MODELIZACIÓN Y SIMULACIÓN DE SISTEMAS COMPLEJOS

Antonio Caselles Moncho

UNIVERSITAT DE VALÈNCIA

Este texto ha sido diseñado para la realización de un curso en la Universidad de Entre Ríos (Argentina) con una ayuda de la AECI en el marco de la convocatoria realizada por Resolución de 6 de julio de 2007, de la Agencia Española de Cooperación Internacional, publicada en el Boletín Oficial del Estado de 23 de julio.

ISBN: 978-84-370-7198-5
Depósito legal: V-3913-2008

Edición digital

Índice

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

1. INTRODUCCIÓN AL PENSAMIENTO SISTÉMICO

El paradigma predominante hasta hace unas pocas décadas era el paradigma mecanicista, basado en las ideas de Descartes y que podríamos resumir con la frase "divide y vencerás". Este paradigma conduce a la especialización. Es claro que un especialista es capaz de resolver un determinado tipo de problema mejor que alguien que no lo es. No obstante, existen problemas, "problemas complejos" que implican a más de una especialidad y para ser resueltos necesitan un equipo interdisciplinario de especialistas. De aquí surge el paradigma sistémico basado en las ideas de Von Bertalanffy y que podríamos resumir con la frase "el todo es más que la suma de las partes". Este paradigma se basa en el concepto de sistema: "conjunto de elementos interrelacionados". El Universo es un sistema, y está compuesto por sistemas de menor entidad o "subsistemas", y estos subsistemas están compuestos pos sub-subsistemas, y así indefinidamente. Cuando aplicamos la técnica "divide y vencerás" no debemos olvidar que cada una de las partes está relacionada con las demás y, si queremos aislarla por conveniencia, no debemos olvidar sus relaciones con el resto del universo.

Hasta aquí hemos hablado de "elementos" de "relaciones" de "partes" y de "el todo". Estas son palabras muy generales. Pero, en un caso concreto ¿qué es un elemento? ¿y una relación? ¿cómo hacemos las partes? ¿a qué estamos llamando "el todo"? La respuesta a estas preguntas depende del problema que estemos estudiando, de los objetivos que persigamos. Normalmente, lo que pretendemos es hacer un "modelo" del sistema de la vida real sobre el que estamos trabajando y sobre el que queremos intervenir para resolver un problema que nos preocupa. Un modelo es como un dibujo, una maqueta, una escultura, una descripción literaria, unas ecuaciones matemáticas, quiere ser una representación aproximada, simplificada, del sistema real, de su estructura y de su comportamiento. El objetivo con el que construimos un modelo es el de obtener de él respuestas que el sistema real tardaría en darnos, sería costoso y quizá peligroso obtener del mismo. A esto se le llama "simulación", hacer experimentos sobre el modelo en lugar de hacerlos sobre el sistema real. Claro que estas respuestas deben ser fiables. Y, si el modelo es una simplificación ¿hasta qué punto nos podemos fiar de sus respuestas? La respuesta a esta pregunta tiene dos partes. Primero, debemos asegurarnos de que el modelo es una representación "válida" de la realidad. Y segundo, conviene que calculemos la fiabilidad con la que se presentan los resultados.

Obviamente, existen métodos y tecnología que nos ayudan a llevar a buen fin lo dicho anteriormente. A continuación presentaremos una metodología (conjunto organizado de métodos) para construir modelos lógico-matemáticos de sistemas complejos de la vida real que nos ayuden a resolver el modo de intervenir sobre los mismos para lograr un determinado objetivo. Llamaremos "modelo" a una lista de variables matemáticas y a una lista de relaciones funcionales entre las mismas. Llamaremos "simulador" a la representación informatizada (programa de ordenador o aplicación informática) del modelo. Trabajaremos con el simulador para tratar de obtener de él respuestas a preguntas del tipo "¿qué pasaría si….?", y a otras preguntas más sofisticadas.

2. METODOLOGIA SISTÉMICA

Objetivo del tema: después de una introducción al pensamiento sistémico y antes de la presentación formalizada de la Teoría General de Sistemas (presentación formalizada de conceptos y relaciones entre los mismos, demasiado ardua de entrada) se pretende dar una

primera visión de la metodología sistémica general, que normalmente conduce a la construcción de modelos ad-hoc para cada tipo de problema, para experimentar sobre ellos y, como consecuencia, tomar decisiones. Se proponen, por tanto, los siguientes objetivos de aprendizaje:
1. distinguir entre modelos mentales y modelos para computadora, modelos cualitativos y modelos cuantitativos, modelos discretos y modelos continuos, etc.;
2. adquirir un conjunto de métodos para la construcción de modelos;
3. aplicar estos métodos a la construcción de algunos modelos de sistemas complejos "sencillos".

2.1 EL MÉTODO CIENTÍFICO Y LA TEORÍA DE SISTEMAS.

La construcción de modelos lógico-matemáticos y de simuladores sigue el método científico. Asumimos que trabajamos "científicamente", pero ¿qué es trabajar científicamente? Para responder a esta pregunta con pocas palabras podríamos decir que trabajamos tratando de asegurarnos de que lo que decimos es cierto o, al menos, tiene un grado de certeza conocido y aceptable. No obstante, somos conscientes de que cualquier explicación que damos a un fenómeno de la vida real es válida siempre provisionalmente. En los siguientes párrafos encontraremos una respuesta más detallada.

El método científico puede ser resumido de la siguiente manera (Wartofsky, 1968):
- Observación del colectivo objeto de estudio.
- Formulación de una teoría que explique las propiedades de sus elementos y de las relaciones entre los mismos.
- Uso de la teoría para la predicción de eventos relacionados con el colectivo.
- Experimentación-para determinar si las predicciones son acertadas o no. Cuando no lo son procede modificar la teoría o aparcar el problema provisionalmente (Kuhn, 1962).

Una teoría es siempre provisional y se mantiene mientras no se observen sucesos importantes que esta teoría prohíbe. Cuando uno de estos sucesos aparece procede modificar la teoría a fin de explicar este suceso y todos los anteriores, usando la imaginación, la observación y la experimentación cuantas veces sea necesario. El punto de comienzo no es necesariamente la observación porque el investigador puede tener suficiente conocimiento previo del colectivo objeto de estudio para poder elaborar su primera teoría. Hay situaciones en las que el concepto "programa de investigación" (Lakatos, 1970) es más adecuado que el concepto de "teoría", pero 1os principios inductivistas, convencionalistas, sistémicos y falsacionistas que subyacen en la anterior síntesis pueden ser mantenidos en su justa medida, por el momento. Es posible conectar la Epistemología con la Teoría de Sistemas (Buckley, 1972). La contribución más importante para la Epistemología que se deriva de esta conexión es la referencia al control. En otras palabras, la Naturaleza es algo que, no solamente puede ser observado, sino que puede ser transformado y controlado por el ser humano. Y, e1 conocimiento es algo más que contemplación y entendimiento. Siguiendo esta idea, es posible decir que un conocimiento es una representación fiable de una parte de la realidad. Y, llegados a este punto, consideramos conveniente recordar los siguientes conceptos:
- Un sistema es un conjunto de de elementos interrelacionados. (Von Bertalanffy, 1972).
- Analizar un sistema es identificar sus elementos y las interrelaciones entre los mismos.
- Describir un sistema es presentar un modelo del mismo, es decir, hacer sus elementos e interrelaciones visibles y comprensibles.

- Para un observador O, un objeto M, es un modelo de otro objeto A, en la medida en que O puede obtener de M respuestas a cuestiones que se refieren a A. (Minsky, véase Melése, 1968).

Así pues, el problema del conocimiento puede ser reducido a uno o más problemas de modelización de sistemas. Puede haber diferentes clases de modelos:
- modelos verbales, que son las descripciones tradicionales, donde los elementos son conceptos y las relaciones son conexiones lógicas; modelos gráficos, como por ejemplo diagramas, dibujos, etc.;
- modelos físicos, como son las maquetas (modelos estáticos), o los simuladores de vuelo para entrenamiento de pilotos (modelos dinámicos);
- modelos de estructura, donde se representan solamente aspectos como la posición relativa o la dependencia causal;
- modelos de comportamiento, donde se representan las reacciones del sistema a las influencias externas, teniendo en cuenta aspectos tales como retroalimentación, control, etc.
- etc.

Consideramos conveniente también puntualizar la diferencia entre el significado de la palabra modelo en Lógica y en Teoría de Sistemas. En Lógica una teoría es, en síntesis, una lista de axiomas y una lista de consecuencias de estos axiomas. Y, un modelo de una teoría es un colectivo que satisface estos axiomas. En otras palabras, una teoría es un "sistema formal" (véase, por ejemplo, Hackstaff, 1966) cuyos elementos y relaciones son:
- símbolos primitivos, sin ningún significado;
- secuencias de símbolos;
- reglas de formación, que indican cuando una secuencia de símbolos "es una fórmula" (está bien formada, pertenece al sistema);
- axiomas, que son las fórmulas iniciales; y
- teoremas, que son fórmulas obtenidas a partir de los axiomas aplicando las reglas de formación;
y un modelo es una "interpretación" de una teoría, o bien, un modelo es un sistema cuyos elementos y relaciones se corresponden con los de la teoría.

En Teoría de Sistemas la noción que ocupa la posición central es, obviamente, la de "sistema" y no la de "teoría". De aquí se pude deducir que una teoría (explicación sistemática, es decir, acertada y formal) es un modelo verbal, y/o matemático de un sistema. Es posible que existan otros sistemas que también estén en correspondencia con la misma teoría pero este es otro problema. El problema que nos interesa aquí es como obtener una buena teoría de un sistema dado. Este es el problema del conocimiento. Y la conclusión a la que creemos haber llegado es que una teoría del conocimiento es una teoría de la modelización de sistemas y que cada rama del conocimiento puede ser obtenida desde esta meta-teoría.

Modelizar un sistema comienza por identificar sus elementos y las relaciones entre ellos. La modelización se dice "multifacética" o "perspectivista", (Zeigler, 1984) porque el modelo que se construya de un sistema real depende del objetivo del modelizador. El grado de detalle al que se llegue en la descripción también depende de este objetivo. Así pues, una primera aproximación puede ser un modelo de tipo "caja negra" o "input-output", donde solo se especifican las entradas y las salidas (los datos y los resultados). Cuando se va incrementando el detalle en la descripción se va transformando esa caja negra en "caja transparente" (Bunge, 197D: y se va ajustando más el modelo a la realidad. No obstante un perfecto ajuste entre modelo y realidad solo es posible en ciertos casos particulares pero no en general. Las "teorías generales" tienen una probabilidad igual a cero, y el progreso en la ciencia consiste en incrementar el detalle y reducir los errores (Popper, 1912). En general un modelo es aceptable cuando las diferencias entre él y el sistema real no son aparentes (incluyendo el error experimental), o cuando siendo aparentes no molestan. Cuando el objetivo del modelizador es el puro conocimiento el proceso de ajuste del modelo a la realidad puede no tener fin. El sistema real puede ser construido y/o descubierto elemento a elemento y relación a relación de modo paralelo al modelo si así lo determina el objetivo del modelizador (o colectivo investigador, obviamente), es decir, la modelización puede ser una ayuda al descubrimiento, a la creatividad, al diseño o a la construcción de sistemas reales.

En general el modelizador debe empezar por concretar: su objetivo y la parcela de la realidad que desea considerar. El "nivel de resolución" o grado de detalle depende de ese objetivo y es difícil de determinar a priori. El proceso de modelización es esencialmente cíclico, es decir, que vuelve atrás para corregir resultados de etapas anteriores como consecuencia de los resultados de etapas posteriores que conllevan contrastes con la realidad. Esto es 1o que tradicionalmente se describe como un proceso de "observación-inducción-deducción-experimentación" con ciertas restricciones (véase Wartofsky, 1968). El avance de este movimiento cíclico es, primero en el sentido de un mejor ajuste del modelo a la realidad y después en el sentido de una mayor desagregación, hasta lograr la satisfacción del modelizador. El punto de partida es probable que no sea la "observación" porque el modelizador generalmente cuenta con suficiente conocimiento del sistema como para construir su primer modelo o "teoría". Este primer paso no es normalmente un paso lógico en el estricto sentido de la palabra "lógico" (véase Lakatos, 197l). La intuición, la imaginación o un "principio inductivo" extra-metodológico, puede intervenir de manera importante en este punto. La primera teoría y las consecutivas son siempre convencionales y provisionales. (Teoría: establecimiento acertado y formal de elementos y relaciones: modelo verbal y/o matemático). De esta primera teoría se pueden sacar muchas consecuencias que deben ser contrastadas con el sistema real. Esto conlleva la realización de una serie de experimentos cuyos resultados determinarán el grado de ajuste entre el modelo y el sistema real. Si este grado satisface al modelizador este aceptará el modelo para sus propósitos, y si no, volverá a la teoría y/o al sistema real y cambiará, encontrará o construirá los elementos y las relaciones que se requieran a fin de que una nueva serie de experimentos muestre un grado de ajuste mejor que el precedente. La palabra "validación" el probablemente una de las más adecuadas para expresar que se desea lograr un grado de ajuste provisionalmente suficiente. Otras palabras como verdadero o falso, corroboración y falsificación o verificación (Popper, I976) también han sido usadas en relación con una teoría. Cuando el sistema real es muy complejo y no está bien definido desde el principio, el proceso de modelización es un auténtico "programa de investigación", donde el modelo se construye sección a sección (bloque, o submodelo), para después conectar estas secciones (también mal definidas) como un rompecabezas. Por otra parte, un proceso muy complejo necesita más de una, quizás muchas personas, y posiblemente más de una generación para su desarrollo. En este caso coexisten diferentes opiniones sobre los principios básicos del proceso de modelización (el "núcleo firme" de Lakatos, o el "paradigma" de Kuhn) y sobre la manera de desarrollarlos (la "heurística positiva" de Lakatos). El proceso de ajuste de un modelo a gran escala al sistema real correspondiente implica muchos cambios en el modelo y/o en el sistema. Estos cambios pueden tener mayor o menor entidad dependiendo del tamaño y de la situación de la parte a renovar. Los cambio que afectan a un "núcleo firme" o a un "paradigma" pueden ser "revoluciones científicas" (Khun, 1962). Los cambios que afectan a hipótesis auxiliares a fin de explicar nuevos "sucesos" se llaman cambios "had hoc". Mucho se ha escrito en relación con el grado de ajuste de un modelo a la realidad. Cuestiones tales como ¿Cómo encontrar la verdad? (Bacon-Descartes), o ¿Es posible la certeza absoluta, o encontrar las explicaciones más profundas? (Hume-Kant), condujeron el problema de evaluar y comparar teorías a la metodología de Popper (véase Watkins, 1980, por ejemplo) y a la metodología de Lakatos. Hay opiniones (Miller, 1976) en el sentido de que la similaridad entre una teoría y la verdad no se puede expresar mediante una definición formal, o en el sentido de que el contenido en verdad o falsedad de dos teorías no es comparable, estas "ven el mundo de maneras diferentes" Feyerabend, el último Wittgenstein). Pero probablemente sí es posible construir un índice que evalúe el grado de ajuste entre un modelo dado y un sistema real dado (Caselles, 1984). Este índice, como todo modelo, será siempre convencional y provisional y perfectible, y permanecerá vigente mientras sea útil o suficiente para los propósitos que se le asignen.

2.2 ETAPAS DE LA MODELIZACIÓN

2.2.1 La toma de decisiones y la resolución de problemas.

Imaginemos que nos encontramos ante un problema en un sistema complejo. Es decir, tenemos que tomar decisiones respecto al sistema a fin de maximizar o minimizar algo. El pilotaje de los sistemas complejos necesita de modelos y de computadoras para ser eficiente. Construir un modelo que estime la fiabilidad de sus resultados, sobre un sistema complejo puede ser una tarea descorazonadora, necesita de muchas horas de trabajo y de muchas personas, y además el modelo necesita una puesta al día constante porque los sistemas reales evolucionan muchas veces de manera imprevisible. Por consiguiente, para que los modelos sean un instrumento eficaz de ayuda en la toma de decisiones deben poder ser construidos y actualizados de manera poco costosa. Para ello y una vez más necesitamos a las computadoras. Pero su uso está condicionado por la metodología que se utilice para la modelización. Existen diferentes metodologías para modelizar sistemas reales con diferente eficiencia y generalidad. Caselles (1993b) propone una metodología que parece bastante general y eficiente. Vamos a estudiarla y a compararla con otras metodologías alternativas.

2.2.2 Metodología general para modelizar

Nosotros entendemos por metodología un conjunto de métodos organizados con un fin determinado. Check1and (1981) hace una discusión bastante extensa sobre los posibles significados de la palabra metodología. Una metodología general para construir modelos sería muy de desear. Gorokov (1985) llama "configurador" a un hipotético modelo universal válido para todos los sistemas. Este configurador debería incluir un esquema general de tipo ontológico y un esquema general de comportamiento, debería asumir y sintetizar las diferentes aproximaciones a la modelización, y debería ser capaz de integrar modelos parciales o específicos en uno más complejo. Este configurador no existe todavía. Hay diferentes metodologías para modelizar pero todas ellas están restringidas en algún sentido (véase por ejemplo, Gelovany, 1985 y O'Keefy, 1989).

Caselles (1993b) propone un proceso modelizador que trata de organizar métodos parciales y trata de integrar las ideas de Forrester (1961 y 1966), Check1and (1981), Morecroft (1982), Balci (1986), Mathewson (1989), Zhang et al. (1990) y Caselles (1992 y 1993a). Este proceso es el siguiente (más adelante se detalla).

1. Descripción del problema. Esto es, especificación de los objetivos y de los condicionantes en lenguaje natural. Se sugieren como métodos adecuados para ello el Brainstorming, el Delphi y similares.

2. Construcción de un modelo conceptual.

2.1 Elección de los objetos, elementos o variables que tengan alguna relación con los objetivos propuestos. Métodos que se sugieren: Brainstorming, Delphi, y similares.

2.2 Identificación de relaciones causa-efecto entre los elementos seleccionados. Métodos sugeridos: Diagrama Causal, Diagrama de Subsistemas, Diagrama hidrodinámico o de Forrester (véase Morecroft, 1982), Análisis Multivariante, y métodos similares.

2.3 Asignación de una representación funcional a las relaciones detectadas. Es decir, escribir dichas relaciones (una variable dependiendo de otras) como ecuaciones y/o tablas y/o reglas lógicas. Métodos sugeridos: Brainstorming, Delphi, Regresión, Ecuaciones Diferenciales, Integración numérica, etc.

3. Programación del modelo para una computadora o instrucción de un grupo de expertos (generalmente el mismo que ha construido el modelo conceptual) sobre la estructura y el comportamiento del mismo. Métodos que se sugieren: Hoja de Cálculo, Generadores de aplicaciones (herramientas CASE), Lenguajes de Simulación, Interpretadores de Descripciones (Davies y O'Keefy, 1989), técnicas de Dinámica de Grupos humanos (para modelos mentales o "soft"), etc.

