como resultado, éstas son mucho más objetivas porque tienen soporte en datos reales contenidos en una base de datos.
- Frente a las redes neuronales, el sistema experto posee la ventaja de que no es un sistema conformado por una caja negra, lo que permite valorar la importancia relativa de cada una de las variables explicativas, mostrando la estructura interior. Esta última muestra la forma lógica de obtener los patrones inmersos dentro de los datos.

Artículo recibido en: marzo de 2011

Manejado por: ABCE

Aceptado en: agosto de 2011

Referencias

1. Abadía, José. (2010). Sistema experto para la bolsa de valores. Universidad del Valle Colombia.
2. Albuquerque de Castro, Rafael. (2007). El estado de bienestar. Cambio de paradigmas. Los derechos sociales. Vicepresidencia de la República Dominicana Bogotá.
3. Cárdenas, Miguel; Choquevillca, P.; Saavedra, J.P.; Torrico, G. y Espinoza, J. (2008). Construcción de mapas de riesgo. Comisión Europea, Fundepco, Oxfam, Bolivia.
4. Cardona, Omar. (2003). Indicadores para la gestión de riesgos. Banco Interamericano de Desarrollo-Universidad Nacional de Colombia. Colombia.
5. Castillo, Enrique; Gutiérrez, J.M. y Hadi, A. (2000). Sistemas expertos y modelos de redes probabilísticas
6. Chavarro, Andrés. (2002). Economía ambiental y economía ecológica. *Journal Ideas Ambientales*, Nº 2.
7. Cohen, William. (2005). Fast Effective Rule Induction. AT&T Bell Laboratories.
8. Duarte, Tito. (2007). Aproximación a la teoría del bienestar. Universidad Tecnológica de Pereira, Colombia.
9. FAN–Bolivia. (2009). Implementation and Validation of a Regional Climate Model for Bolivia, Bolivia
10. García, Andrea. (2008). Amenazas, riesgos, vulnerabilidades y adaptación frente al cambio climático. Naciones Unidas-Universidad Nacional de Colombia. Colombia.
11. Gómez, Javier. (2001). Vulnerabilidad y medio ambiente. Naciones Unidas–CELADE. Chile.
12. INDH. (2004). Índice de desarrollo humano en los municipios. Informe Nacional de Desarrollo Humano, Bolivia.

13. Instituto Nacional de Estadística, INE. (2010). Anuario 2010.
14. Menger, Carl. (2001). Economía y bienestar económico. Barcelona: Ediciones Orbis.
15. Muguerza, Javier. (2006). Construcción de un árbol de clasificación basado en múltiples submuestras sin renunciar a la explicación. Tesis doctoral, Universidad del País Vasco, Donostia.
16. Quintana, Michel. (2004). Modelos híbridos para los procesos de Data Mining en el apoyo a la toma de decisiones basados en tecnologías inteligentes conexionistas. Universidad Nacional de San Agustín, Perú.
17. Quiroga, Roger; Torrico, G.; Salamanca, L. A.; Quiroga, R. y Espinoza, J. C.(2010). Atlas de amenazas, vulnerabilidades y riesgos de Bolivia. Oxfam, Fundepco, VIDEICODI, Bolivia.
18. Saulnier, Galán. (2000). Aplicaciones de inteligencia artificial y sistemas expertos a la arqueología del conocimiento. Aplicación a la economía. Universidad Autónoma de Madrid, España.
19. Trincado, Estrella. (2008). Economía del desarrollo y economía del bienestar. Universidad Complutense de Madrid.
20. Zucchetti, Anna; Ramos, V.; Alegre, M.; Aguilar, Z.; Arroyo, R. y Tribut, E. (2008). Guía metodológica para el ordenamiento territorial y la gestión de riesgos. PNUD-UN-Habitat, Perú.

Anexo

Estimación de un sistema de ecuaciones para el análisis de las relaciones entre bienestar, vulnerabilidades y amenazas

En la presente investigación se presentan los modelos de ecuaciones estructurales, una técnica de análisis estadístico multivariante utilizada para contrastar modelos que proponen relaciones causales entre las variables que conforman el bienestar, las vulnerabilidades y las amenazas. A continuación se discute la estructura general que tiene un modelo, los tipos de variables que se pueden utilizar en ellos y su representación mediante diagramas estructurales. Construido y diagnosticado el sistema de ecuaciones del modelo y estimados los parámetros, se discute el concepto de causalidad, para entender su utilización en un contexto de política económica, social y finalmente la toma de decisiones.

El calentamiento global se expresa en el corto plazo en la variabilidad climática, la misma que se muestra en los cambios atmosféricos y comportamientos climáticos que las sociedades perciben en el cotidiano vivir.