4. Calibrado del modelo. Algunos tipos de modelos requieren esta operación, que consiste en asignar un valor lo más adecuado posible a los parámetros del modelo una vez construido este. Ello se suele lograr con los métodos de prueba y error o con estudios específicos.

5. Análisis de sensibilidad. Algunos tipos de modelos y concretamente los que requieren de la operación de calibrado necesitan también del análisis de la sensibilidad de las variables endógenas frente a pequeñas variaciones de los valores de los parámetros. También esto se suele lograr por los métodos de prueba y error o por estudios específicos.

6. Evaluación de la validez o utilidad del modelo para el logro de los objetivos propuestos. Métodos que se sugieren: establecimiento de reglas y/o ecuaciones y/o tablas para determinar el grado de ajuste entre el modelo y el sistema real.

7. Diseño de experimentos o de procedimientos de optimización para ser realizados sobre el modelo. Métodos sugeridos: los de la Estadística y la Investigación Operativa, escenarios y estrategias, etc.

8. Realización de los experimentos o procedimientos de optimizaci6n diseñados. Métodos que se sugieren: simulación sobre computadora, simulación sobre un grupo humano (para la aproximación "soft") y similares.

9. Presentación de los resultados obtenidos. Métodos sugeridos: elaboración de tablas, gráficos, dibujos, documentos, etc.

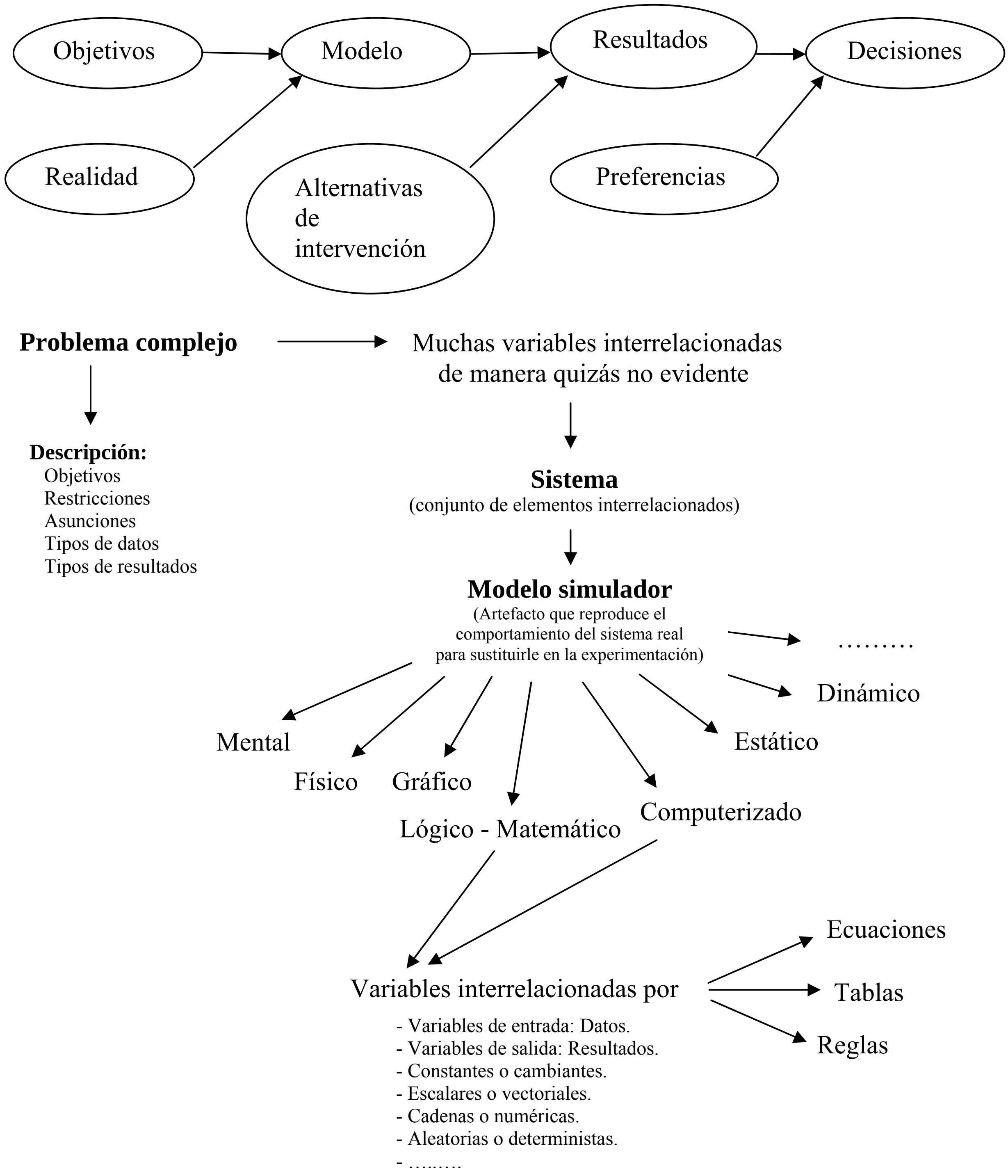
10. Toma de decisiones. Métodos sugeridos: los de la Teoría de la Decisión.

El proceso anterior es secuencial, pero:

a. Es necesario un proceso de verificación en cada paso. Es decir, hay que comprobar que cada paso ha sido bien ejecutado. Esta comprobación conduce al equipo, además de a corregir errores, a volver atrás y rehacer y/o modificar los pasos anteriores y sus resultados. Cada paso conduce al equipo modelizador a descubrir nuevos elementos, relaciones y/o objetivos que no habían sido tenidos en cuenta en los pasos anteriores, y como consecuencia a modificarlos o a reconsiderarlos. Los datos de campo son necesarios en la mayoría de los pasos. Por tanto, la adquisición de datos, su gestión, análisis y validación están siempre presentes.

Así pues, la situación podría resumirse de la siguiente manera:

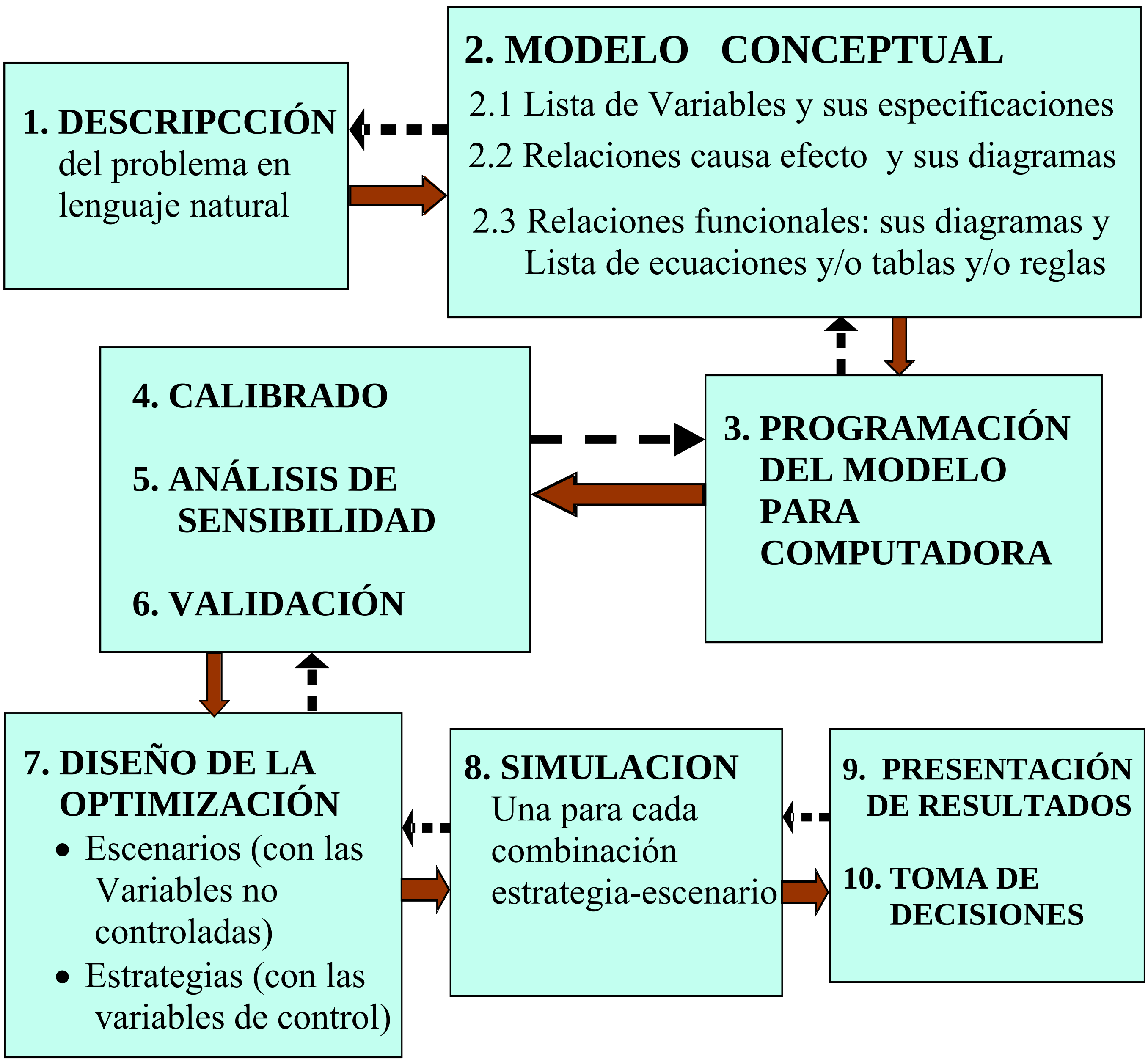
Figura 1. Problema complejo y su simulador



Obsérvese que el proceso modelizador que se propone trata de integrar las metodologías "soft", basadas en modelos mentales, (Checkland, 1981), y las "hard", y con respecto a esta últimas, la aproximación "inductiva", basada en los datos de campo, (Klir, 1985, por ejemplo), y la aproximación "deductiva" o realista, basada en estructuras y comportamientos hipotéticos propuestos para el sistema (Zeigler, 1984 por ejemplo).

El grafico de la Figura 2 puede también ayudar a visualizar el proceso descrito.

Figura 2. METODOLOGÍA DE MODELIZACIÓN GENERAL: trama que guía un conjunto organizado de métodos



2.2.3 Ayudas informáticas

Cada uno de los pasos descritos en el punto 2.2.2 puede ser realizado con métodos que pueden incluir ayuda informática. El proceso completo podría ser realizado dentro de un entorno informático. La necesidad y las características de esta clase de entorno son estudiadas por Balci (1986). Un entorno de este tipo ya comercializado pero restringido a una clase de problemas muy específica es descrito por Standbridge (1985). En el futuro, interfaces que usarán el lenguaje natural guiarán el proceso completo dentro de los entornos de modelización. Hoy en día las herramientas de modelización más modernas inciden en los pasos 2.2, 3 y 9. Las que inciden en los pasos 2.2 y 9 son, generalmente, herramientas de uso general en los campos gráfico y/o estadístico. Las que inciden en el paso 3 pueden ser clasificados en:

- Lenguajes de simulación,
- Generadores de aplicaciones,
- Interpretadores de descripciones.

La programación de un modelo para una computadora puede realizarse con un lenguaje de programación de tipo general (C, PASCAL, BASIC, FORTRAN, etc.) o con un lenguaje especialmente diseñado para ahorrar tiempo de programación cuando tratamos con modelos de determinados tipos. Por ejemplo, un tipo muy común de problemas que se dan en la industria son los que conducen a modelos llamados de redes de colas, que se asocian generalmente con la llamada "simulación de sucesos discretos" (DEVS) (véase por ejemplo, Davies y O'Keefe, 1989). Alrededor de este tipo de simulación se ha desarrollado una metodología muy completa (Zeigler, 1984, 1987, 1989, 1990). Los lenguajes de simulación más populares son SIMSCRIPT, GPSS y SIMULA. Otros lenguajes importantes son QNAP2 y STIMS. Todos ellos tienen facilidades para definir entidades, usar números aleatorios, manejar listas, elaborar tablas y gráficos, etc. Lenguajes de tipo algo más general que tienen facilidades para programar simuladores son el Mathematica y el Matlab/Simulink.

Para nosotros, generadores de aplicaciones o generadores de programas son programas que dialogan con el usuario, interpretan sus respuestas y construyen otros programas que pueden ser compilados y ejecutados de manera independiente, fuera del generador, con el fin de ejecutar las simulaciones diseñadas por el usuario. El uso de generadores de programas produce un considerable ahorro de tiempo, sin embargo, normalmente exigen adaptarse a una metodología muy rígida. Como ejemplos de generadores de programas adaptados a la DEVS tenemos: GRAFT (Matheuson, 1984) que puede producir programas escritos en varios lenguajes, eLSE, que genera programas en PASCAL (Crookes et al., 1986). Un generador de programas de simulación orientados a objetos llamado GASPE es descrito por Simonot et al. (1990). Zhang et al. (1990) compara un generador que usa interface de tipo dialogo con otros que usan interfaces de tipo gráfico, todos generando programas en GPSS, llagando a la conclusión de que la interface de tipo dialogo es más eficiente.

Los "interpretadores de descripciones" suelen presentar al usuario una "hoja" o menú donde el usuario da una descripción semi-formal del sistema. Esta descripción es interpretada inmediatamente. No se genera ningún programa en lenguaje fuente ni de forma que pueda ser ejecutado fuera del entorno en que ha sido producido. Tampoco se mantiene dialogo alguno con el usuario. Como ejemplos de este tipo de entornos tenemos Inter-SIM (O'Keefy, 1987) y HOCUS (Ponle y Szymankiewicz, 1987), ambos adaptados a DEVS. En esta categoría podríamos incluir también a las hojas de cálculo (EXCEL por ejemplo) y a los modernos entornos como son el Stella y el Vensim.

En general, las ayudas informáticas a la modelización pueden considerarse como demasiado restringidas y poco eficientes respecto a lo que sería de desear. Se hace necesario progresar hacia una metodología con un máximo de ayuda informática y capaz de enfocar un amplio espectro de tipos de problemas, siendo realmente eficiente. Nosotros hemos estado trabajando en esta línea desde 1986.

El generador de programas SIGEM, fue presentado por primera vez por Caselles (1988). El amplio espectro de problemas que pueden ser tratados en la actualidad por esta ayuda informática se describe en otro trabajo (Caselles, 1991 y 1994). Este espectro va desde sistemas estáticos como bases de datos con variables numéricas y/o literales a sistemas dinámicos con variables numéricas y/o literales, con o sin incertidumbre, dimensionadas o no, y con relaciones funcionales dadas mediante ecuaciones y/o reglas lógicas y/o tablas, con o sin incertidumbre. Los programas construidos por SIGEM dialogan con el usuario, dándole indicaciones sobre cómo proceder, planteándole preguntas y menús y ejecutando sus órdenes. Con respecto a la eficiencia de SIGEM diremos que con un análisis previo del problema que se materialice en un gráfico de conexiones entre variables o en un diagrama de Forrester y en una lista desordenada de relaciones funcionales, todo ello escrito a mano en una o varias hojas de papel, en un tiempo entre varios minutos y varias horas, con un ordenador personal, el modelo se convierte en programa de ordenador listo para ser usado como un amigable sistema experto que puede ayudar al usuario a tomar decisiones sobre el problema del que entiende. Como el programa producido esta en lenguaje fuente (Visual Basic), puede considerarse como un prototipo y puede ser modificado si se considera conveniente. SIGEM está enteramente fundado sobre la Teoría General de Sistemas, con una formalización novedosa que se describe en otros dos artículos de Caselles (1992b, y 1993a).

2.3 DETALLES DE LA METODOLOGIA PARA LA MODELIZACIÓN

La metodología propuesta en 2.2 de manera esquemática, pretende ser capaz de abordar problemas tanto desde el punto de vista "soft" (con modelos mentales y discusión en equipo) como desde el punto de vista "hard" (con modelos matemáticos y computadoras). Y, en este último caso, tanto desde el punto de vista inductivo (con modelos fenomenológicos obtenidos exclusivamente a partir de datos de campo) como desde el punto de vista deductivo (con modelos realistas, propuestos por el equipo como hipotéticos, y posteriormente validados), o desde un punto de vista mixto (con subsistemas de un tipo o del otro). Detallando un poco más, con esta metodología se pueden elaborar modelos de sucesos discretos, modelos continuos, cualitativos, cuantitativos, deterministas, estocásticos, etc., puesto que las cuestiones específicas de cada tipo de modelo se corresponden con métodos específicos en cada una de las fases del proceso.

A continuación vamos a detallar algo más en cada una de las fases del proceso referido, presentando las líneas generales de los principales métodos utilizables en cada caso y ofreciendo bibliografía para un estudio más profundo de los mismos.

2.3.1 Descripción del problema

Es comúnmente aceptado que la modelización es "multifacética" o "perspectivista", es decir el modelo que se construya dependerá de los objetivos del modelizador y de las asunciones que se hagan sobre el sistema real. Así, con diferentes objetivos y/o asunciones se obtendrán diferentes modelos del mismo sistema real. Aquí estamos considerando al sistema real como sinónimo de la porción de la realidad que nos suministra datos.

Como consecuencia de este perspectivismo, el primer paso del proceso de modelización debe ser la especificación de: (a) los ***objetivos del modelizador*** y de (b) las ***asunciones***, (c) ***restricciones***, (d) ***tipos de datos*** y (e) ***tipos de resultados,*** y de todo lo que pueda tener alguna influencia sobre los mencionados objetivos. En esta fase es conveniente utilizar el lenguaje natural ayudado de tablas, gráficos, y de lo que se estime conveniente, con la máxima libertad. No obstante, suele ser suficiente para empezar una descripción somera de los ítems (a), (b), (c), (d), (e) recientemente mencionados, dado que en las fases sucesivas se encontrarán ocasiones de mejorar esta descripción. No olvidemos que la metodología incluye continuas vueltas atrás para corregir las fases anteriores como consecuencia de lo que se encuentra en las fases siguientes.

Ejemplo 1:

Objetivos.
1. Una empresa que comercializa aparatos eléctricos de gran potencia desea optimizar el número de aparatos a guardar en su almacén.

Restricciones.
1. La demanda varia de un día a otro.
5. Todas las unidades solicitadas pero no disponibles se dejan en espera y se satisfacen cuando llega una nueva remesa.

Tipos de datos.
1. El aparato que comercializa le cuesta 5000 € por unidad.
2. El costo de mantener un aparato en el almacén es del 20% de su valor al año.
3. El costo de procesamiento de un pedido de cualquier número de unidades es de 200 € por pedido.
4. Se dispone de un registro de unidades vendidas día a día y de tiempos de espera desde que el cliente solicitó hasta que se le sirvió.