El riesgo se va configurando a lo largo del tiempo, por una interacción entre los procesos que generan condiciones de vulnerabilidad acentuados por las amenazas naturales. Estas amenazas cambian por la existencia de variabilidad climática, alterando sus patrones de comportamiento. Al mismo tiempo, en el largo plazo la variabilidad climática es afectada por el calentamiento global. Entonces se puede inferir que el cambio climático, por un lado, aumenta las amenazas de origen meteorológico y climático y, por otro, incide negativamente en la resiliencia de muchos hogares y comunidades, las cuales ven afectado su bienestar por las pérdidas sufridas.

Por lo tanto, el cambio climático supone un factor global y muy importante de riesgo, que en realidad se genera gracias a una construcción social (asentamientos inadecuados, mala planificación urbana, etc.). Este cambio climático actuará como inyector de potencia para la relación entre riesgo y vulnerabilidad, aumentando de manera drástica el impacto de los desastres en las personas.

Para reducir los riesgos, los gobiernos, y en particular el de Bolivia, deben ser sensibles a las necesidades de las personas expuestas a los desastres naturales y además deben ser capaces de

tomar las decisiones de manera oportuna, equitativa y estratégicamente coherente en materia de toma de decisiones y planteamiento de políticas.

Es por las razones expresadas que el presente trabajo se orienta a identificar y cuantificar las variables correspondientes al bienestar de las personas, y a las vulnerabilidades y amenazas ambientales. De esta manera se busca cuantificar e interpretar las relaciones existentes entre las diferentes variables para apoyar a la toma de decisiones y estructuración de políticas.

Para lograr el objetivo planteado se hace uso de los modelos SEM (Structural Equation Modeling), también llamados modelos de “Análisis de la Estructura de Covarianzas”, los mismos que consisten en una técnica confirmatoria que trata principalmente de comprobar si un cierto modelo empírico es válido estadísticamente. El análisis factorial, la regresión y el “*pathanalysis*” pueden considerarse como casos particulares de esta técnica.

El primero de los conceptos relacionados con SEM que se aplica es lo que se conoce como variable latente, constructo o factor. Así, en la presente investigación se tienen por un lado variables medibles, observables y variables latentes o no observables. Además, según el papel que jueguen los factores o variables no observables en el modelo se tienen factores endógenos y exógenos.

Modelos de ecuaciones estructurales

Los Modelos de Ecuaciones Estructurales (en inglés SEM) pertenecen a la clase de los Modelos de la Estructura de Covarianza (CSM), que es una técnica general estadística que incluye un gran número de conocidas técnicas utilizadas en el análisis multivariado, como ser el análisis factorial exploratorio, el análisis factorial confirmatorio, el análisis de senderos “*path*” y la regresión múltiple, entre otras.

El análisis factorial, como parte del SEM, trata principalmente de encontrar factores explicativos. Es decir, con los campos que conforman una tabla de datos y las relaciones entre las tablas, se estructuran un conjunto de potenciales relaciones entre los campos, también denominados variables, para luego estadísticamente mostrar si existe un conjunto de relaciones significativas.

Existen dos enfoques para estudiar el análisis factorial: el enfoque exploratorio y el enfoque confirmatorio. El enfoque exploratorio no tiene restricciones en cuanto a normalidad,

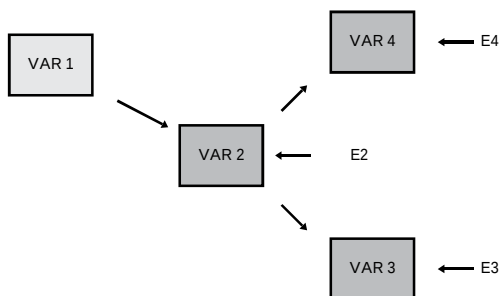
homocedasticidad o linealidad entre las variables; de hecho, algunas veces la multicolinealidad favorece los resultados, ya que se agrupa variables en función de su correlación. Para la estimación usa el análisis de componentes principales. La perspectiva confirmatoria, sin embargo, permite aplicar máxima verosimilitud y se propone como una manera óptima de validar el análisis factorial confirmatorio.

El análisis exploratorio no tiene en cuenta los errores de medida de las variables. Normalmente el análisis de componentes principales considera la varianza total y estima los factores que contienen proporciones bajas de varianza única. El análisis confirmatorio, a parte de las cargas factoriales, que también son una salida del exploratorio, nos provee los valores de la estimación de los errores de medida y la bondad de ajuste del modelo.

Dos son las principales ventajas de usar SEM. La primera de ellas es que permite tener en cuenta los errores de medida de las variables. Es decir, al trabajar con variables latentes, como es el caso de la sociología, SEM permite considerar que existe un error de medida en las variables observadas y tiene ese error en cuenta al hacer la estimación.

Otra de las grandes ventajas de SEM es que se puede explicar más de una variable a la vez, es decir, mientras que en regresión múltiple y regresión simple sólo teníamos una variable dependiente, con SEM podemos tener más de una variable dependiente y variables mediadoras. Esto significa que si VAR1 influye a VAR2 y VAR3, a su vez tiene un efecto sobre VAR3; SEM nos permite estimar ambos efectos a la vez, algo que no era posible con la regresión. Además VAR2 puede influir a su vez en VAR4 y estimar todos los parámetros del modelo de manera simultánea.

Gráfico 1: Estructura del "Pathanalysis"



El anterior gráfico muestra cómo se representan las relaciones entre las variables propuestas y los errores. El modelo general de ecuaciones estructurales debe ser composición de un modelo de medida y de un modelo estructural.

De manera formal, el modelo estructural o “*innermodel*” se presenta como sigue:

$$\eta = B\eta + \Gamma\xi + \varsigma$$

donde:

η representa al vector de variables aleatorias latentes endógenas de dimensión $m \times 1$.

ξ representa al vector de variables aleatorias latentes exógenas de dimensión $n \times 1$.

B representa la matriz de coeficientes que rigen las relaciones entre las variables endógenas $m \times m$.

Γ representa la matriz de coeficientes que rigen las relaciones exógenas y cada una de las endógenas, o, dicho de otro modo, los efectos de ξ sobre η . Su dimensión es $m \times n$.

ς representa al vector de perturbaciones o errores.

En este modelo hemos definido como variable latente exógena toda aquella que afecta a otra variable latente y que por el contrario no es afectada por alguna otra. Así, una variable latente endógena será aquella que sea afectada por otra independiente que afecte o no a otra.

El modelo de medida o “*outermodel*” se rige por dos ecuaciones. Una que mide las relaciones entre las variables latentes endógenas y sus variables observables:

$$y = \Lambda_y \eta + \varepsilon$$

donde:

y es el vector de p variables observables ($p \times 1$)

Λ_y es la matriz de coeficientes que muestran las relaciones entre las variables latentes y las observadas ($p \times m$) (también llamada matriz de cargas).

ε es el vector de errores ($p \times 1$).

La segunda ecuación del modelo de medida es la que rige las relaciones entre las variables latentes exógenas y sus variables observables:

$$x = \Lambda_x \eta + \delta$$

donde:

x es el vector de p variables observables (qx1)

Λ_x es la matriz de coeficientes que muestran las relaciones entre las variables latentes y las observadas (qxm) (también llamada matriz de cargas).

δ es el vector de errores (qx1).

En general, podemos decir que la relación causal entre dos variables implica que ambas variables covarían, permaneciendo constantes el resto de las variables. Pero lo contrario no es cierto: la covariación entre dos variables no implica necesariamente que exista una relación causal entre ambas; la relación puede ser espuria, falsa, ficticia.

Modelos y las relaciones causales directa e indirecta

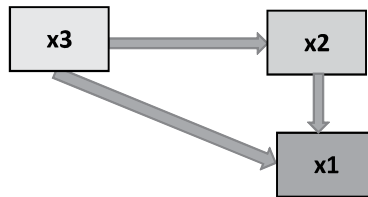
Hasta ahora sólo hemos mencionado relaciones causales directas. Una relación causal indirecta implica la presencia de tres variables. Existe una relación indirecta entre dos variables cuando una tercera variable modula o mediatiza el efecto entre ambas. Es decir, cuando el efecto entre la primera y la segunda pasa a través de la tercera. A las variables que median en una relación indirecta se las denomina también variables moduladoras.

Consideremos la relación entre la x_1 , x_2 y la x_3 . Si teóricamente se define que x_1 está en función de x_2 y esta última de x_3 , la relación puede representarse gráficamente como:



El modelo del gráfico propone que existe un efecto directo de la variable x_3 sobre la x_2 y de la x_2 sobre la x_1 . Además, existe un efecto indirecto entre la x_3 y la x_1 . El efecto indirecto de la variable x_3 sobre la x_1 puede ser potenciado (o atenuado) por la variable moduladora x_2 .

La existencia de un efecto indirecto entre dos variables no anula la posibilidad de que también exista un efecto directo entre ellas. Así, las relaciones propuestas en el siguiente gráfico pueden hacerse más complejas.



Relación causal recíproca

La relación causal entre dos variables puede ser recíproca o unidireccional. Cuando la relación es recíproca (bidireccional), la variable causa es a su vez efecto de la otra. Este tipo de relaciones se representa como dos flechas separadas orientadas en sentidos contrarios. Una relación recíproca es en definitiva un bucle de retroalimentación entre dos variables. La relación causal recíproca puede ser directa o indirecta, implicando a otras variables antes de cerrarse el bucle.