De tal registro se han obtenido las siguientes tablas:

Demanda	Frecuencia relativa (%)	Tiempo de espera (días)	Frecuencia relativa (%)
0	40	1	25
1	30	2	50
2	20	3	25
3	10		

Tipos de resultados.
1. Modelo de Dinámica de Sistemas hipotético, a validar en la medida de lo posible con los datos disponibles (No será posible hacerlo por el método de predicción del pasado, dado que no existen datos de pedidos efectuados, ni de existencias día a día, ni de demanda en espera. Se validará por opiniones de expertos, es decir, “si se creen el modelo deberán creerse sus resultados”).
2. Simulaciones de la evolución del número de aparatos vendidos, en almacén y coste de mantenimiento del inventario a lo largo del tiempo durante 30 días.
3. Número óptimo de aparatos a pedir a la fábrica cada vez.
4. Existencias en almacén que debe haber en el momento idóneo para hacer un pedido a fábrica.

Ejemplo 2:

Objetivo: Determinar los flujos óptimos de agua para transvase entre las cuencas hidrográficas de un país.

Restricciones:
- Ámbito nacional y zonas por cuadrículas del mapa.
- Existen dessaladoras y se pueden eliminar y/o construir otras.
- Se clasificará el agua por destinos: Población, Agricultura, Industria y Turismo.

Datos:
- Precipitaciones.
- Agua que producen y pueden producir las dessaladoras actuales.
- Caudales de los ríos.
- Necesidades de la población, agricultura, industria y turismo.

Resultados:
- Trasvases. Flujos óptimos de agua entre cuencas

2.3.2 Selección de las variables más relevantes

Nosotros dividimos este paso en dos:

- identificación de los elementos involucrados en el problema; y
- asignación de una forma a dichos elementos.

La primera parte intenta detectar qué elementos u objetos intervienen o influyen en el problema, es decir, en los objetivos recientemente planteados. Al final de la misma estos elementos quedan descritos de una manera poco rigurosa, generalmente en lenguaje natural. La segunda parte persigue dar a cada elemento identificado una descripción más rigurosa y detallada o bien asociarle una variable matemática, es decir, darle un nombre explicativo preciso y conciso, un nombre codificado (abreviado), una unidad de medida (cuando sea necesario), y un rango de valores posibles (que pueden ser números, palabras, conceptos, principios, reglas, métodos, etc.), una serie de valores observados en la realidad (si es posible), y si se trata de una variable con incertidumbre, determinar, si se puede, su media, su desviación típica, clase de distribución de sus valores, y/o otros parámetros o factores definitorios que se consideren adecuados.

En esta fase se suelen utilizar: (a) los métodos de creatividad en grupos y (b) el blanqueo de cajas negras: (c) en algún caso de modelización inductiva o mixta, el Análisis Multivariante. Este último método nos permite, dado un conjunto de variables numéricas con una serie de valores para cada una de ellas, determinar cuales están relacionadas y cuáles no tienen nada que ver la una con la otra a efectos prácticos. Estas últimas pueden ser eliminadas aunque en un principio pareciera que debían ser tomadas en consideración.

A continuación detallamos estos enfoques.

2.3.2.1 Métodos de creatividad en grupo

Existen métodos que ayudan a conseguir una mayor eficiencia en el proceso de búsqueda de los factores que intervienen en el problema que se está estudiando, cuando es un equipo de personas el que lo estudia. Entre estos métodos destacamos el Brainstorming y el Delphi que son descritos con cierto detalle por ejemplo por Sage (1977).

Son métodos de creatividad grupal que tratan de corregir los **defectos del grupo de discusión tradicional**. Entre estos defectos cabe destacar los siguientes:

– La opinión del grupo acaba siendo la del miembro que más fuerte y con más frecuencia habla. Existe una cierta presión grupal hacia la conformidad, bien por el prestigio del que prestó la idea, bien por existir una mayoría clara, por cansancio, temor, etc.

– Se tratan temas de interés personal para algunos o la totalidad de los miembros del grupo antes que el tema objeto de la reunión, lo que se suele llamar "ruido".

– Intervienen multitud de factores psicológicos que pueden enmascarar o afectar a la solución del problema así como a la duración de la reunión.

En el **brainstorming** el grupo debe nombrar un moderador, un secretario, y es conveniente que disponga de una pizarra. Es función del moderador:

– Explicar al grupo en qué consiste el brainstorming.

– Exponer lo que se desea conseguir.

– Dar la palabra a quien la tenga pedida, de modo ordenado, cuando proceda.

– Inyectar ideas o leer notas tomadas cuando el grupo se paraliza.

Es función del secretario: tomar nota de lo que se va diciendo y de las conclusiones a las que el grupo va llegando. El método tiene dos fases:

1ª. Una vez el moderador ha expuesto al grupo lo que se pretende conseguir, permite a los miembros del mismo aportar ideas. Ideas que deberán ser expuestas con una o muy pocas palabras y anotadas seguidamente. No está permitido ningún tipo de discusión o dialogo. Ninguna idea se considera descabellada en esta fase.

2ª. Cuando el grupo llega al punto de no aportar ya nada nuevo, el moderador insta a sus miembros a dar razones en pro y en contra de cada una de las ideas de la lista y el secretario toma nota de estas razones de forma ordenada. Agotadas las posibilidades de la primera idea se pasa a la segunda y sucesivas hasta finalizar. En la recapitulación subsiguiente es cuando se logra un mayor o menor grado de consenso sobre el resultado de la sesión.

Respecto al **método Delphi**, diremos que sigue un procedimiento análogo al brainstorming pero por escrito y por correo. Existe un equipo director y unos expertos diseminados. El equipo director elabora unos cuestionarios que envía a los expertos y analiza, sintetiza y presenta de forma ordenada las respuestas. Esta síntesis es enviada a los expertos junto con el siguiente cuestionario. En total suelen elaborarse unos cuatro cuestionarios, los dos primeros se corresponden con la primera fase del Brainstorming y los dos últimos con la segunda. El primer cuestionario presenta el problema y solicita ideas en formato muy breve (1-3 palabras). El segundo presenta una fusión y depuración de las respuestas del primero y solicita nuevas ideas. El tercero presenta la fusión y depuración de las ideas del segundo y solicita pros y contras de cada una de ellas. El cuarto funde y depura las respuestas del tercero y solicita nuevos pros y contras. Para detalles sobre cómo elaborarlos cuestionarios, consideraciones psicológicas, etc. véase la bibliografía especializada (por ejemplo, Listone y Turoff, 1975). Las características más importantes del método son las siguientes:

– Obtiene respuestas simultáneas y anónimas. Ello elimina los dos primeros inconvenientes del grupo de discusión tradicional.

– Permite la interacción entre los miembros del grupo y el retorno a cada miembro de las conclusiones parciales del grupo a pesar de la separación. La separación de los miembros del grupo consigue eliminar el diálogo irrelevante y los temas de interés personal (el "ruido"), y también otros problemas psicológicos derivados del contacto entre las personas.

– El hecho de que cada respuesta a cada pregunta se tenga de modo aislado permite la interpretación estadística de los resultados.

Ejemplo:

Para el problema de las dessaladoras, a través de un brainstorming, se pidió al grupo en la primera fase los “factores implicados en el problema”. En la segunda fase se trató de depurar la lista obtenida y obtener de ella una lista de variables con su unidad de medida (explícita o implícita). Conseguimos la siguiente lista:

- Necesidades de la población (litros/habitante y día)
- Población por zonas (matriz)
- Precipitaciones por zonas (matriz)
- Mermas (% de pérdidas por kilómetro) (zonas, matriz)
- Superficie agraria (zonas, matriz)
- Necesidades agua agraria (m^3/habitante y año)
- Política: votos (zonas, matriz)
- Flujos entre cuencas o entre depósitos:
    - Flujo que circula
    - Casilla de salida
    - Casilla de llegada
    - Capacidad de la conducción.
- Cuencas hidrográficas (matriz, número de cuenca en casillas)
- Capacidad dessaladoras (matriz, capacidad en casilla)
- Ubicación de los depósitos (matriz, capacidad en casilla)
- Caudales ecológicos por cuencas (9 cuencas)
- Rentabilidades de agricultura, industria y turismo (zonas, matriz)
- Necesidades industria
- Capacidad embalses
- Capacidad conducciones
- Acuíferos
- Superficie de bosque y montes (zonas, matriz)

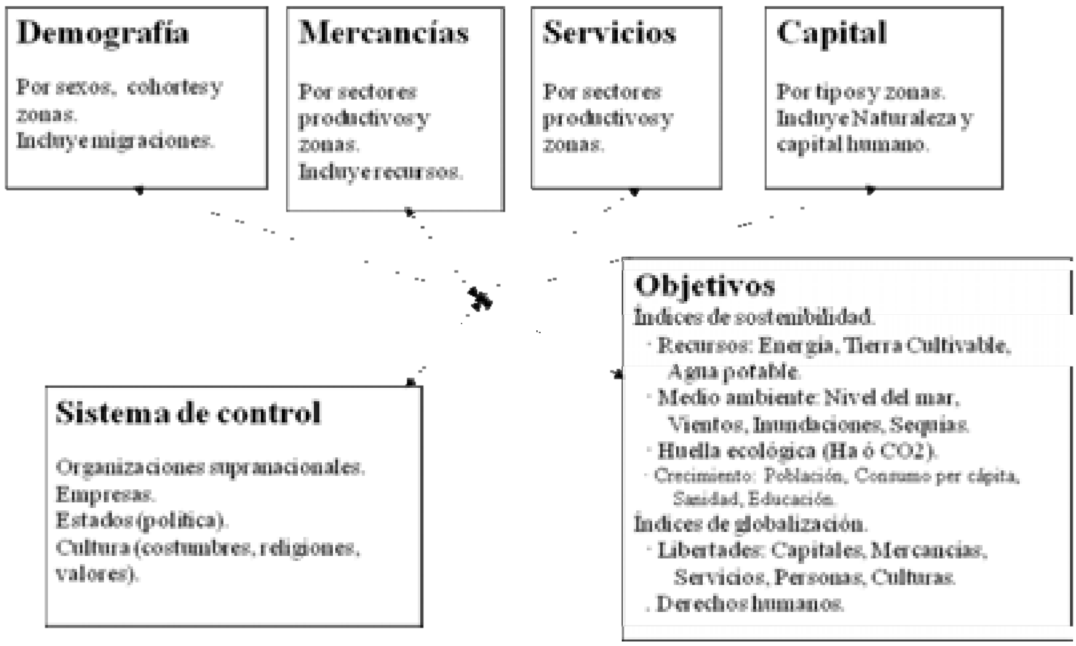
2.3.2.2 El blanqueo de cajas negras.

Entendemos por caja negra un sistema del que solo se conoce su relación con el entorno pero no su estructura y comportamiento interno. Por consiguiente, de él lo único que sabemos son las entradas y las salidas.

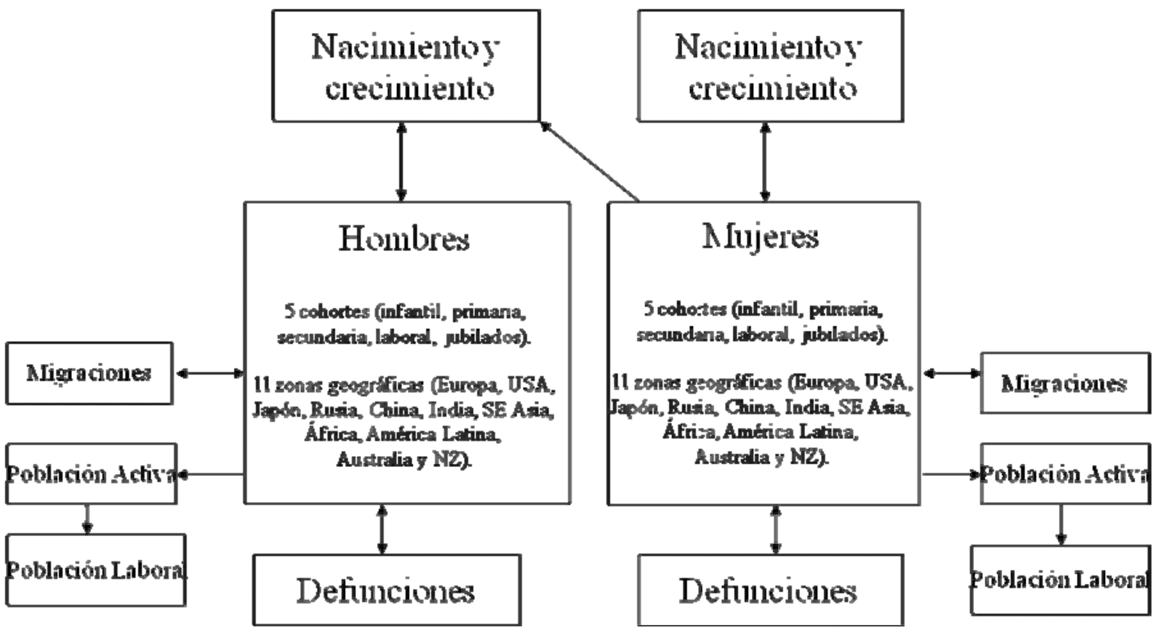
Si tratamos de blanquear una caja negra estaremos tratando de identificar otras cajas negras en su interior y las relaciones entre las mismas y con el entorno.

Ejemplo:

Para el problema de la globalización, que considera el mundo como una caja negra sin relaciones con el entorno (tendrían que ser interplanetarias), tendíamos como primer nivel de desagregación:



Y, como segundo nivel de desagregación, cada uno de estos bloques daría lugar a un nuevo gráfico. Por ejemplo el bloque demográfico se podría desagregar de la siguiente manera:



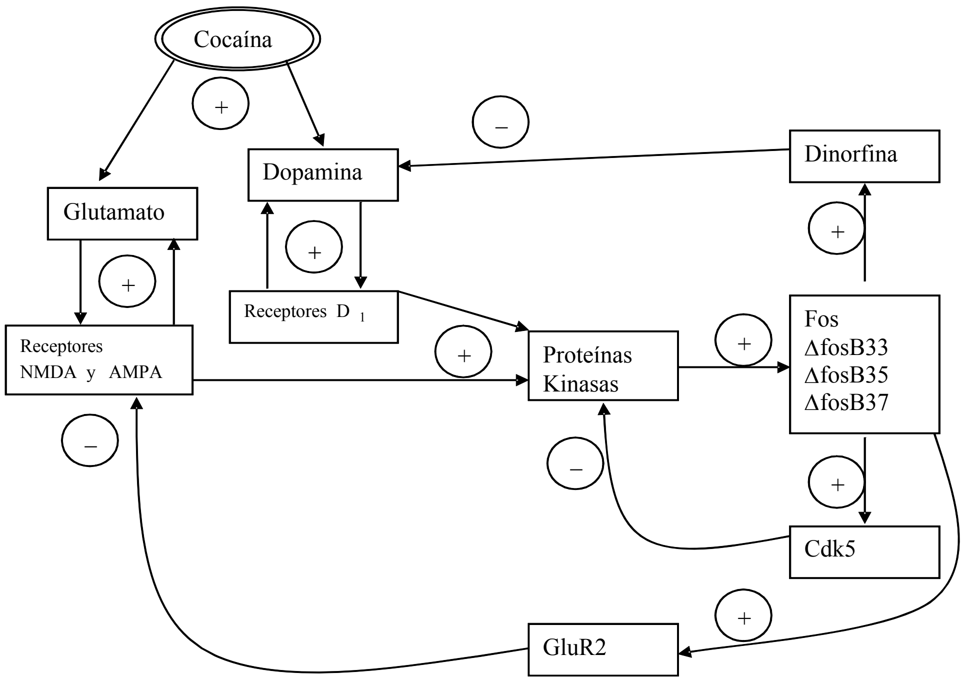
Y así sucesivamente.

2.3.3 Identificación de las conexiones causa-efecto entre las variables o elementos.

La misión de este paso es construir una representación gráfica de la estructura del modelo (todavía no entramos en el estudio del comportamiento). Los elementos serán los vértices del grafo y las relaciones causa-efecto serán las flechas del mismo. Un aditamento interesante a este diagrama son los signos + y – escritos junto a las flechas, indicando que se trata de una influencia positiva o negativa. Haciendo esto construiremos el **diagrama causal**.

Ejemplo de diagrama causal

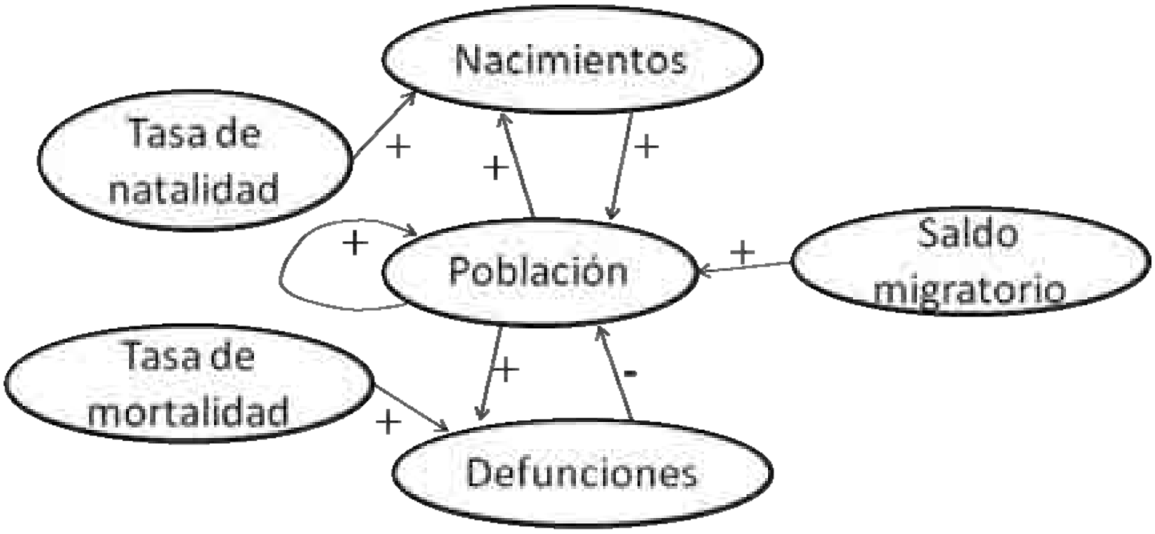
(Diagrama causal del modelo REGEN de regulación génica de los efectos de la cocaína. A. Caselles, S. amigó, J.C. Micó, 2006, pag 1 -14. Revista Internacional de Sistemas)



© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Demografía elemental

Diagrama causal



Antonio Caselles. Universitat de València. España

Del diagrama causal se deduce la matriz de conexiones, ambos definidores de la estructura del sistema. La siguiente es la matriz de conexiones correspondiente a “demografía elemental” (en ella la variable que encabeza una fila recibe la influencia de la variable que encabeza una columna, pero podría haberse escrito al revés).

	Población	Nacimientos	Defunciones	Tasa de Natalidad	Tasa de Mortalidad
Población	x	X	x	-	-
Nacimientos	x	-	-	x	-
Defunciones	x	-	-		x
Saldo migratorio	-	-	-	-	-
Tasa de Natalidad	-	-	-	-	-
Tasa de Mortalidad	-	-	-	-	-

Otro tipo de diagrama que interesa construir es el diagrama hidrodinámico o diagrama de Forrester (Forrester, 1961). Con este fin, lo primero que tenemos que hacer es identificar los siguientes tipos de variables en la lista elaborada en el paso anterior.

– Variables de estado. Son las que necesitan un valor inicial porque influyen sobre sí mismas. Estos valores iniciales son datos del problema o variables de entrada del sistema y, por consiguiente necesitan un nombre a parte. Es decir, cada variable de estado está asociada con otra variable (variable de entrada) que representa su valor inicial o previo. Conviene, por resultar más intuitivo, representar en el gráfico ambas variables, la de estado y su valor inicial, en la misma casilla. El colocar o no una flecha que indique la influencia de la casilla sobre ella misma es innecesario por obvio. Se suele usar para este tipo de variables como icono un recuadro.

- Variables de flujo o tasas. Normalmente, pero no siempre, las variables de estado son comparables a las variables "de nivel" de la metodología de Forrester (1961 y 1966) o a las "colas" en la metodología DEVS (Zeigler, 1984 etc.). Cada nivel o cola representa algo que se acumula, por consiguiente está asociada con variables que aportan ese algo y con variables que lo extraen. Estas son las que llamamos variables de flujo (las tasas, "rates" de Forrester). Una vez que las variables de estado han sido identificadas, y transportadas al campo del gráfico encerradas en un recuadro, es fácil identificar en la lista las variables de flujo relacionadas con cada una de ellas y transportarlas al grafo colocándolas en su proximidad encerradas en un recuadro unido a un aspa, que es el icono tradicional para este tipo de variables.

- El resto de las variables de la lista. Estas variables se transportan al grafo encerradas en una elipse y serán variables auxiliares o variables de entrada. Las variables de entrada, tanto si son constantes como variables exógenas o variables de control conviene distinguirlas con una doble línea en su elipse o símbolo de flujo (puede haber flujos que sean variables de entrada, es decir, datos del problema).

- En ocasiones interesa considerar un nuevo tipo de variables llamadas “retrasos” de las que trataremos más adelante.

La descripción del problema, el sentido común, u otros métodos de los que hablaremos más adelante, nos permitirán colocar flechas entre los símbolos, que expresen: (a) movimiento de algo comparable con un fluido (la población en un modelo demográfico, los materiales en un modelo de inventario, etc.) y (b) conexiones de pura influencia o causalidad, es decir, de información en general. Se suelen utilizar flechas continuas en el caso (a) y flechas discontinuas en el caso (b). No obstante, en ocasiones puede ser interesante utilizar flechas de más tipos (continuas gruesas para equipos, continuas dobles para mano de obra, etc.).

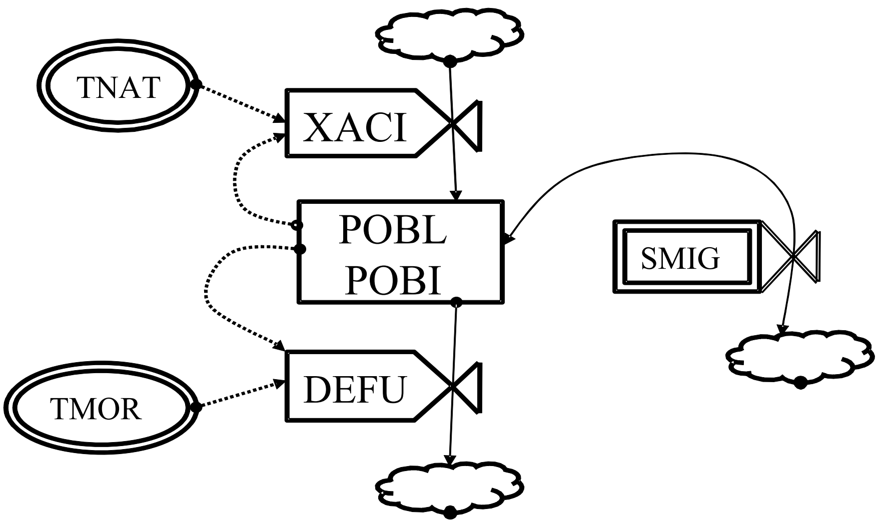
Ejemplo:

Demografía elemental
Lista de variables

- POBI Población a principio de año
- POBL Población a final de año
- XACI Nacimientos
- DEFU Defunciones
- TNAT Tasa de natalidad
- TMOR Tasa de mortalidad
- SMIG Saldo migratorio

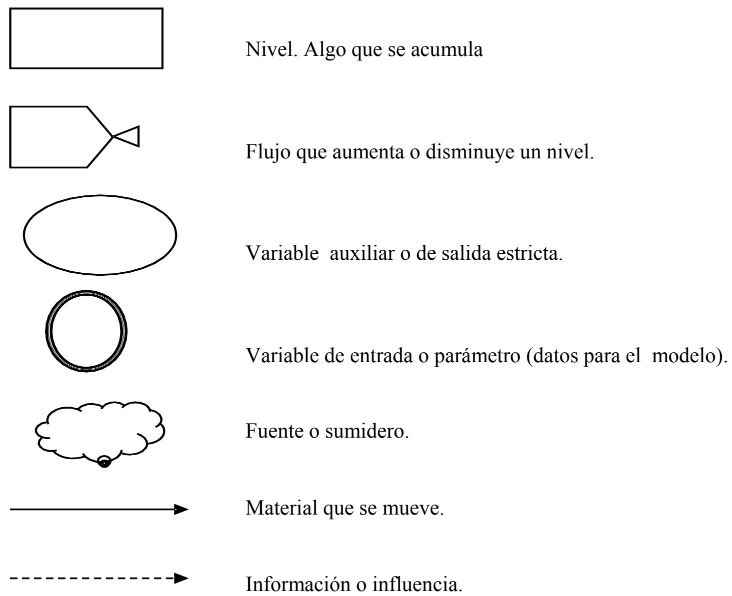
Demografía elemental

Diagrama hidrodinámico



© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Interpretación de los iconos del diagrama hidrodimámico o diagrama de Forrester

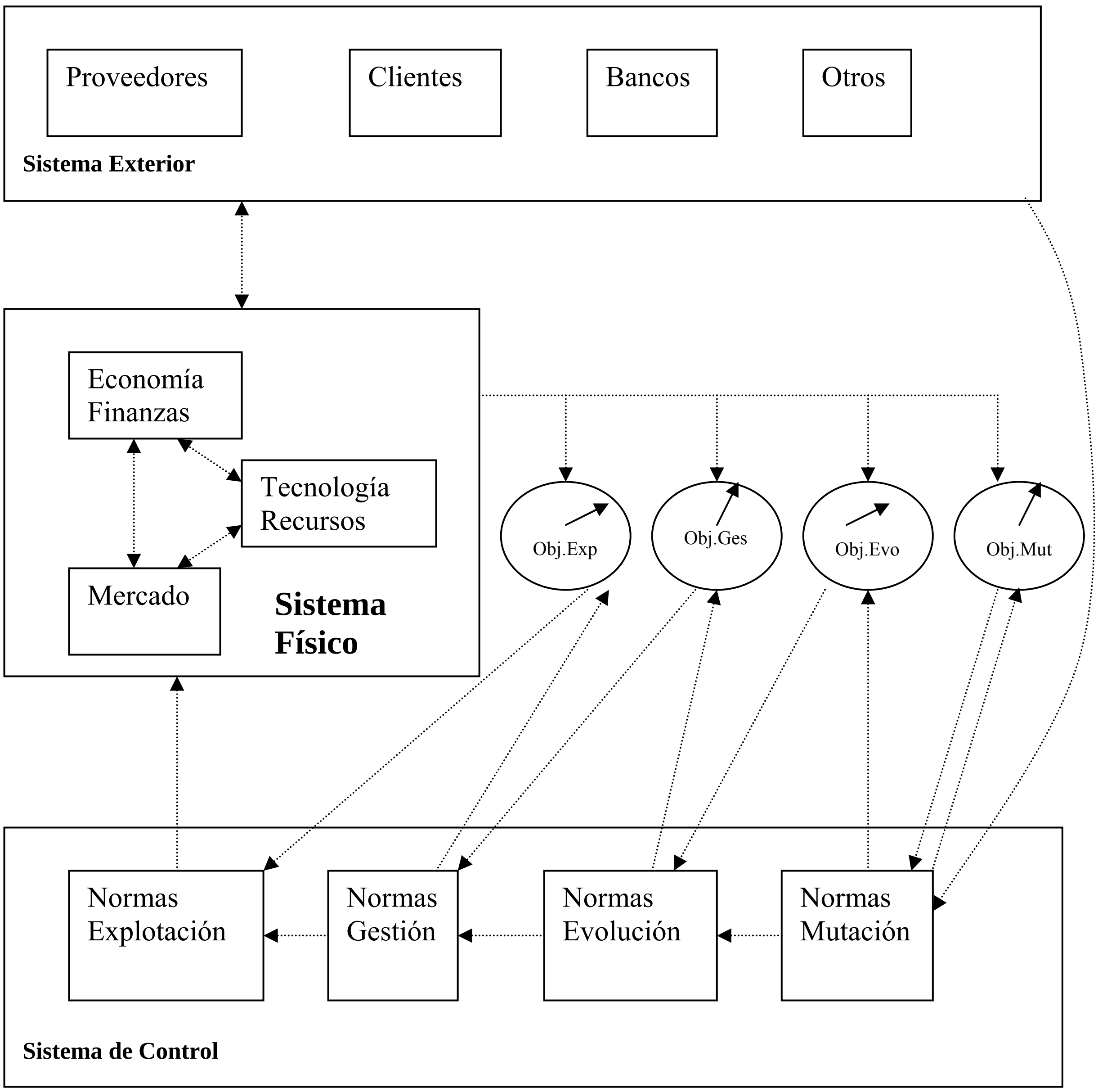


En relación con los métodos para detectar las conexiones entre elementos o variables diremos que cuando se construyen modelos de tipo mental o modelos para computadora de tipo deductivo, dependiendo de la complejidad del modelo y de los medios disponibles, se puede utilizar el sentido común, la consulta a expertos, el Brainstorminsg, el Delphi o similares. Cuando se construyen modelos para computadora de tipo inductivo (a partir de datos reales) o mixto, con variables numéricas se pueden utilizar métodos llamados de "análisis causal" (*) que pueden encontrarse por ejemplo en Bollen (1989).

En ocasiones, especialmente cuando el sistema es muy complejo, puede ser interesante en esta fase construir además del descrito otros tipos de diagramas pues ello ayuda a entender y a precisar la estructura del sistema. Concretamente uno a varios diagramas de subsistemas. Los subsistemas, como veremos más adelante, los define el usuario a su conveniencia, pero siempre existe uno o más conjuntos de subsistemas que resultan más intuitivos, y pueden en esta fase ser representados como elementos en un grafo de conexiones similar al descrito anteriormente, donde también pueden figurar, si se estima oportuno, las variables comunes representadas por cada flecha. Como ejemplo significativo de este tipo de diagramas está el diagrama de Ashby descrito por Melèse (1976), especialmente indicado para los sistemas con objetivos o sistemas “cibernéticos”. En el diagrama del “sistema ultraestable” de Ashby se contemplan cuatro niveles de control anidados y está pensado para las empresas, y cualquier otro tipo de organización. Los sistemas vivos son un caso particular de sistemas con objetivos y para ellos Miller (1978) desarrolló toda una teoría. Mas detalles sobre esta teoría y los sistemas con objetivos pueden encontrarse en el Apéndice 1.

Diagrama del sistema ultraestable de Ashby (para sistemas con objetivos) (Aplicado al caso de una empresa)

Nivel de Control	Conoce:	Determina:	Decide a plazo:	Su actividad se llama:
Mutación	Universo Exterior	Objetivos	Muy largo	Política
Evolución	Objetivos	Medios	Largo	Estrategia
Gestión	Objetivos y Medios	Procedimientos	Medio	Táctica
Explotación	Objetivos, Medios y Procedimientos	Correcta ejecución	Muy corto o inmediato	Ejecución



2.3.4 Representación funcional de las relaciones.

La representación gráfica de la estructura del modelo muestra, por ejemplo, que una variable o elemento A depende de B, C y D. Esto significa que para obtener un valor de A debemos conocer los valores de B, C y D así como el mecanismo que los liga, es decir la función A=f(B,C,D), si se utiliza el lenguaje matemático. Esta función describe el comportamiento de la variable A y puede ser representada por una lista de reglas, por una tabla, una ecuación o una lista de reglas y ecuaciones.

La forma de esta función en cada caso particular puede determinarse: por el sentido común (tautologías), por consulta a expertos individuales o en grupo (Brainstorming o Delphi), usando los métodos de la Regresión (para funciones de tipo numérico en modelos inductivos o mixtos), o bien por el método de prueba y error planteando formas hipotéticas y probándolas con datos reales (esto formaría parte del "calibrado del modelo").

A continuación se detalla las funciones más comunes que nos pueden relacionar las variables en un sistema complejo.

a) **Tautologías**: igualdades evidentes.
Ejemplos: 4 = 2 +2. Lo que hay = lo que había + lo entrado – lo salido

b) **Reglas lógicas**: Si "esto" entonces "lo otro" y si no "otra cosa".
Que, en lenguaje de programación, sería: If "…." then "…….." else "………."

c) **Tablas:** la función $Y=f(X_1, X_1, \ldots, X_n)$ podría venir dada por una tabla parecida a esta:

X_1	X_2	…	X_n	Y
3	-1.1		0.08	24.1
5	-0.4		0.10	26.4
8	0		0.16	29.3
11	2.7		0.21	29.8
15	3.1		0.29	33.1

d) **Ecuaciones diferenciales:** una ecuación diferencial ordinaria de orden n es una función de este tipo: $f(x, y, y', y'', \ldots, y^{(n)}) = 0$, donde x es la variable independiente y $y, y', y'', \ldots, y^{(n)}$ son las derivadas sucesivas de y respecto de x.

Una ecuación diferencial de primer orden solo tendría la derivada primera $y'=dy/dx$. Y si la variable independiente fuese el tiempo y la ecuación incluyese otras variables a las que llamaríamos $x_1, x_2, \ldots x_n$, podríamos escribir:

$$dy/dt = f(x_1, x_2, \ldots x_n, y, t) \qquad (1)$$

De una ecuación diferencial nos suele interesar una "integral particular", es decir, una función $y = F(x_1, x_2, \ldots x_n, t)$ cuya derivada es $f(x_1, x_2, \ldots x_n, y, t)$ y que además cumple otra condición. El método numérico más sencillo para obtener los puntos de esta F es el

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

método de Euler que, en síntesis, consiste en considerar los diferenciales *dy* y *dt* como incrementos finitos Δy e Δt. Con ello la ecuación diferencial se transforma en una ecuación en diferencias finitas, que se integra como vemos a continuación.

e) **Ecuaciones en diferencias finitas:** la ecuación (1) escrita en diferencias finitas quedaría:

$$\Delta y / \Delta t = f(x_1, x_2, \ldots x_n, y(t-\Delta t), t)$$

O bien:

$$\Delta y = f(x_1, x_2, \ldots x_n, y(t-\Delta t), t) \cdot \Delta t$$

De dónde:

$$y(t) = y(t-\Delta t) + f(x_1, x_2, \ldots x_n, y(t-\Delta t), t) \cdot \Delta t$$

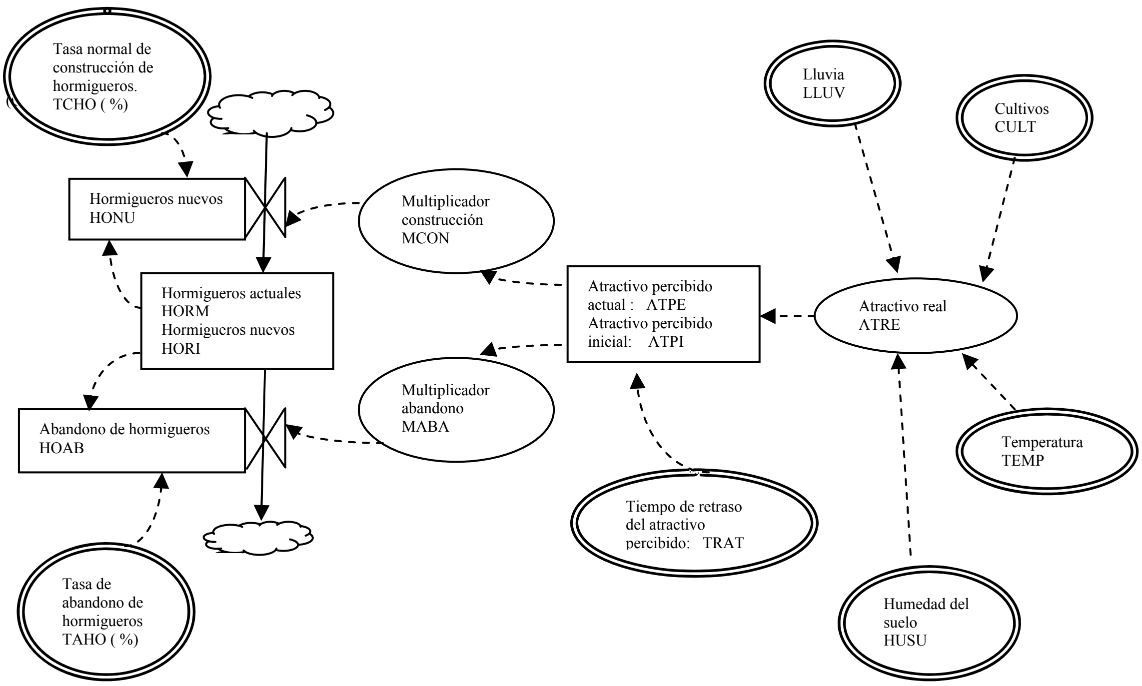
En el Apéndice 4 encontraremos más detalles y ejercicios sobre integración de ecuaciones diferenciales ordinarias..

f) **Retrasos:** en ocasiones, algunas variables influyen con cierto retraso temporal respecto a otras. Los retrasos se suelen formular de la siguiente manera:

$$V_{retrasada}(t) = V_{retrasada}(t-\Delta t) + (V_{original}(t) - V_{retrasada}(t-\Delta t)) \cdot \Delta t / T_{retraso}$$

Es decir, se asume que la diferencia entre el valor real de la variable *V* en el instante *t*, al que hemos llamado $V_{original}(t)$, y el valor de la misma variable percibida con retraso en el instante anterior, $V_{retrasada}(t-\Delta t)$, tarda un tiempo en percibirse, $T_{retraso}$, luego en el intervalo Δt se percibirá la fracción $\Delta t / T_{retraso}$ de esa diferencia.