La relación entre la x_1 y la x_2 puede representarse como un bucle recíproco: cuanto mayor es x_1 , menor o mayor es x_2 ; o cuanto menor es x_1 , menor o mayor es x_2 .



Efectos totales

El efecto total se refiere a los efectos no analizados. En la representación gráfica son las flechas que podrían estar representadas y que no lo están. Estas ausencias pueden obedecer a dos motivos: primero, puede ocurrir que se hayan dejado fuera del modelo variables importantes para explicar la covariación presente en los datos (error de especificación); y segundo, puede ser que se haya asumido que el resto de las variables no consideradas en el modelo se compensan entre sí, incorporándose su efecto en los términos de error del modelo. A la suma de los efectos espurios más los efectos no analizados se denomina efectos no causales.

Marco metodológico

La estimación de un modelo comienza con la formulación de la teoría que lo sustenta. Dicha teoría debe estar formulada de manera que se pueda poner a prueba con datos reales. En concreto, debe contener las variables que se consideran importantes y que deben medirse a los sujetos.

El modelo teórico debe especificar las relaciones que se espera encontrar entre las variables (correlaciones, efectos directos, efectos indirectos, bucles). Si una variable no es directamente observable, deben mencionarse los indicadores que permiten medirla. Lo normal es formular el modelo en formato gráfico; a partir de ahí es fácil identificar las ecuaciones y los parámetros.

Una vez formulado el modelo, cada parámetro debe estar correctamente identificado y ser derivable de la información contenida en la matriz de varianzas-covarianzas. Existen estrategias para conseguir que todos los parámetros estén identificados, como por ejemplo, utilizar al menos tres indicadores por variable latente e igualar la métrica de cada variable latente con uno de sus indicadores (esto se consigue fijando arbitrariamente al valor 1 el peso de uno de los indicadores). Aun así, puede suceder que el modelo no esté completamente identificado, lo que querrá decir que se está intentando estimar más parámetros que el número de piezas de información contenidas en la matriz de varianzas-covarianzas. En ese caso habrá que imponer más restricciones al modelo (fijando el valor de algún parámetro) y volver a formularlo.

Por otra parte, una vez seleccionadas las variables que formarán parte del modelo, hay que decidir cómo se medirán las variables observables. Estas mediciones (generalmente obtenidas mediante escalas o cuestionarios) permitirán obtener las varianzas y las covarianzas en las que se basa la estimación de los parámetros de un modelo correctamente formulado e identificado (asumimos que estamos trabajando con una muestra representativa de la población que se desea estudiar y de tamaño suficientemente grande).

Una vez estimados los parámetros del modelo se procede, en primer lugar, a valorar su ajuste. Si las estimaciones obtenidas no reproducen correctamente los datos observados, habrá que rechazar el modelo y con ello la teoría que lo soportaba, pudiendo pasar a corregir el modelo haciendo supuestos teóricos adicionales.

En segundo lugar se procede a hacer una valoración técnica de los valores estimados para los parámetros: su magnitud debe ser la adecuada, los efectos deben ser significativamente distintos de cero, no deben obtenerse estimaciones impropias (como varianzas negativas), etc.

Puede ocurrir que alguna de las estimaciones tenga un valor próximo a cero; cuando ocurre esto es recomendable simplificar el modelo eliminando el correspondiente efecto. Por último, el modelo debe interpretarse en todas sus partes. Si el modelo ha sido aceptado como una buena explicación de los datos, será interesante validarlo con otras muestras y, muy posiblemente, utilizarlo como explicación de teorías de mayor complejidad que se desee contrastar. El proceso expuesto se resume en el gráfico 6.

Tipos de relaciones

En las técnicas multivariantes estamos acostumbrados a estudiar la relación simultánea de diversas variables entre sí. En estas técnicas las relaciones entre variables dependientes e independientes son todas del mismo nivel o del mismo tipo. En un modelo de ecuaciones estructurales podemos distinguir distintos tipos de relaciones. Entender estos distintos tipos de relaciones puede ser de gran ayuda a la hora de formular los modelos a partir de las verbalizaciones en lenguaje común. A continuación vamos a discutir estos tipos de relaciones.

Covariación vs causalidad

Decimos que dos fenómenos covarían, o que están correlacionados, cuando al observar una mayor cantidad de uno de los fenómenos también se observa una mayor cantidad del otro (o menor, si la relación es negativa). De igual forma, a niveles bajos del primer fenómeno se asocian niveles bajos del segundo. Así, por ejemplo, cuando decimos que la aptitud y el rendimiento correlacionan entre sí, esperamos que los sujetos con un mayor nivel de aptitud manifiesten un mejor rendimiento, y viceversa. Sin embargo, ya hemos enfatizado que covariación y causalidad no son la misma cosa. Cuando se observa una alta relación (covariación) entre dos variables, no debemos interpretarla como una relación causal entre ambas. Pueden existir otras variables que no hemos observado y que potencien o atenúen esta relación. Por ejemplo, es posible que la motivación y el rendimiento estén relacionados y que esa relación esté condicionando la relación de la aptitud con el rendimiento (potenciándola o atenuándola). Un ejemplo tal vez más claro es el propuesto por Saris. Si recolectamos datos sobre el número de vehículos y el número de aparatos telefónicos en distintas poblaciones, es seguro que encontraremos una covariación entre ambas variables. Pero no por ello pensamos que un mayor número de vehículos es el causante de que haya un mayor número de aparatos telefónicos.