Ejemplo: en una zona con hormigueros, las hormigas pueden colonizar o abandonar la zona. Hay atractivos de la zona que pueden tardar en percibirse. Sobre el atractivo puede influir la lluvia, nuevas fuentes de alimento, etc. Estúdiese el diagrama siguiente:



La lista de variables y la lista de funciones serían:

HORM cantidad de hormigueros
HORI cantidad inicial de hormigueros
HONU nuevos hormigueros
LLUV cantidad de lluvia media de a zona
CULT cantidad de cultivos (%)
TEMP temperatura media de a zona
HUSU humedad media del suelo en la zona
HOAB abandono de hormigueros
ATPE atractivo de la zona percibido por las hormigas
ATPI atractivo de la zona percibido inicialmente por las hormigas
TCHO tasa normal de construcción de hormigueros
TAHO tasa normal de abandono de hormigueros
MCON multiplicador para la construcción de hormigueros
MABA multiplicador para el abandono de hormigueros
TRAT tiempo de retraso en la percepción del atractivo de la zona por las hormigas
ATRE atractivo real de la zona para las hormigas

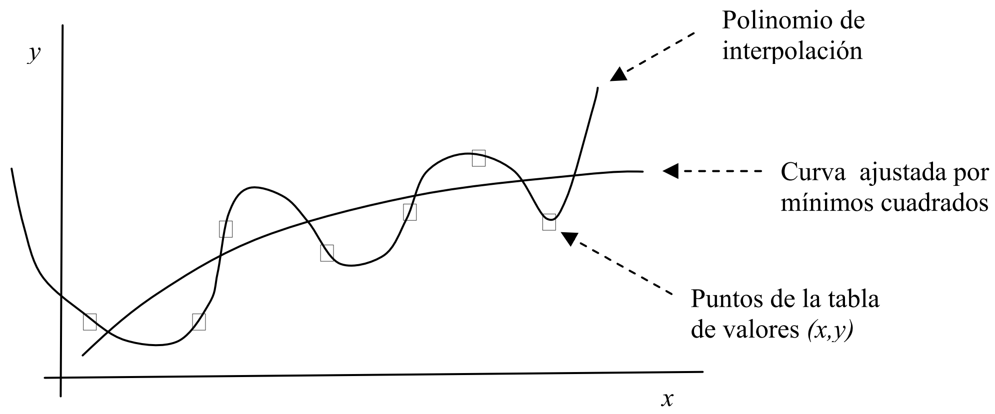
HONU	honu = hori * tcho/100 * mcon	
HOAB	hoab = hori * taho/100 * maba	
HORM	horm = hori + honu – hoab	
ATPE	atpe = atpi + 1/trat * (atre-atpi)	
ATRE	atre = f(lluv, cult, temp, husu)	‘f obtenida por ajuste
MCON	mcon = tabla(atpi)	‘tabla dada por expertos y calibrada
MABA	maba = tabla(atpi)	‘tabla dada por expertos y calibrada

g) **Regresión múltiple y ajuste por mínimos cuadrados:** si se trata de una función de varias variables, $y = f(x_1, x_2, \ldots x_n)$, dada por una tabla de valores, la regresión permite obtener una curva que se ajuste a esos puntos por el método de mínimos cuadrados. La curva no pasará, probablemente, por ninguno de esos puntos pero su distancia a ellos será la menor posible. Este tipo de funciones puede ser lineal o no lineal. En este último caso es necesario un algoritmo buscador de funciones que nos encuentre la mejor entre un abanico de funciones posibles, asumiendo que es una variable aleatoria normal cuya media calculamos con la función ajustada y cuya desviación típica también se puede calcular. Ese buscador podría ser REGINT (Caselles, 1998). Con la función ajustada obtendremos el valor medio y REGINT nos proporciona también los datos para calcular la desviación típica. En los modelos deterministas haremos uso solamente de la función ajustada pero en los modelos estocásticos será necesario calcular también la desviación típica (Caselles, 1992a). En el Apéndice 4 encontraremos más detalles y ejercicios sobre ajuste por mínimos cuadrados.

h) **Interpolación:** cuando una función de varias variables dada por una tabla de valores tiene que pasar por todos los puntos de la tabla, es decir, no hay ruido ninguno en las medidas, no vale el juste, es necesaria la interpolación. La interpolación más frecuente es la interpolación polinómica. En ella, la función desconocida $y = f(x_1, x_2, \ldots x_n)$ es

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

sustituida por un polinomio de grado *n-1*. El polinomio de interpolación, generalmente, no puede usarse para extrapolar, es decir, para valores de las variables independientes que estén fuera del rango de valores usado en la tabla original dado que fuera de rango los polinomios tienen formas imprevisibles. En el Apéndice 4 encontraremos más detalles y ejercicios sobre interpolación.



i) En ausencia de datos históricos no queda más remedio que utilizar métodos que aprovechen la información existente en las mentes de los expertos en el tema. El método más interesante bajo nuestro punto de vista **es** el de los impactos cruzados (cross-impact). Este y otros métodos son de los llamados tradicionalmente “métodos prospectivos” y se detallan en el Apéndice 3.

En el siguiente ejemplo, que concuerda con otros anteriores sobre demografía elemental, todas las ecuaciones son tautologías (relaciones evidentes).

Ejemplo:

Demografía elemental

Lista de funciones (ecuaciones, tablas o reglas lógicas)

- XACI xaci=pobi*tnat/1000
- DEFU defu=pobi*tmor/1000
- POBL pobl=pobi+xaci -defu+smig

2.3.5 Programación

Tenemos la lista de los elementos del sistema, su estructura (las conexiones entre los mismos) y su comportamiento (representado por las relaciones funcionales), es decir, ya tenemos la esencia del modelo. Ahora queda hacerlo funcionar y para ello necesita un soporte dinámico.

Si es un modelo mental o tipo "soft", debe estar en las mentes del equipo que lo ha elaborado además de en el papel. El proceso de elaboración ha servido para introducirlo en sus mentes y en ellas debe funcionar, así que, teóricamente, el equipo ya está en condiciones de contestar con conocimiento de causa a preguntas tipo "que pasaría si...", referentes al sistema. Como métodos adecuados para llegar a responder a estas preguntas están, como tantas veces, el Brainstorming y el Delphi.

Si es un modelo para computadora o tipo "hard", es necesario transformarlo en un programa de ordenador. Para elaborar este programa existen multitud de herramientas de las que ya hemos tenido ocasión de hablar:

- Lenguajes de programación de tipo general (C, PASCAL, BASIC, etc.).
- Lenguajes de simulación (GPSS, SIMSCRIPT, SIMULA, QNAP2, STIMS, etc.)
- Cajas de herramientas (MATLAB, MATHEMATICA, etc.)
- Interpretadores de descripciones (Inter-SIM, HOCUS, STELA, VENSIM, etc.)
- Generadores de aplicaciones (DRAFT, eLSE, GASPE, SIGEM, etc.)

Por sus ventajas en cuanto a generalidad (número de tipos de modelos que pueden ser construidos), y facilidad de uso vamos a describir con algo más de detalle el generador de aplicaciones SIGEM.

El generador de aplicaciones SIGEM recaba del usuario la información correspondiente a la lista de nombres de los elementos del sistema, a las características de los mismos y a las relaciones funcionales entre los mismos. Esta información se puede introducir mediante especificaciones en forma de ficheros de texto o mediante un diálogo interactivo. El diálogo interactivo puede ser útil en modelos de pequeño tamaño o para principiantes. La información concreta incluida en este dialogo o en las especificaciones es la siguiente:

- Información sobre las variables de entrada, una a una: si son literales o numéricas, si son constantes o si varían, si llevan incertidumbre o no, y si tienen dimensiones o no (si son escalares o vectores, matrices, etc.)). Las variables con incertidumbre pueden entrar en el modelo generado, bien como una pareja valor-medio/desviación-típica (en el caso de que se consideren normalmente distribuidas) o bien como una tabla de frecuencias (en el caso general), pero esto no es necesario especificarlo porque siempre son posibles las dos opciones.

- Información sobre las variables de salida y las funciones que permiten determinar su valor, una a una: si son literales o numéricas, si son dimensionadas (matrices con uno o varios subíndices) o no, si llevan incertidumbre o no, si son de estado (necesitan un valor inicial) o no, y si la función viene dada por una tabla o por un conjunto de ecuaciones y/o reglas lógicas. Las funciones que de definen de manera explícita como con incertidumbre se suponen correspondientes a distribuciones normales multivariadas, debiendo introducirse al menos dos ecuaciones, una que permita calcular el valor medio y otra la desviación típica correspondiente. Para más detalles véase el trabajo de Caselles (1992a) y algún libro sobre Regresión múltiple, lineal y no lineal.

Para usuarios no principiantes la inducción de la mencionada información se traduce en la creación de dos ficheros de texto denominados "L*nombre*.txt" y "G*nombre*.txt" donde "*nombre*" es el nombre que nosotros le queramos dar a muestro modelo. El fichero L incluye la lista de variables con sus especificaciones y el fichero G la lista de funciones (sin necesidad de que estén ordenadas), ambos en el formato que se detalla en el ejemplo que aparece a continuación.

Ejemplo:

Contenido de los ficheros "Ldemo1.txt" y "Gdemo1.txt" correspondientes al modelo que hemos llamado "demografía elemental":

Fichero "Ldemo1.txt":

```
POBI Población a principio de año [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
POBL Población a final de año [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=s;inic=POBI;tabl=n;npun=..;]
XACI Nacimientos [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
DEFU Defunciones [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TNAT Tasa de natalidad [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TMOR Tasa de mortalidad [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
SMIG Saldo migratorio [ince=n;cval=00;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
```

Obsérvese que:

- El nombre completo va precedido del nombre codificado (que es el que se usará en los programas). Debe tener siempre el mismo número de caracteres. Las mayúsculas no son obligatorias pero son convenientes para distinguir bien el nombre codificado del nombre completo.
- Las especificaciones de cada variable van al final entre corchetes. Cada especificación se detalla con un código de 4 caracteres es minúsculas, un signo =, y una letra (s ó n) o un número seguido de un signo ";". Así, *ince=n* significa que la variable correspondiente no lleva incertidumbre (no es una variable aleatoria), *cval=00* significa que, si es variable de entrada, no lleva cambios de valor (es constante), *ndim=0* significa que no tiene dimensiones (es escalar), "*d1=.*" significa que su primera dimensión no existe, si existiese habría un número en lugar del ".", *esta=n* significa que la variable no es de estado, *inic=POBI* significa a que el valor inicial de la variable POBL es POBI, *tabl=n* significa que la función no viene dada por una tabla, "*npun=..*" significa que el número de puntos de la posible tabla no existe (si existiese podríamos un número en lugar de ..).

Fichero "Gdemo1.txt":

```
XACI xaci=pobi*tnat/1000
DEFU defu=pobi*tmor/1000
POBL pobl=pobi+xaci-defu+smig
```

Obsérvese que:

- Cada ecuación va precedida por el nombre codificado de la variable que se calcula con ella y un espacio en blanco. Si una función, en lugar de venir determinada por una ecuación o línea de código, necesitase más líneas de código, las siguientes líneas de código deben ir precedidas por espacios en blanco (no tabuladores) hasta que la sangría sea de 6 o más espacios (caso de tener 4 caracteres los nombres codificados de las variables).
- Las funciones no necesitan estar ordenadas, SIGEM las ordenará adecuadamente.

- Las funciones suelen venir dadas por ecuaciones pero muchas veces son pequeños algoritmos que incluyen ecuaciones y/o reglas lógicas y a veces bucles de repetición. Todo ello necesita ser escrito en Visual Basic 6. Las nociones mínimas necesarias de este lenguaje se encuentran en el Apéndice 2.

Otra facilidad que ofrece SIGEM al usuario es la posibilidad de interrumpir el trabajo en cualquier momento y reanudarlo desde el punto en que quedó interrumpido. Para ello crea un fichero de texto denominado *modelo*.txt* donde se guarda el dialogo previo (explícito o implícito). En lugar de * escribiremos un número entre 0 y 9.

Ejemplo:

Contenido del fichero "modelo1.txt" correspondiente al modelo *demo1*.

```
"Crear, Ensamblar, formar Bloques, Analogía","c"
"Nombre del simulador","DEMO1"
"Servira para","demografia elemental"
"Lista al completo","s"
"Ver la lista por Pantalla, Fichero, o No ver","n"
"Fichero G si y fichero C no","s"
"Ver la matriz por Pantalla, Impresora, o No ver","n"
"Listado en orden de calculo","s"
"Ordenar fichero de especificaciones: ","n"
"¿Sistema dinamico?","s"
"  Unidad dinamica","año"
"  Algunas, Todas o Ninguna variables de entrada con incertidumbre: ","n"
"POBI Población a principio de año"
"   Cambios de valor",0
"   ¿Dimensional?","n"
"   ¿Correcto?",""
"TNAT Tasa de natalidad"
"   Cambios de valor",4
"   ¿Dimensional?","n"
"   ¿Correcto?",""
"TMOR Tasa de mortalidad"
"   Cambios de valor",4
"   ¿Dimensional?","n"
"   ¿Correcto?",""
"SMIG Saldo migratorio"
"   Cambios de valor",4
"   ¿Dimensional?","n"
"   ¿Correcto?",""
"¿Funciones grabadas?","s"
"XACI Nacimientos"
"   ¿De estado?","n"
"   ¿Dimensional?","n"
"   ¿Tabla?","n"
```

" ¿Correcto?",""

"DEFU Defunciones"

" ¿De estado?","n"

" ¿Dimensional?","n"

" ¿Tabla?","n"

" ¿Correcto?",""

"POBL Población a final de año"

" ¿De estado?","s"

" Su valor inicial es","POBI"

" ¿Dimensional?","n"

" ¿Tabla?","n"

" ¿Correcto?",""

El menú inicial de SIGEM tiene cuatro opciones: Crear un modelo muevo, dividir un modelo en submodelos ensamblables, ensamblar submodelos, y búsqueda analógica.

La opción *Crear*, con cada modelo, elabora tres programas Visual Basic 6: el gestor de datos, el simulador y el productor de informes. Más adelante hablaremos sobre cómo trabajan estos programas.

La opción *Dividir* permite al usuario jerarquizar las variables del modelo (construir el árbol de dependencias), y construir estructuras de subsistemas también jerárquicas, para que sean posteriormente ensamblables. En esta opción se pregunta al usuario sobre los límites entre los que desearía que estuviese el tamaño de los subsistemas, o construye unos llamados "subsistemas naturales" cuando el tamaño no importa demasiado. Una vez aceptados por el usuario los subsistemas, elabora la lista de nombres, la matriz de conexiones, además del sistema de subsistemas con su lista de nombres, su matriz de conexiones, las variables comunes de cada conexión, y otros detalles, y lo deja todo preparado para que cuando los subsistemas hayan sido validados de manera independiente, sus correspondientes programas puedan ser fácilmente ensamblados. Para más detalles véase el artículo de Caselles (1993a).

La opción *Ensamblar* solo necesita del usuario el nombre del sistema de subsistemas a ensamblar. Obviamente cuando se intente ensamblar un sistema de subsistemas todos los módulos de todos los programas de los subsistemas deben estar presentes en la unidad implícita del ordenador, así como los ficheros elaborados con la opción *Dividir*. En caso de no estar presentes estos últimos también pueden introducirse sus datos por medio de un diálogo.

Las ventajas más importantes de la descomposición de un sistema en subsistemas para el posterior acoplamiento de los mismos reconstruyendo el sistema global son casi obvias:

– un sistema de muchas variables es difícil y tedioso de construir, de verificar y de validar, y las oportunidades de error son mayores;

– cuando tiene que intervenir un equipo grande de personas en la elaboración de un modelo es necesario un criterio adecuado para repartir el trabajo y posteriormente unir los resultados del trabajo individual.

No obstante la utilidad de dividir y posteriormente ensamblar, existe otra posibilidad para lograr el mismo fin: unir en un solo fichero todos los ficheros L y en otro fichero todos los ficheros G y optar por crear un modelo nuevo. Cuando no existen problemas de memoria (lo normal en estos tiempos), esta segunda opción es preferible.

2.3.6 Análisis y gestión de datos.

Para construir un modelo muevo, en principio, se requieren tres tipos de información:

(a) información sobre los objetivos y la estructura del modelo (elementos que intervienen y sus conexiones);

(b) información sobre el comportamiento de las variables (relaciones funcionales);

(c) información sobre los experimentos, optimizaciones o pruebas a realizar con el modelo (que se derivan de los objetivos inicialmente propuestos y del tipo de modelo construido).

Las informaciones (a) y (b) ya hemos dicho de qué tipo son y cómo se obtienen y elaboran. La información de tipo (c) será estudiada cuando hablemos del diseño de experimentos.

Las cuestiones de tipo general sobre adquisición de datos, análisis y gestión de los mismos son de gran interés dado que consumen una proporción considerable del tiempo del equipo modelizador y pueden llegar a paralizar y hasta hacer fallar el proyecto (con sorpresa para el equipo). Claro que esto ocurre si el equipo se empeña en validar un determinado modelo previamente construido. La alternativa es diseñar un nuevo modelo que use los datos existentes una vez detectados estos. Y estos datos adecuados solo podrán detectarse, normalmente, cuando ya se ha construido un modelo tentativo (recuérdese lo de la vuelta atrás que se explicó al tratar la metodología general de modelización).

No vamos entrar en temas de diseño de experimentos para determinar valores de los diferentes tipos de variables en las diferentes áreas de conocimiento, ni en métodos para ordenar y clasificar datos. Esto, obviamente, escapa de los objetivos de este trabajo. Solamente queremos hacer hincapié en que en los sistemas socio-económicos, y medioambientales, la escasez de datos y la imposibilidad de realizar una toma de los mismos en el tiempo que dura un estudio, hace fracasar muchos intentos. Por consiguiente, es necesario pensar en el tipo de datos que van a ser necesarios y en las posibilidades de localizarlos o determinarlos antes de comprometerse a realizar un trabajo determinado en unas condiciones determinadas.

Respecto a la gestión de datos (valores de las variables de entrada) en los modelos que se construyen, es necesario que exista un modulo específico para esta misión, que haga fácil la introducción inicial de los datos y la modificación de los mismos, y que pueda preparar varios conjuntos de datos alternativos de forma accesible para el módulo simulador.