Otro nivel de análisis es la causalidad. Si recogemos información sobre el número de fumadores en una habitación y la cantidad de humo existente en la habitación, observaremos que existe una alta covariación entre ambas variables. Parece razonable dar un paso más en la interpretación de este resultado y argumentar, conceptualmente, que la cantidad de fumadores causa la cantidad de humo y que los cambios en la cantidad de fumadores causarán un cambio en la cantidad de humo.

El cambio de perspectiva desde la covariación observada a la causalidad atribuida a dos variables lo lleva acabo el investigador, que es quien hipotetiza la causalidad. Es una buena costumbre que las verbalizaciones, o enunciados, sean explícitos respecto al tipo de relación que deseamos probar entre dos variables.

Análisis empírico

La utilidad de los modelos de ecuaciones estructurales radica en la aportación de una visión sistémica de los aspectos del fenómeno estudiado, en contraposición a otro tipo de herramientas estadísticas que se centran en el análisis individual de cada factor. Asimismo, reducen la cantidad de información que debe ser analizada, ya que su fundamento es agrupar las relaciones entre un gran número de variables en unos pocos factores, poniendo de relieve los aspectos esenciales del fenómeno. En el caso del estudio de constructos o variables no medibles directamente, estos modelos tienen la ventaja de carecer del error de medición, pero el inconveniente es que el investigador debe proceder a la explicación objetiva de relaciones causales entre variables que se caracterizan por su abstracción y/o subjetividad.

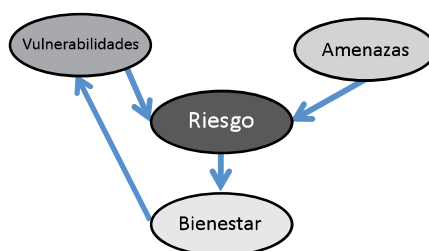
En lo que respecta al estudio de la causalidad, la función de los modelos de ecuaciones estructurales no es corroborar las relaciones causales entre las distintas variables, sino facilitar su análisis y toma de decisiones, para lo cual es necesario un análisis exploratorio de los datos y que el proceso de modelización sea seguido con rigor. En ocasiones la confirmación de un modelo de este tipo se ha considerado como una prueba de validez, sin tener en cuenta que podrían ser igualmente válidos otros modelos econométricos alternativos, puesto que las pruebas de significación sólo son efectivas cuando se cumplen las condiciones especificadas.

Modelo Estructural Subyacente del ECSI

Con el modelo estructural se quiere analizar las relaciones de causalidad, los impactos y efectos de las vulnerabilidades, las amenazas y las variables económicas para dar a conocer si estas con datos de corte transversal pueden mostrar patrones de comportamiento.

Definición de las variables del modelo y sus relaciones causales

La economía del bienestar estudia todo lo conducente a la formulación de proposiciones y juicios que permitan ordenar situaciones económicas alternativas, calificadas en términos de mejor o peor. El primer paso para elaborar un modelo estructural es la conceptualización de las variables latentes y sus relaciones:



El anterior gráfico muestra que se plantea una relación entre vulnerabilidades y amenazas con riesgo, el riesgo con el bienestar y finalmente el bienestar con las vulnerabilidades.

Operacionalización de las variables latentes

Los cuatro componentes del sistema son por tanto variables latentes, medidas cada una por dos o tres variables, en cada caso. El plan de operacionalización de las variables latentes en indicadores se realiza mediante la utilización de un análisis de cada una de las variables y sus componentes, primeramente teóricos, y luego por variables próxi.

Variables latentes

Vulnerabilidades:

Población
Salud
Educación
Vivienda

Amenazas:

Inundaciones
Sequías
Heladas
Incendios

Bienestar:

Índice de Desarrollo Humano
Índice de Pobreza Humana
Necesidades Básicas Insatisfechas

Riesgo:

Vulnerabilidades
Amenazas

Variables próximas que se utilizan para el análisis empírico

Vulnerabilidades:

- ♦ (A4) Población (población total por municipio)
- ♦ (A13) Salud (déficit de personal calificado en salud por cada 1000 habitantes)
- ♦ (A9) Educación (cobertura de educación)
- ♦ (A18) Vivienda (calidad de la construcción)

Amenazas:

- ♦ (A32) Inundaciones (riesgo de inundación)
- ♦ (A33) Sequías (riesgo de sequía)
- ♦ (A34) Heladas (riesgo de helada)

Bienestar:

Índice de Desarrollo Humano (IDH)

- i) (A8) Analfabetos menores a 15 años como próxi de tasa de alfabetización y tasa combinada de matriculados en educación primaria, secundaria y superior
- ii) (A16) Consumo *per cápita* como próxi de PIB *per cápita* por municipio.