2.3.7 Calibrado del modelo.

En determinados modelos, una vez construidos, quedan parámetros sin determinar y para determinarlos se recurre al método de prueba y error (véase por ejemplo Meyer et al., 1979). El ajuste de los parámetros desconocidos, en algunos tipos de relaciones e incluso de modelos dinámicos completos ya construidos, se puede realizar por el método de los mínimos cuadrados (véase 2.3.4). Otras veces se recurre a opiniones de expertos, que pueden ser posteriormente convertidas en valores numéricos. Incluso relaciones entre variables, presentadas en forma de tablas de valores, pueden ser obtenidas a través de opiniones de expertos (véase Forrester, 1970) y

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

después acopladas al modelo y rectificadas hasta que el modelo produzca resultados adecuados.

2.3.8 Verificación del modelo.

Este paso consiste en comprobar que el modelo (el programa de ordenador o modelo mental) da los resultados conocidos cuando trabaja con los correspondientes datos conocidos. Por datos y resultados conocidos entendemos los que han servido para calibrar o elaborar el modelo. Con este test se pueden detectar errores de programación, o cuando el programa ha sido producido por un generador de aplicaciones, se pueden detectar errores en las relaciones funcionales introducidas por el usuario. Para dar una idea clara de lo que es verificar un modelo podríamos decir que es: “comprobar que el resultado de 2 + 2 es efectivamente 4 y no otra cosa”.

2.3.9. Validación del modelo.

La idea intuitiva de validación consiste en que el modelo se construye para ser aplicable a una cierta generalidad o conjunto de sistemas, definidos por los objetivos propuestos en la primera etapa del proceso de modelización, pero se construye a partir de datos tomados de un pequeño subconjunto de tales sistemas, a veces de uno solo. Por tanto, es necesario asegurarse en cierta medida de que el modelo va a servir para cualquiera de los sistemas del conjunto correspondiente a los objetivos planteados. Caselles (1984 y 1992b) propone una aproximación difusa al concepto de similaridad entre sistemas que puede ser aplicada a la validación de modelos. Para dar una idea clara de lo que es validar un modelo podríamos decir que es: “comprobar que con el tipo de datos que tenemos vamos a obtener el tipo de resultados que deseamos de una manera fiable”. El método clásico de validación es el de “predicción del pasado” también llamado “simulación ex post”, que consiste en suministrar al modelo datos históricos conociendo también los correspondientes resultados históricos. Resultados estos que debe reproducir el modelo con una aproximación aceptable para el usuario (coeficiente de variación menor del 5%, por ejemplo). Este método asume que “en el futuro las cosas van funcionar como en el pasado”, es decir, que la estructura y comportamiento del modelo no va sufrir modificaciones. Si se conocen estas posibles modificaciones, ¿por qué no introducirlas en el modelo? Y, si no se conocen, no queda más remedio que actualizar el modelo cuando se presenten.

2.3.10 Gestión y manejo de modelos.

Esta etapa hace referencia a la posibilidad de guardar, modificar y rehusar modelos y partes de modelos. Cada metodología tiene unas capacidades específicas en este sentido. Por ejemplo, MATLAB/SIMULINK, STIMS y SIGEM, entre otros, tienen la capacidad de ensamblar modelos parciales. Un modelo parcial puede formar parte de varios modelos globales distintos. En cada metodología los modelos parciales se definen, se construyen, se guardan y se ensamblan de una forma distinta. La manera como lo hace SIGEM ha sido descrita someramente en 2.3.5. Otros paquetes requieren más nociones previas para poder aludir a este tipo de mecanismos de modo inteligible.

En los modelos construidos por SIGEM, pequeñas modificaciones correspondientes al comportamiento de las variables pueden realizarse directamente en el programa fuente del módulo simulador, y las modificaciones estructurales, es decir, las que afectan al número de variables que se consideran y a las conexiones entre las mismas, se realizan modificando la lista de elementos, la lista de ecuaciones/reglas/tablas, y regenerando el modelo a través de una nueva sesión de diálogo con SIGEM. Muchas modificaciones pequeñas pueden realizarse en el fichero donde se almacena el dialogo (*modelo*.txt*) con un procesador de textos y regenerando el modelo.

SIGEM dispone en el menú principal de la opción *búsqueda analógica.* Esta opción es capaz de explorar una biblioteca de modelos almacenados en forma compacta, es decir, con el fichero *Lnombre.txt*, el fichero *Gnombre.txt* y el fichero *Mnombre.txt* (que es el fichero *modelo*.txt* cambiado de nombre), con unas palabras clave proporcionadas por el usuario, y generar un nuevo modelo que será operativo pero que, obviamente, no tendrá ningún sentido, no obstante, podrá servir de base para la elaboración de un nuevo proyecto.

2.3.11 Diseño de experimentos y procedimientos de optimización.

Los modelos pueden ser diseñados de tal manera que incorporen procedimientos de optimización o que realicen procesos de búsqueda o toma de muestras, o que persigan un objetivo, pero lo más común es que el modelo tenga una o varias variables que tengan que alcanzar un valor máximo, un valor mínimo o, al menos, un valor aceptable. A estas las llamaremos "variables objetivo" o “variables esenciales”. Los valores de las variables de entrada que están en correspondencia con ese valor o valores óptimo o quasi-óptimo se suelen determinar realizando ensayos tipo prueba-y-error con el modelo. Hay dos clases de variables que deben ser tenidas en cuenta: las "variables de escenario" y las "variables de control".

– Llamamos variables de escenario o exógenas a las que son determinadas por factores desconocidos o imprevisibles, y respecto de las que únicamente podemos hacer hipótesis sobre sus posibles valores o conjuntos de valores. A las combinaciones de hipótesis sobre diferentes variables se les suele llamar "escenarios".

– Llamamos variables de control o de acción a aquellas cuyos valores pueden ser determinados por el usuario del modelo.

Por consiguiente, la situación más frecuente es aquella en la que el usuario desea encontrar los valores que más le convienen para las variables de control en cada uno de los escenarios más probables para las variables exógenas. Para lograrlo se suelen elaborar conjuntos de valores para las variables de entrada que combinan los diferentes escenarios con distintas opciones de control factibles a priori, y se rueda el modelo con cada uno de ellos. A la vista de los resultados en las variables objetivo se elige la opción idónea. Una buena manera de seleccionar una opción de control, que llamaremos estrategia, válida para cualquiera de los escenarios consiste en recabar de un grupo de expertos opiniones sobre las probabilidades de los diferentes escenarios construidos y, posteriormente, sumar los productos variable-objetivo por probabilidad en los diferentes escenarios para la misma estrategia, quedándonos con la estrategia que alcance el valor mayor como estrategia óptima.

En otros casos es necesario diseñar un experimento adecuado para un tipo de análisis estadístico determinado como puede ser una "prueba de hipótesis", "análisis de varianza", o similar.

Otra posibilidad es tomar una muestra aleatoria entre las combinaciones posibles de los valores posibles de las variables de entrada y obtener los valores correspondientes de la variable objetivo para, posteriormente, seleccionar la que maximice la variable objetivo.

Veámoslo de una forma más sistemática y formal:

Definición de optimización:

$$\left.\begin{array}{l} Y = f(x_1 , x_2 , \ldots , x_n) \quad \rightarrow \quad \text{Max ó min} \\ \left.\begin{array}{l} g_1 (x_1 , x_2 , \ldots , x_n) \; <=> 0 \\ g_2 (x_1 , x_2 , \ldots , x_n) \; <=> 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_m (x_1 , x_2 , \ldots , x_n) <=> 0 \end{array}\right\} \text{Restricciones} \end{array}\right\} \begin{array}{c}\text{Modelo lógico- matemático del} \\ \text{sistema}\end{array}$$

Tipos de optimización y nombres que reciben**:**

En general: Programación matemática.

Si f es lineal y las g son lineales: Programación lineal.

Si f es cuadrática y las g son lineales: Programación cuadrática.

Si f es no-lineal: Programación no lineal.

Si las x son enteras: Programación entera.

Si f ó las g incluyen funciones no lineales y/o discontinuas y/o estocásticas y su número de variables es grande: Optimización de sistemas complejos.

Métodos de optimización para sistemas deterministas

- Escenarios y estrategias.
- Algoritmos genéticos: se utilizan en los casos más complejos. Se trata de construir un algoritmo que imite a la selección natural.
- Muestreo: extraer una muestra aleatoria de tamaño *n* para que la probabilidad de obtener al menos un valor de la variable objetivo que esté dentro del *p%* de los mejores sea del *P%*.
    - Sin posterior ajuste de una función: Es evidente que $P/100 = 1 - (1 - p/100)^n$. De aquí se despeja *n.* Para p=5 y P=5 sale aproximadamente n≈60. Y si p=1 y P=1 n≈400.
    - Con posterior ajuste de una función y búsqueda de su máximo o mínimo. Teniendo los valores de las variables de entrada y su correspondiente valor de la variable

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

objetivo se puede ajustar una función no lineal sobre la que buscar máximos o mínimos.

- ....

Métodos de optimización para sistemas estocásticos

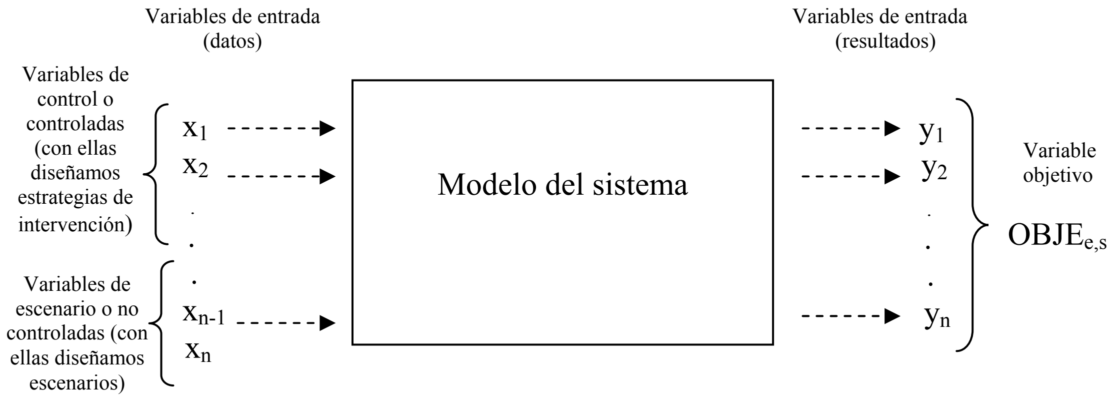
Los sistemas estocásticos incluyen variables aleatorias. Si son de entrada y normales se pueden introducir en el modelo con su media y su desviación típica. En otro caso se pueden introducir en forma de tabla de frecuencias. Si son de salida, una función calculará su valor medio y otra su desviación típica, asumiéndose normalmente distribuidas. En el caso de funciones ajustadas a una tabla de datos, el programa REGINT, además de realizar el ajuste, proporciona la información necesaria para calcular el valor medio estimado con dicha función y su correspondiente desviación típica. Las fórmulas para tales cálculos pueden verse en el Apéndice 5. En este caso son aplicables los siguientes métodos:

- Escenarios y estrategias
- Pruebas de hipótesis
- Análisis de varianza
- ....

En el Apéndice 5 puede verse un resumen de los métodos estadísticos que pueden ser especialmente útiles cuando tratamos con modelos estocásticos.

Formalización del método de los escenarios y las estrategias

Normalmente la situación con la que nos encontramos es la siguiente:



Y lo que tenemos que hacer es lo siguiente:

1. Definir la variable objetivo a partir de las variables de salida: la llamaremos OBJE.
2. Definir la variable objetivo intertemporal (si estamos en un sistema dinámico): la llamaremos $Y_{e,s}$ pues tiene un valor para cada escenario *e* y para cada estrategia *s*.
3. Definir los escenarios y las estrategias.

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

- Escenario: conjunto de valores distribuidos a lo largo del tiempo para las variables no controlables o no controladas. Se le da un nombre asociado a una situación del entorno.
- Estrategia: conjunto valores de las variables de control distribuidos a lo largo del tiempo. Se le da un nombre asociado a un tipo de acción.

4. Calcular *Ye,s* . Se obtiene sumando o promediando los valores de la variable objetivo a lo largo del tiempo.

$$Y_{e,s} = \sum_{t} OBJE_{e,s}$$

5. Asignar probabilidades a los escenarios (normalmente con opiniones de expertos): *Pe.*
6. Calcular el valor de cada estrategia. Se obtiene sumando los productos *Ye,s · Pe.*

$$Z_s = \sum_{e} Y_{e,s} \cdot P_e$$

7. Seleccionar la estrategia óptima. Se consigue encontrando el valor mayor de Z_s

$$Z_{opt} = \max_{s} Z_s$$

Ejemplo:

Veamos como quedaría el método de los escenarios y estrategias con el modelo que hemos denominado "demografía elemental".

Vamos a considerar como objetivo la población y como variables de control las tasas de natalidad y mortalidad. Ambas se podrían controlar con ayudas económicas y campañas publicitarias. Consideraremos el saldo migratorio como variable de escenario (no controlada). Simularemos 5 años.

Estrategia 1: Campaña de fomento de la natalidad.

Estrategia 2: Campaña de reducción de la mortalidad (anti tabaco y anti accidentes de tráfico).

Definición de las estrategias:

	Estrategia 1: Pro-natalidad		Estrategia 2: Pro-salud	
Años	TNAT	TMOR	TNAT	TMOR
1	10.1	9.1	10.1	9.1
2	10.2	9.1	10.1	9.0
3	10.3	9.0	10.2	8.9
4	10.5	9.0	10.2	8.7
5	10.7	8.9	10.3	8.5

Definición de los escenarios

AÑOS	Escenario 1: invasión	Escenario 2: asimilación
	SMIG	SMIG
1	40000	20000
2	45000	20000
3	50000	25000
4	60000	25000
5	80000	27000

Combinación de escenarios y estrategias: crearemos cuatro ficheros de datos.

	Invasión	Asimilación
Pro-natalidad	Datos 1	Datos 2
Pro-salud	Datos 3	Datos 4

En otros casos no tan sencillos la definición de los escenarios puede costar un poco más. La idea base de su construcción y combinación con las estrategias de control es la siguiente:

1. suponemos que tenemos una serie histórica de datos sobre cada una de las variables de escenario, llamémosles E_i a estas variables;
2. ajustamos una curva (por mínimos cuadrados) a cada una de esas series: $E_i=f_i(t)$; para ello podemos utilizar el buscador de funciones REGINT (Caselles, 1998);
3. extrapolamos la serie obteniendo para cada periodo futuro una estimación por intervalo y su valor promedio; para ello podemos utilizar el extrapolador por intervalo EXTRAPOL (veasé Caselles y Romero, 2004 y también el Apéndice 8);
4. asignamos un nombre al extremo superior, al valor promedio y al extremo inferior de los intervalos de confianza de cada variable de escenario; estos nombres pueden ser por ejemplo: optimista, tendencial y pesimista (o a la inversa, según sean favorables o desfavorables a nuestros objetivos);
5. creamos un escenario optimista utilizando los extremos optimistas de los respectivos intervalos, un escenario tendencial utilizando los valores promedios y, un escenario pesimista utilizando los extremos pesimistas;
6. diseñamos, con las variables de control, las estrategias de intervención que consideremos adecuadas y les damos un nombre (por ejemplo, estrategia agresiva, estrategia conservadora y, estrategia regresiva);
7. combinamos estos escenarios (u otros a nuestra conveniencia) con las estrategias de intervención que hayamos considerado posibles que, si son tres, y también tenemos tres escenarios, nos permitirán crear nueve ficheros de datos que llamaremos Datos1 … Datos9.

2.3.12 Realización de los experimentos o pruebas con el modelo.

Una vez diseñados los experimentos o pruebas procede realizarlos sobre el modelo construido. En los modelos construidos por SIGEM, los conjuntos de datos para las diferentes pruebas son preparados por el módulo gestor de datos (*Dnombre.bas*) en ficheros diferentes que se llaman datos1, datos2, etc. El modulo simulador (*Snombre.bas*) pregunta al usuario los números de las pruebas (simulaciones) a realizar, que se deben corresponder con los nombres de esos ficheros y elabora los correspondientes ficheros de resultados que se llaman resul1, resul2, etc., y que son filas indias de números cuya interpretación requiere del modulo generador de informes (*Rnombre.bas*).

2.3.13 Análisis de sensibilidad y control del caos.

En ocasiones, en modelos deterministas, a la vista de los resultados de una simulación, se observa que una determinada variable (o variables) toma una sucesión de valores alarmante: valores muy altos seguidos de otros muy bajos con apariencia de una total aletoriedad. Este fenómeno es lo que se ha llamado caos determinista y se presenta en los sistemas dinámicos cuando uno o más parámetros toman determinados valores, valores poco realistas en general (una tasa de natalidad enorme por ejemplo).

Todo sistema dinámico, aunque sea sencillo, es capaz de producir caos. El origen del caos es la sensibilidad a las condiciones iniciales (valores iniciales de las variables de estado). Esto significa que con un cambio muy pequeño en las condiciones iniciales se pueden obtener unos resultados muy distintos (el famoso efecto mariposa).

El fenómeno del caos se ha estudiado profusamente en uno de los sistemas dinámicos más sencillos que se conocen: la curva logística de crecimiento poblacional, cuya fórmula es

$$x_{n+1} = \mu \, x_n \, (1 - x_n)$$

Fórmula que nos dice que la población del periodo siguiente es proporcional a la del anterior y a lo que le falta para llegar a su valor máximo (que en el caso presente es 1). Cuando el valor de μ es bajo, x_n tiende a un valor fijo, pero cuando el valor de μ es muy elevado presenta grandes oscilaciones y no se puede predecir su valor final.

Dada una serie temporal de una determinada variable, con apariencia aleatoria, se puede diferenciar si es realmente aleatoria o es caótica analizando la existencia de auto-correlación, es decir, analizando si existe correlación entre cada valor y varios de los valores anteriores. Si hay correlación la serie es caótica (lo que implica que hay un sistema dinámico detrás) y si no es aleatoria (puro ruido). Existen procedimientos para determinar el número máximo de variables de estado que tendría el sistema dinámico que soporta a una variable caótica (a este número se le llama dimensión de correlación, y a la base teórica de lo anterior teorema de Whitney) (véase por ejemplo el libro de Solé y Manrubia, 1996).

Para nosotros lo interesante es saber que el caos existe, que puede aparecer cuando determinados parámetros del modelo toman valores poco habituales, que puede ser necesario identificar esos valores (imaginemos una empresa que un año se enriquezca y al siguiente se arruine) y, que dada una serie lo suficientemente larga de valores de una variable (más de mil valores), existen métodos para saber si es aleatoria (puro ruido) o caótica (forma parte de un sistema dinámico) y, en este último caso, para saber cuántas variables de estado tendría ese sistema dinámico.