Índice de Pobreza Humana (IPH)

- i) (A12) Tasa de mortalidad infantil de menores de 1 año, como próxi de la probabilidad de vida al nacer.
- ii) (A6) Población adulto mayor.
- iii) (A17) Razón de dependencia de la población entre los 14 y 64 años

Necesidades Básicas Insatisfechas (NBI)

- i) (A9) Cobertura de educación.
- ii) (A20) Sin cañería como próxi de agua potable
- iii) (A21) Sin baño como próxi de servicios higiénicos.
- iv) (A19) Habitabilidad de la vivienda

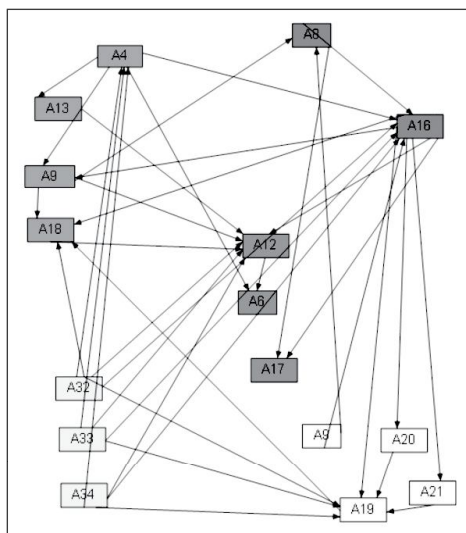
Riesgo:

Vulnerabilidades

Amenazas

Análisis por trayectorias

En el análisis por trayectorias o “Path Análisis” se construye una red en la que las variables se unen por flujos o flechas, las que muestran relaciones entre las variables como variables endógenas o exógenas.



Elaboración propia

Modelos de medición: Especifica las ecuaciones que vinculan las variables latentes a las observadas o indicadores x e y , conformando un sistema de ecuaciones, que expresado de forma matricial tiene la siguiente forma:

$$x = \Lambda_x \xi + \delta$$

$$y = \Lambda_y \eta + \varepsilon$$

Λ_x : Es la matriz de coeficientes de los indicadores de la variable exógena

Λ_y : Son las matrices de coeficientes de los indicadores de variables endógenas

ξ : Son las variables latentes exógenas

η : Son las variables latentes endógenas

δ y ε Son los errores de medida

Modelo estructural: Especifica las ecuaciones causales lineales entre las variables latentes del modelo.

$$\eta = \beta\eta + \tau\xi + v$$

siendo:

η : variables endógenas

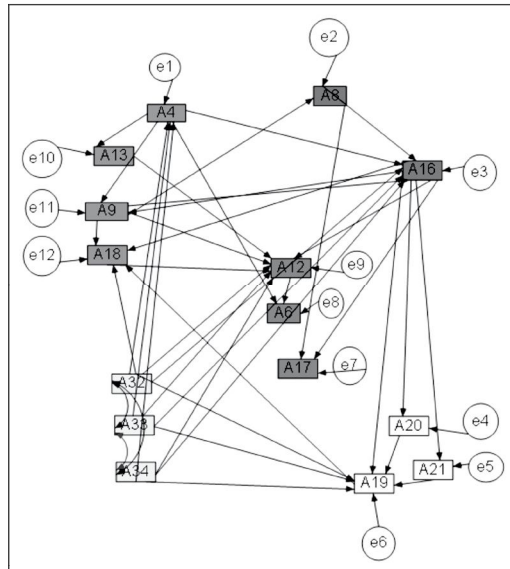
ξ : variable exógena

β : matriz de coeficientes de las variables endógenas

τ : matriz de coeficiente de la variable exógena

v : término perturbación aleatoria

El modelo estructural se encuentra conformado por un conjunto de ecuaciones estructurales, a las cuales se las expresa de la siguiente manera:



Elaboración propia

A continuación se muestra el sistema de ecuaciones correspondiente al modelo estructural:

$$A4 = \beta_0 + \beta_1 A32 + \beta_2 A33 + \beta_3 A34 + e_1$$

$$A13 = \beta_4 + \beta_5 A4 + e_{10}$$

$$A9 = \beta_6 + \beta_7 A4 + \beta_8 A16 + e_{11}$$

$$A18 = \beta_9 + \beta_{10} A9 + \beta_{11} A16 + \beta_{12} A19 + \beta_{13} A32 + e_{12}$$

$$A19 = \beta_{14} + \beta_{15} A32 + \beta_{16} A33 + \beta_{17} A34 + \beta_{18} A16 + \beta_{19} A20 + \beta_{20} A21 + e_6$$

$$A20 = \beta_{11} + \beta_{22} A16 + e_4$$

$$A21 = \beta_{23} + \beta_{24} A16 + e_5$$

$$A16 = \beta_{25} + \beta_{26} A8 + \beta_{27} A4 + \beta_{28} A32 + \beta_{29} A33 + \beta_{30} A34 + e_3$$

$$A8 = \beta_{31} + \beta_{32} A9 + e_2$$

$$A12 = \beta_{33} + \beta_{34} A13 + \beta_{35} A9 + \beta_{36} A18 + \beta_{37} A32 + \beta_{38} A33 + \beta_{39} A34 + \beta_{40} A16 + e_9$$

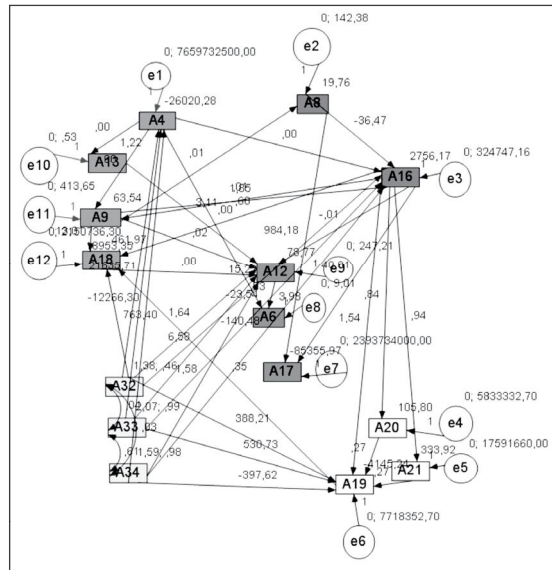
$$A6 = \beta_{41} + \beta_{42} A4 + \beta_{43} A12 + e_8$$

$$A17 = \beta_{44} + \beta_{45} A8 + \beta_{46} A16 + e_7$$

Estimación de los parámetros del modelo

El modelo seleccionado, por criterios de estimación, es el realizado mediante máxima verosimilitud, aunque previamente, han sido aplicados sobre los mismos datos otros tres métodos de estimación, con objeto de verificar la bondad del ajuste: Mínimos Cuadrados Parciales, Mínimos Cuadrados Generalizados y Distribución Libre Asintótica. El procedimiento ha sido estimar a la vez los valores de las variables independientes a partir de los valores estimados de las variables latentes dependientes, empleando un procedimiento de iteración en el que el algoritmo asigna valores a los ponderadores de los indicadores, hasta encontrar una convergencia estable.

Los parámetros estimados del sistema, mediante máxima verosimilitud, se muestran en el siguiente diagrama.



Elaboración propia

Ajuste del modelo

Existen diversos programas informáticos para ajustar modelos: EQS, LISREL, AMOS, CALIS, etc. En la elaboración del ECSI se ha utilizado el programa AMOS, por la posibilidad de trabajar con diagramas. La evaluación del modelo se ha efectuado utilizando diversos índices:

- Índice de ajuste normado como medida de discrepancia entre el modelo ajustado y el modelo base (NFI: NormedFitIndex).
- Índice de bondad de ajuste; similar al anterior, compara las discrepancias entre el modelo ajustado y el modelo anterior al ajuste (GFI: Goodness of FitIndex).
- Índice ajustado de bondad del ajuste; es el mismo indicador que el anterior pero ponderado por un ratio de los grados de libertad del modelo base y ajustado (AGFI: Adjusted Goodness of FitIndex).
- Índice del ajuste parsimonioso; obtenido a partir del índice NFI y ponderado por el cociente de los grados de libertad del modelo ajustado y el modelo base (PNFI: Parsimonius NormedFitIndex).

- Índice del radical del error de aproximación medio, que se obtiene como la raíz cuadrada de la ratio del parámetro no central ajustado por los grados de libertad (RMSEA: Root Mean Square Error Aproximation).

Dados los siguientes cuadros, se muestra que el modelo tiene la mayor parte de los parámetros no significativos.