Se ha venido llamando “análisis de sensibilidad” al estudio de la repercusión que pequeños cambios en los valores de los parámetros del modelo tienen sobre determinadas variables del mismo, aparezca caos o no. El control del caos es la extensión del análisis de sensibilidad hasta detectar caos en determinadas variables.

2.3.14 Interpretación de los resultados.

En la mayoría de los casos prácticos el simulador produce como resultados los valores de cada una de las variables de salida a lo largo del tiempo. Si el modelo tiene varios centenares de variables, visualizar todo ello puede ser además de largo y enrevesado, innecesario. Normalmente solo interesan los valores de algunas variables, su evolución, y su relación con algunas otras variables. En algunos casos unos gráficos pueden ayudar a percatarse mejor del significado de unos resultados. En los modelos con incertidumbre cada resultado numérico suele venir dado por un intervalo de confianza entre cuyos extremos se debe encontrar el valor real con una probabilidad del 95%, del 99% (supuesta la distribución normal), u otra. También puede venir dado por un valor medio y una desviación típica, o por una distribución de frecuencias de valores.

En los modelos elaborados por SIGEM, el modulo preparador de informes (*Rnombre.bas*) permite al usuario definir que variables desea visualizar, si lo desea en forma de tabla o de gráfico, que periodo de tiempo y con qué intervalo entre valores. De este modo el usuario del modelo puede seleccionar las variables esenciales, juntarlas en una tabla o un gráfico, relacionarlas con algunas otras variables, y así prepararse los resultados de forma que a la vista de los mismos pueda determinar qué conjunto de valores de las variables de entrada es el más adecuado para sus fines.

2.3.15 Elaboración de documentos explicativos del estudio realizado.

Es frecuente el caso en que el usuario del modelo es un consultor o un miembro del equipo asesor de la dirección de una empresa u organismo público. Entonces va a tener que elaborar un documento explicando los planteamientos realizados y los resultados obtenidos, y posiblemente también el modo de usar el modelo para repetir las pruebas o realizar otras.

En los modelos elaborados con SIGEM los programas no necesitan ser documentados porque las explicaciones necesarias aparecen tanto en el programa fuente como en la pantalla al usarlo. El programa elaborado se comporta como un experto que dialoga con el usuario aproximándose a como lo haría en realidad el experto. Las tablas y gráficos elaborados por el modulo generador de informes son de texto, fáciles de definir y rápidos de obtener, y además ofrece la posibilidad de elaborar los resultados de tal manera que puedan servir como entrada a otros paquetes comerciales más especializados (hoja de cálculo EXCEL, por ejemplo) con el fin de poder presentar tablas y gráficos con mayor calidad visual.

2.4 APLICACIONES

El objetivo de esta sección es poner en acción la metodología general para construir modelos. Obviamente no vamos poder ver un ejemplo de cada tipo de modelo y como se construiría con cada uno de los enfoques particulares de la metodología general. Veremos solamente algunos modelos con algunos enfoques.

Hemos visto en los ejemplos previos como se enfocaría el caso más sencillo de “sistema complejo” que hemos podido imaginar, el que hemos denominado “demografía elemental”. Ahora veremos algunos casos más en orden de complejidad creciente: 1.

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Demografía por cohortes. 2. Lista de espera en un hospital. 3. Un bar que solo sirve bebidas. Los enfocaremos pensando que disponemos del generador de programas SIGEM.

2.4.1 Caso 1: demografía por cohortes

Se trata en este caso de introducir el manejo de variables dimensionadas (vectores y matrices). Para ello consideraremos a la población clasificada por grupos de edad (cohortes). Así la primera cohorte podrían ser los niños de 0 a 10 años, la segunda cohorte la gente de 11 a 20 años, etc. De esta manera, no solo podemos tratar de controlar la natalidad y la mortalidad general sino que podemos tratar de incidir sobre la mortalidad de los jóvenes (accidentes de tráfico, alcohol, etc.), gente de mediana edad (campañas anti-tabaco, etc.).

2.4.1.1 Planteamiento de objetivos y restricciones.

Objetivos.

1. Determinar la evolución del número de personas de cada grupo de edad en unas condiciones determinadas de tasas de natalidad y mortalidad y de llegada y salida de familias migrantes, que también pueden cambiar a lo largo del tiempo.
2. Se desea conocer la repercusión sobre la estructura de la población a largo plazo de determinadas campañas publicitarias y otras acciones de fomento de la natalidad y de control del tráfico, tabaco y alcohol.

Restricciones y asunciones.

1. Interesa la evolución año a año durante 10 años al menos.
2. Se considera un país determinado en un momento determinado.
3. Se considera un país con población estabilizada (ni creciendo ni en declive).
4. No interesa la distinción de sexos.
5. Los movimientos migratorios se consideran no controlables.

Tipos de datos.

1. Se conoce la población histórica por grupos de edad de 10 en 10 años.
2. Se conoce el saldo migratorio por número de familias, así como la composición media de la familia migrante.
3. Se conocen las tasas históricas de natalidad y de mortalidad por cohortes de 10 años.

Tipos de resultados.

1. Modelo de Dinámica de Sistemas hipotético, a validar con un error medio máximo del 10% con los datos históricos disponibles.
2. Simulaciones de la evolución del número de personas de cada cohorte de edad a lo largo del tiempo durante 20 años. También de los nacen y de los que mueren año a año durante el periodo de simulación.

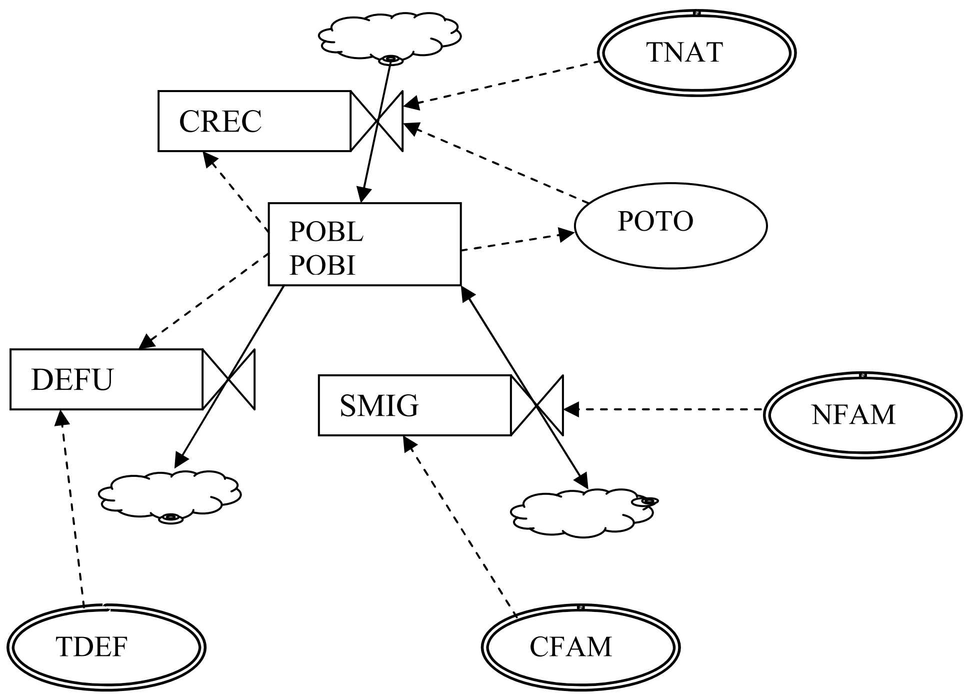
2.4.1.2 Selección de las variables relevantes.

Llamaremos al modelo: DEMO2. El contenido del fichero DEMO2.txt es el siguiente:

```
POBI Población a principio de año
     [ince=n;cval=00;ndim=1;d1=7;d2=.;esta=n;inic=....;tabl=n;npun=..;]
POBL Población a final de año
     [ince=n;cval=..;ndim=1;d1=7;d2=.;esta=s;inic=POBI;tabl=n;npun=..;]
CREC Nacimientos y crecimiento
     [ince=n;cval=..;ndim=1;d1=7;d2=.;esta=n;inic=....;tabl=n;npun=..;]
```

```
DEFU Defunciones
     [ince=n;cval=..;ndim=1;d1=7;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TNAT Tasa de natalidad
     [ince=n;cval=03;ndim=0;d1=.;d2=.;esta=sn;inic=....;tabl=sn;npun=..;]
TMOR Tasa de mortalidad
     [ince=n;cval=03;ndim=1;d1=7;d2=.;esta=sn;inic=....;tabl=sn;npun=..;]
SMIG Saldo migratorio
     [ince=n;cval=..;ndim=1;d1=7;d2=.;esta=n;inic=....;tabl=n;npun=..;]
POTO Población total a principio de año
     [ince=n;cval=..;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
POBF Población total a final de año
     [ince=n;cval=..;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
NFAM Numero de familias migrantes
     [ince=n;cval=03;ndim=0;d1=.;d2=.;esta=sn;inic=....;tabl=sn;npun=..;]
CFAM Composición de la familia migrante
     [ince=n;cval=00;ndim=1;d1=7;d2=.;esta=sn;inic=....;tabl=sn;npun=..;]
```

2.4.1.3 Identificación de las relaciones de dependencia.



2.4.1.4 Representación funcional de las relaciones.

A continuación se presenta el contenido del fichero GDEMO2.txt. Para su interpretación téngase en cuenta que:

1. Está escrito en Visual Basic 6.

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

2. Cada función va precedida de una etiqueta (en mayúsculas) que especifica la variable que se calcula con dicha función.
3. Al contador de cohortes se le llama “i1”, y especifica el número de la cohorte que estamos considerando. Obviamente, en el cálculo de las variables dimensionadas, no está escrita la sentencia “For” que inicia la cuenta ni la sentencia “Next” que permite pasar al siguiente valor de “i1”. Ambas sentencias las escribirá SIGEM automáticamente. Sí están escritas en POTO y en POBF porque ellas no son variables dimensionadas.

```
CREC if i1=1 then
        crec(1)=poto*tnat
       else
        crec(i1)=pobi(i1-1)/10-pobi(i)/10
       endif
DEFU defu(i1)=pobi(i1)*tmor(i1)
SMIG smig(i1)=nfam*cfam(i1)/100
POTO poto=0:for i1=1 to 7
      poto=poto+pobi(i1):next
POBL pobl(i1)=pobi(i1)+crec(i1)-defu(i1)+smig(i1)
POBF pobf=0:for i1=1 to 7
      pobf=pobf+pobl(i1):next
```

2.4.1.5 Programación para la computadora.

El contenido del fichero Modelo1.txt, donde se almacenan las preguntas y las respuestas al dialogo inicial con SIGEM es el siguiente:

```
"Crear, Ensamblar, formar Bloques, Analogia","c"
"Nombre del simulador","demo2"
"Servirá para","demografia por cohortes"
"Lista al completo","s"
"Ver la lista por Pantalla, Fichero, o No ver","n"
"Fichero G si y fichero C no","s"
"Ver la matriz por Pantalla, Fichero, o No ver","n"
"Listado en orden de calculo","s"
"Ordenar fichero de especificaciones: ","n"
"¿Sistema dinámico?","s"
"  Unidad dinámica","año"
"  Algunas, Todas o Ninguna variables de entrada con incertidumbre: ","n"
```

Después nos pregunta si deseamos utilizar los códigos entre corchetes sin comprobar nada. Contestamos que sí. Vamos repasando las ventanas emergentes comprobando que las variables y funciones que aparecen son las que deberían aparecer. Si se detecta algún error, procede corregirlo en el los ficheros L y G, editar el fichero modelo1.txt dejándolo como acabamos de ver, y empezar de nuevo con SIGEM. Como resultado obtendremos un módulo gestor de datos *Ddemo2.bas*, un módulo simulador *Sdemo2.bas* y un módulo gestor de resultados *Rdemo2.bas.*

2.4.1.6 Diseño de experimentos.

Procederíamos de modo análogo al del caso “demografía elemental” (véase 2.3.11 y 2.3.12) con un modo distinto de introducir los datos que es consecuencia de que la tasa de mortalidad y el saldo migratorio son ahora vectores cuyas componentes varían con el tiempo. El procedimiento operativo que seguiríamos es el siguiente:
Entramos en la hoja EXCEL. Presionamos ALT-F11 y aparece el VisualBasic-6. En el menú *Archivo* seleccionamos *importar*. Así cargamos los módulos *Ddemo2.bas*, *Sdemo2.bas* y

Rdemo2.bas. Con F5 ejecutamos *Ddemo2.bas*. En el menú principal seleccionamos la opción 5 (grabar datos), y después *numero de la simulación para grabar datos = 0.* Con ello obtendremos un fichero denominado *datos0* en el que podremos introducir nuestros datos con un editor de texto (*notebook* de Windows por ejemplo). Esta opción es más eficiente que seguir el diálogo que nos ofrece *Ddemo2.bas.* Guardamos estos datos como *datos1.* Ahora podemos ir cambiando el valor de las variables ***tnat, tdef*** y ***nfam*** de acuerdo con los escenarios y estrategias que hayamos diseñado y guardando los datos como *datos2, datos3, datos4, datos5.*
Cuando los ficheros de datos están preparados ejecutamos el módulo *Sdemo2.bas* donde diremos que queremos efectuar las simulaciones de la 1 a la 9 (suponiendo 3 estrategias y 3 escenarios). Este módulo simulador produce unos ficheros denominados *resul1*, *resul2, resul3,…, resul9, y r1ic, r2ic, r3ic,…, r9ic.* Estos ficheros contienen información que solo el módulo *Rdemo2.bas* puede interpretar.

Ahora ejecutamos el módulo *Rdemo2.bas.* En su menú principal seleccionamos la opción 2 (definir salidas) pues lo primero que tenemos que hacer es decirle cómo queremos que nos presente los resultados (variables a visualizar, cuadros o tablas, gráficos, número de años, etc.). Con ello creamos un fichero con las especificaciones de cada informe (tabla o gráfico) al que podemos dar el nombre que queramos (*salidas*, por ejemplo). Con las salidas definidas elegimos la opcion1 (ejecutar salidas) del menú principal. Nos preguntará el número del primer fichero EXCEL. Le diremos que 1 (por ejemplo). Con ello nos creará unos ficheros llamados *excel1, excel2, excel3, etc.* Estos ficheros son ficheros de texto que pueden ser visualizados con cualquier editor de texto. También pueden ser importados por la hoja EXCEL en la que estamos trabajando y producir a partir de ellos gráficos de calidad.

2.4.2 Caso 2: lista de espera en un hospital

Con este caso aprenderemos a tratar con sistemas estocásticos donde la incertidumbre se encuentra en las variables de entrada. Si una variable de entrada es aleatoria y su distribución es Normal podemos utilizar como datos su media y su desviación típica. Si su distribución no es Normal o no está clasificada, podemos utilizar como datos su tabla de frecuencias relativas. Este último caso es el que se considera en el problema siguiente.

2.4.2.1 Planteamiento de objetivos y restricciones.

Objetivos.
1. Determinar la evolución del número de pacientes en lista de espera, del número de camas ocupadas, del número de camas libres y de la nueva ocupación de camas, a lo largo del tiempo en unas condiciones determinadas de número total de camas y de llegada y salida de pacientes, que también pueden cambiar a lo largo del tiempo.

Restricciones.
1. Interesa la evolución día a día durante 30 días al menos.
2. Se considera un departamento o servicio de nueva creación dentro de un hospital.
3. Se desea encontrar el número de camas óptimo para que la lista de espera sea mínima y, a la vez, el número de camas desocupadas sea también mínimo, una vez alcanzado el equilibrio.
4. A pesar de que ese servicio hospitalario es de nueva creación, existen datos estadísticos de llegada y salida de pacientes del tipo que los que serían atendidos en el mismo pero, no existen datos adecuados de listas de espera ni de camas ocupadas.

Tipos de datos.

1. El número total de camas que podrían instalarse, entre un mínimo de 5 y un máximo de 25.
2. Tablas de frecuencias absolutas especificando el número de pacientes que entrarían y que saldrían en un día determinado.

Llegan	Frecuencia (%)		Salen	Frecuencia (%)
1	25		0	10
2	35		1	15
3	30		2	25
4	10		3	30
			4	10

Tipos de resultados.

1. Modelo de Dinámica de Sistemas hipotético, a validar en la medida de lo posible con los datos disponibles (No será posible hacerlo por el método de predicción del pasado, dado que no existen datos de camas ocupadas ni de listas de espera. Se validará por opiniones de expertos, es decir, "si se creen el modelo deberán creerse sus resultados").
2. Simulaciones de la evolución del número de camas libres y de camas ocupadas a lo largo del tiempo durante 30 días. También de los pacientes en lista de espera y de los que se reciben una cama. Todo ello para distintos valores del número total de camas.
3. Determinación del número óptimo de camas deducido como consecuencia de las simulaciones efectuadas.

2.4.2.2 Selección de las variables relevantes.

Llamaremos al modelo: HOSP. El contenido del fichero LHOSP.txt es el siguiente:

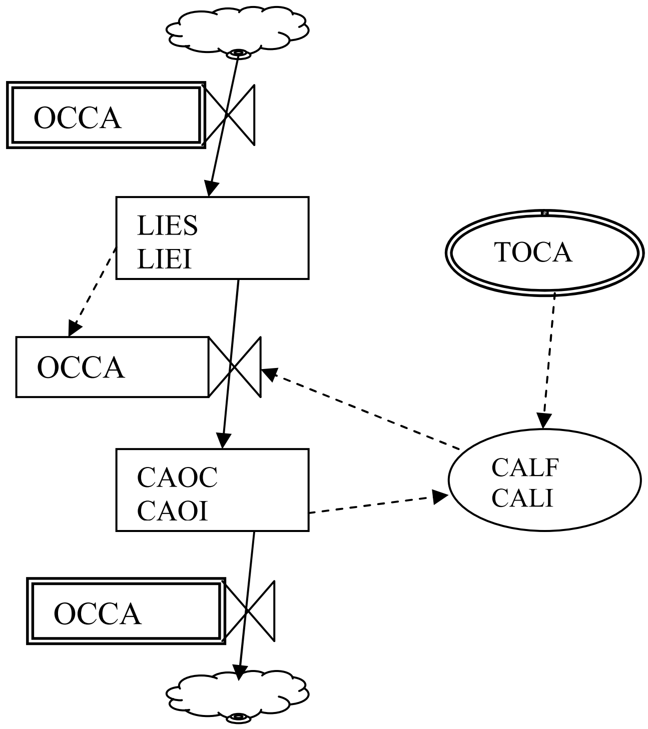
```
LLPA Llegada de pacientes
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
LIES Numero de pacientes en lista de espera
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=s;inic=LIEI;tabl=n;npun=..;]
LIEI Numero de pacientes inicial en lista de espera
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
OCCA Ocupacion de camas
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
CAOC Camas ocupadas
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=s;inic=CAOI;tabl=n;npun=..;]
CAOI Camas ocupadas inicialmente
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
SAPA Salida de pacientes
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TOCA Total camas disponibles
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
CALF Camas libres al final del día
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=s;inic=CALI;tabl=n;npun=..;]
CALI Camas libres al principio del día
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
```

Obsérvese que hemos declarado como variables con incertidumbre a LLPA y SAPA que son variables de entrada y a la hora de introducir sus datos introduciremos sus respectivas tablas de frecuencias con el siguiente formato, que para la variable LLPA será:

"1,25/2,35/3,30/4,10". Obsérvese también que hemos declarado como variables con incertidumbre a LIEI, CAOI y CALI, que son los valores iniciales de las variables de estado. Esta clase de variables, en los modelos con incertidumbre, deben ser declaradas como variables con incertidumbre aunque no la tengan (por requerimientos del SIGEM).