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
A4	<—	A32	18953,349	7149,607	2,651	,008	
A4	<—	A33	21635,706	6237,731	3,469	***	
A4	<—	A34	-12266,300	6285,702	-1,951	,051	
A16	<—	A32	15,219	47,057	,323	,746	
A16	<—	A33	-23,545	41,371	-,569	,569	
A16	<—	A34	-140,481	41,605	-3,377	***	
A16	<—	A4	,004	,000	10,098	***	
A9	<—	A4	,000	,000	-,200	,841	
A20	<—	A16	,845	,163	5,179	***	
A21	<—	A16	,944	,283	3,332	***	
A19	<—	A16	1,536	,202	7,599	***	
A19	<—	A21	,274	,037	7,481	***	
A19	<—	A20	,270	,064	4,240	***	
A19	<—	A32	388,214	227,604	1,706	,088	
A19	<—	A33	530,734	198,352	2,676	,007	
A19	<—	A34	-397,625	202,832	-1,960	,050	
A18	<—	A32	763,402	120,294	6,346	***	
A13	<—	A4	,000	,000	-,966	,334	
A18	<—	A16	-,598	,118	-5,080	***	
A18	<—	A19	,348	,026	13,316	***	
A18	<—	A9	13,013	3,994	3,258	,001	
A12	<—	A16	-,013	,001	-11,144	***	
A12	<—	A9	,024	,043	,543	,587	
A12	<—	A13	3,114	1,193	2,610	,009	
A12	<—	A32	1,638	1,358	1,206	,228	
A12	<—	A33	6,575	1,126	5,839	***	
A12	<—	A34	1,585	1,150	1,378	,168	
A12	<—	A18	,000	,000	-,155	,876	
A6	<—	A12	,035	,008	4,272	***	
A17	<—	A8	984,184	269,018	3,658	***	
A6	<—	A4	,000	,000	-1,892	,058	
A17	<—	A16	40,007	3,905	10,245	***	
A8	<—	A9	,010	,091	,109	,913	
A9	<—	A16	,008	,007	1,104	,270	
A16	<—	A8	-36,470	4,053	-8,998	***	
A16	<—	A9	1,848	5,572	,332	,740	

**Standardized Regression Weights:
(Group number 1 - Default model)**

			Estimate
A4	<—	A32	,143
A4	<—	A33	,239
A4	<—	A34	-,134
A16	<—	A32	,013
A16	<—	A33	-,029
A16	<—	A34	-,169
A16	<—	A4	,427
A9	<—	A4	-,025
A20	<—	A16	,276
A21	<—	A16	,181
A19	<—	A16	,354
A19	<—	A21	,329
A19	<—	A20	,191
A19	<—	A32	,074
A19	<—	A33	,149
A19	<—	A34	-,110
A18	<—	A32	,265
A13	<—	A4	-,053
A18	<—	A16	-,251
A18	<—	A19	,634
A18	<—	A9	,143
A12	<—	A16	-,502
A12	<—	A9	,024
A12	<—	A13	,109
A12	<—	A32	,054
A12	<—	A33	,315
A12	<—	A34	,075
A12	<—	A18	-,007
A6	<—	A12	,231
A17	<—	A8	,207
A6	<—	A4	-,102
A17	<—	A16	,580
A8	<—	A9	,018
A9	<—	A16	,290
A16	<—	A8	-,529
A16	<—	A9	,048

Means: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
A32			1,379	,038	36,648	***	
A33			2,067	,055	37,468	***	
A34			1,593	,055	29,130	***	

Intercepts: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
A4			-26020,276	14586,982	-1,784	,074	
A9			63,540	13,098	4,851	***	
A8			19,762	7,158	2,761	,006	
A16			2756,168	518,163	5,319	***	
A20			105,800	352,492	,300	,764	
A21			333,921	612,132	,546	,585	
A19			-4145,244	607,729	-6,821	***	
A13			1,219	,042	29,068	***	
A18			461,967	374,097	1,235	,217	
A12			78,774	4,628	17,019	***	
A6			3,977	,655	6,071	***	
A17			-85355,974	12031,268	-7,095	***	

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
A32	<->	A34	,032	,037	,858	,391	
A33	<->	A34	,615	,064	9,565	***	
A32	<->	A33	,045	,038	1,193	,233	

Correlations: (Group number 1 - Default model)

			Estimate
A32	<->	A34	,048
A33	<->	A34	,625
A32	<->	A33	,066

Variances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
A32			,462	,036	12,767	***	
A33			,992	,078	12,767	***	
A34			,975	,076	12,767	***	
e1			7659732532,235	599956554,999	12,767	***	
e11			413,654	33,505	12,346	***	
e2			142,379	13,285	10,717	***	
e3			324747,159	37730,792	8,607	***	
e4			5833332,674	456901,878	12,767	***	
e5			17591660,064	1377885,158	12,767	***	
e6			7718352,658	604548,038	12,767	***	
e10			,531	,042	12,767	***	
e12			2150736,309	168458,669	12,767	***	
e9			247,208	19,363	12,767	***	
e8			9,008	,706	12,767	***	
e7			2393733979,646	187491715,405	12,767	***	