2.4.2.3 Identificación de las relaciones de dependencia.

Este sería el diagrama de Forrester correspondiente a este caso:



2.4.2.4 Representación funcional de las relaciones.

El contenido del fichero GHOSP.txt es el siguiente:

```
LIES lies=liei+llpa-occa
      if lies<0 then lies=0
OCCA if cali <= liei then occa=cali else occa=liei
CAOC caoc=caoi+occa-sapa
      if caoc<0 then caoc=0
CALF calf=toca-caoc
```

Obsérvese que estas funciones incluyen tautologías (ecuaciones evidentes) y relaciones lógicas de puro sentido común.

2.4.2.5 Programación para la computadora.

El contenido del fichero Modelo1.txt, donde se almacenan las preguntas y las respuestas al dialogo inicial con SIGEM es el siguiente:

```
"Crear, Ensamblar, formar Bloques, Analogia","c"
"Nombre del simulador","HOSPITAL"
"Servirá para","simular camas hospital"
"Lista al completo","s"
"Ver la lista por Pantalla, Fichero, o No ver","n"
"Fichero G si y fichero C no","s"
"Ver la matriz por Pantalla, Fichero, o No ver","n"
"Listado en orden de calculo","s"
"Ordenar fichero de especificaciones: ","n"
"¿Sistema dinámico?","s"
"  Unidad dinámica","dia"
"  Algunas, Todas o Ninguna variables de entrada con incertidumbre: ","a"
```

Después nos pregunta si deseamos utilizar los códigos entre corchetes sin comprobar nada. Contestamos que sí. Vamos repasando las ventanas emergentes comprobando que las variables y funciones que aparecen son las que deberían aparecer. Si se detecta algún error, procede corregirlo en el los ficheros L y G, editar el fichero modelo1.txt dejándolo como acabamos de ver, y empezar de nuevo con SIGEM.

2.4.2.6 Diseño de experimentos.

Este es un modelo con incertidumbre en dos variables (flujos) de entrada, la llegada y la salida de pacientes. La simulación con este tipo de modelos se basa en repetir la corrida del modelo un buen número de veces con extracciones aleatorias de valores de las variables con incertidumbre, guardar los resultados y después calcular las medias y las desviaciones típicas de cada resultado. A continuación, procede hacer la prueba de normalidad en las distribuciones de cada resultado (con *Chi-cuadrado*). Si resultan ser normales podremos realizar su estimación por intervalo utilizando la *t de Student.* Si no son normales nos conformaremos con la desviación típica como medida del grado de dispersión de cada resultado.

Respecto al tipo de experimentos que podemos realizar con este modelo, observemos que lo que interesa es el número óptimo de camas y al final el número de camas que van a estar libres y ocupadas cuando el sistema se estabilice. Simularemos 30 días, efectuaremos 100 repeticiones (un número alto pero asumible; a más repeticiones más estrechos los intervalos de confianza). Consideraremos el hospital inicialmente vacío y simularemos que tenemos 5 camas, después 10, etc., hasta 25. Con los datos disponibles seleccionaríamos la opción con la mayor relación entre camas ocupadas y camas libres, aunque el ideal sería contar con datos económicos para poder seleccionar la opción que maximice el beneficio. Como ayuda a la selección dibujaremos juntas las gráficas de evolución temporal de ***cali*** y de ***caoc***. También es interesante la evolución de la lista de espera si es que la hay.

2.4.2.7 Ejecución de los experimentos

Entramos en la hoja EXCEL. Presionamos ALT-F11 y aparece el VisualBasic-6. En el menú *Archivo* seleccionamos *importar*. Así cargamos los módulos *Dhospital.bas*, *Shospital.bas* y *Rhospital.bas*. Con F5 ejecutamos *Dhospital.bas*. En el menú principal seleccionamos la opción 5 (grabar datos), y después *numero de la simulación para grabar datos = 0.* Con ello obtendremos un fichero denominado *datos0* en el que podremos introducir nuestros datos con un editor de texto (*notebook* de Windows por ejemplo). Esta opción es más eficiente que

seguir el diálogo que nos ofrece *Dhospital.bas.* Guardamos estos datos como *datos1.* Ahora podemos ir cambiando el valor de la variable ***toca*** desde 5 hasta 25 y guardando los datos como *datos2, datos3, datos4, datos5.*
Cuando los ficheros de datos están preparados ejecutamos el módulo *Shospital.bas* donde diremos que queremos efectuar las simulaciones de la 1 a la 5. Este módulo simulador produce unos ficheros denominados *resul1*, *resul2, resul3, resul4, resul5, y r1ic, r2ic, r3ic, r4ic, r5ic.* Estos ficheros contienen información que solo el módulo *Rhospital.bas* puede interpretar.

Ahora ejecutamos el módulo *Rhospital.bas.* En su menú principal seleccionamos la opción 2 (definir salidas) pues lo primero que tenemos que hacer es decirle cómo queremos que nos presente los resultados (variables a visualizar, cuadros o tablas, gráficos, número de años, etc.). Con ello creamos un fichero con las especificaciones de cada informe (tabla o gráfico) al que podemos dar el nombre que queramos (*salidas*, por ejemplo). Con las salidas definidas elegimos la opcion1 (ejecutar salidas) del menú principal. Nos preguntará el número del primer fichero EXCEL. Le diremos que 1 (por ejemplo). Con ello nos creará unos ficheros llamados *excel1, excel2, excel3, etc.* Estos ficheros son ficheros de texto que pueden ser visualizados con cualquier editor de texto. También pueden ser importados por la hoja EXCEL en la que estamos trabajando y producir a partir de ellos gráficos de calidad.

2.4.3 Control de la estabilidad de las parejas

El objetivo de aprendizaje que tenemos en este caso es el manejo de la incertidumbre en las funciones. Suponemos que una variable determinada depende de otras y no sabemos cómo depende. No obstante disponemos de datos históricos de las variables implicadas. Con el buscador/ajustador de funciones REGINT encontramos la función más adecuada (véase 2.3.4 -g). Los ficheros de entrada de datos y de salida de resultados de REGINT correspondientes a este caso pueden verse en el Apéndice 7.

2.4.3.1 Planteamiento de objetivos y restricciones.

Objetivos.

1. Fomentar la estabilidad de las parejas.

Restricciones.

1. Nos restringiremos a España y a la próxima década.
2. Entendemos por pareja cualquier tipo de convivencia estable (compartir domicilio) entre personas no consanguíneas.
3. Como unidad de tiempo tomaremos el año.

Asunciones.

1. Asumimos que la estructura social de los últimos 15 años se perpetuará durante la próxima década.

Tipos de datos.

1. Demográficos y sociológicos del Instituto Nacional de Estadística de España de los últimos 15 ó 16 años (los únicos existentes).

Tipos de resultados.

1. Buscamos la estrategia óptima con las variables que resulten controlables por el gobierno central.

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

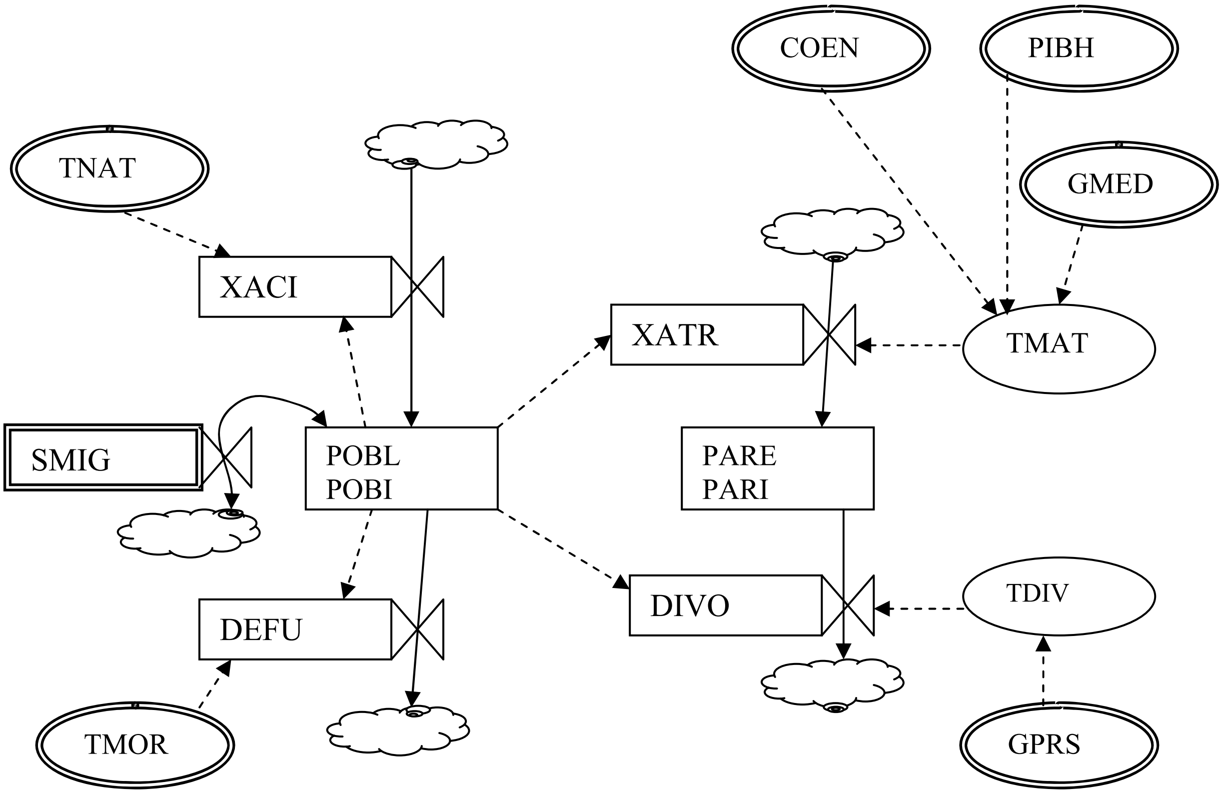
2.4.3.2 Selección de las variables relevantes.

Si llamamos PAREJAS al futuro simulador del problema, el contenido del fichero Lparejas.txt que los métodos conocidos seleccionaron es el siguiente. Obsérvese que la tasa de divorcios TDIV y la tasa de matrimonios TMAT se declaran como variables con incertidumbre (van a ser calculadas con funciones determinadas por REGINT a partir de tablas de datos históricos). También han sido declaradas como variables con incertidumbre POBI y PARI que son los valores iniciales de las variables de estado (esto se hace así por requerimientos de SIGEM). Se tomaron en consideración más variables de las que aparecen a continuación pero fueron eliminadas por el propio REGINT al no incluirlas en las ecuaciones ajustadas a los datos históricos. No obstante dichas variables sí que aparecen en el fichero de datos de REGINT que puede verse en el Apéndice 7.

```
TDIV Tasa de divorcios (div+sep+nul)*1000/población
     [ince=s;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
POBL Poblacion a final de año
     [ince=n;cval=..;ndim=0;d1=.;d2=.;esta=s;inic=POBI;tabl=n;npun=..;]
POBI Poblacion a principio de año
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TMAT Tasa de matrimonios por cada 1000 habitantes
     [ince=s;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
COEN Consumo de energia(Tm petroleo/1000 Hab)
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
PIBH Producto Interior Bruto por Habitabte (€/Año)
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
GMED Gasto medio por habitante (€/Año)
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
GPRS Gasto protección social (€/Año)
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
XACI Nacimientos
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
DEFU Defunciones
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TNAT Tasa de Natalidad
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
TMOR Tasa de Mortalidad
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
SMIG Saldo Migratorio
     [ince=n;cval=14;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
XATR Matrimonios
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
DIVO Divorcios
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
PARE Parejas a fin de año
     [ince=n;cval=0;ndim=0;d1=.;d2=.;esta=s;inic=PARI;tabl=n;npun=..;]
PARI Parejas a principio de año
     [ince=s;cval=0;ndim=0;d1=.;d2=.;esta=n;inic=....;tabl=n;npun=..;]
```

2.4.3.3 Identificación de las relaciones de dependencia.

Diagrama de Forrester correspondiente a este caso:



2.4.3.4 Representación funcional de las relaciones.

La lista de ecuaciones y/o reglas que describen el comportamiento de las variables del modelo (contenido del fichero Gparejas.txt) y que se deducen de la interpretación del planteamiento de objetivos y restricciones y del ajuste de TMAT y TDIV con REGINT son las siguientes. Obsérvese que:

a) escribimos $h=a+b_1*T_1+b_2*T_2+...+b_n*Tn$ siendo a la constante y las b_1, b_2, *etc.,* los coeficientes de las funciones transformadas de las variables independientes. Así, `coen/10*pibh/1000` es una función transformada donde *coen* y *pibh* son variables independientes en el cálculo de TMAT.

b) Usamos dos variables auxiliares: A y B para almacenar el valor de la diferencia entre cada transformada y su valor medio (obtenido al igual que la constante a y las b_1, b_2, *etc.,* del fichero de resultados de REGINT).

c) Escribimos $s=S*sqr(1+1/n+\tau'C\tau)$, donde S es la desviación típica de regresión, *sqr(...)* significa raíz cuadrada de (…), n es el número de puntos de la tabla de datos que utiliza REGINT, τ es el vector cuyas componentes son A y B (en el caso de dos transformadas), τ' es el vector transpuesto de τ, y C es la matriz de los numeradores de las varianzas y covarianzas (también calculada por REGINT) (Véase 3.3 del Apéndice 5).

d) En la tabla de datos que utiliza REGINT hemos dividido *pibh* por 1000 (y *coen* por 10) porque se trabaja mejor con números de un mismo o parecido orden de magnitud (no es bueno mezclar números muy grandes con números muy pequeños). No obstante, cuando se utilicen *pibh* y *coen* como datos en el simulador definitivo entrarán con su valor verdadero.

```
TMAT h=15.6063986+0.0056761*coen/10*pibh/1000-2.9454791*sqr(gmed/1000)
      A=coen/10*pibh/1000-403.368005: B=sqr(gmed/1000)-4.333173
      s=0.080517*sqr(1+1/15+0.000094296810*A^2+20.206693726304*B^2-
        2*0.043011114874*A*B)
TDIV h=11.591269-12.030491*exp(-0.1*gprs)
      A=exp(-0.1*gprs)-0.745145
      s=0.080180*sqr(1+1/16+23.606967*A^2)
XACI xaci=pobi*tnat/1000
DEFU defu=pobi*tmor/1000
POBL pobl=pobi+xaci-defu+smig
XATR xatr=pobi*tmat/1000
DIVO divo=pobi*tdiv/1000
PARE pare=pari+xatr-divo
```

2.4.3.5 Programación para la computadora.

El contenido del fichero Modelo1.txt, donde se almacenan las preguntas y las respuestas al dialogo inicial con SIGEM es el siguiente:

```
"Crear, Ensamblar, formar Bloques, Analogia","c"
"Nombre del simulador","PAREJAS"
"Servirá para","fomentar la estabilidad de las parejas"
"Lista al completo","s"
"Ver la lista por Pantalla, Fichero, o No ver","n"
"Fichero G si y fichero C no","s"
"Ver la matriz por Pantalla, Fichero, o No ver","n"
"Listado en orden de calculo","s"
"Ordenar fichero de especificaciones: ","n"
"¿Sistema dinámico?","s"
"  Unidad dinámica","año"
"  Algunas, Todas o Ninguna variables de entrada con incertidumbre: ","a"
```

Después nos pregunta si deseamos utilizar los códigos entre corchetes sin comprobar nada. Contestamos que sí. Vamos repasando las ventanas emergentes comprobando que las variables y funciones que aparecen son las que deberían aparecer. Si se detecta algún error, procede corregirlo en el los ficheros L y G, editar el fichero modelo1.txt dejándolo como acabamos de ver, y empezar de nuevo con SIGEM.

2.4.3.6 Diseño de experimentos.

Consideremos que el gobierno central puede tratar de intervenir o controlar el Producto Interior Bruto por habitante (PIBH) fomentando la actividad empresarial con leyes adecuadas y, que controla totalmente los gastos de protección social (GPRS). En cambio no controla el consumo de energía (COEN) ni el gasto medio de las familias (GMED) así como las variables demográficas TNAT, TMOR y SMIG. Tenemos pues como variables de estrategia a PIBH y a GPRS, y como variables de escenario a COEN, GMED, TNAT, TMOR y SMIG. Ahora podemos diseñar estrategias y escenarios y plantear combinaciones de escenarios y estrategias de modo similar a como se ha hecho en 2.3.11. Trabajaremos con los valores medios, no obstante obtendremos la evolución de las variables de salida con su intervalo de confianza o, al menos, con su desviación típica. La construcción de escenarios con las variables no controladas la realizaremos ajustando con REGINT una función del tiempo (años) a la serie histórica de cada variable de escenario y posteriormente extrapolando esta función con intervalos de confianza con EXTRAPOL (Véase Apéndice 8). Si diseñamos tres escenarios (lo más sencillo) que podemos llamar "escenario optimista", "escenario tendencial" y "escenario pesimista", podemos asignar al escenario optimista los extremos superiores de los intervalos de confianza extrapolados, al tendencial los valores medios y al pesimista los extremos inferiores. En el Apéndice 8 puede verse cómo se ha realizado el ajuste y la extrapolación para el caso de la tasa de natalidad TNAT.

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

2.4.4 Un bar que solo sirve bebidas

2.4.4.1 Planteamiento de objetivos y restricciones.

- Conociendo el tiempo que necesita una camarera para servir una bebida y el tiempo que necesita para lavar un vaso;
- sabiendo que los clientes llagan siguiendo una ley exponencial negativa de media 10, que el tiempo que necesitan para consumir una bebida está uniformemente distribuido entre 5 y 8, y que su necesidad de otra bebida está uniformemente distribuida entre 1 y 4;
- queremos determinar cómo varía con el tiempo el número de clientes esperando para que les sirvan, el número de clientes servidos, y el porcentaje de su tiempo en que la camarera está sirviendo, lavando o desocupada.

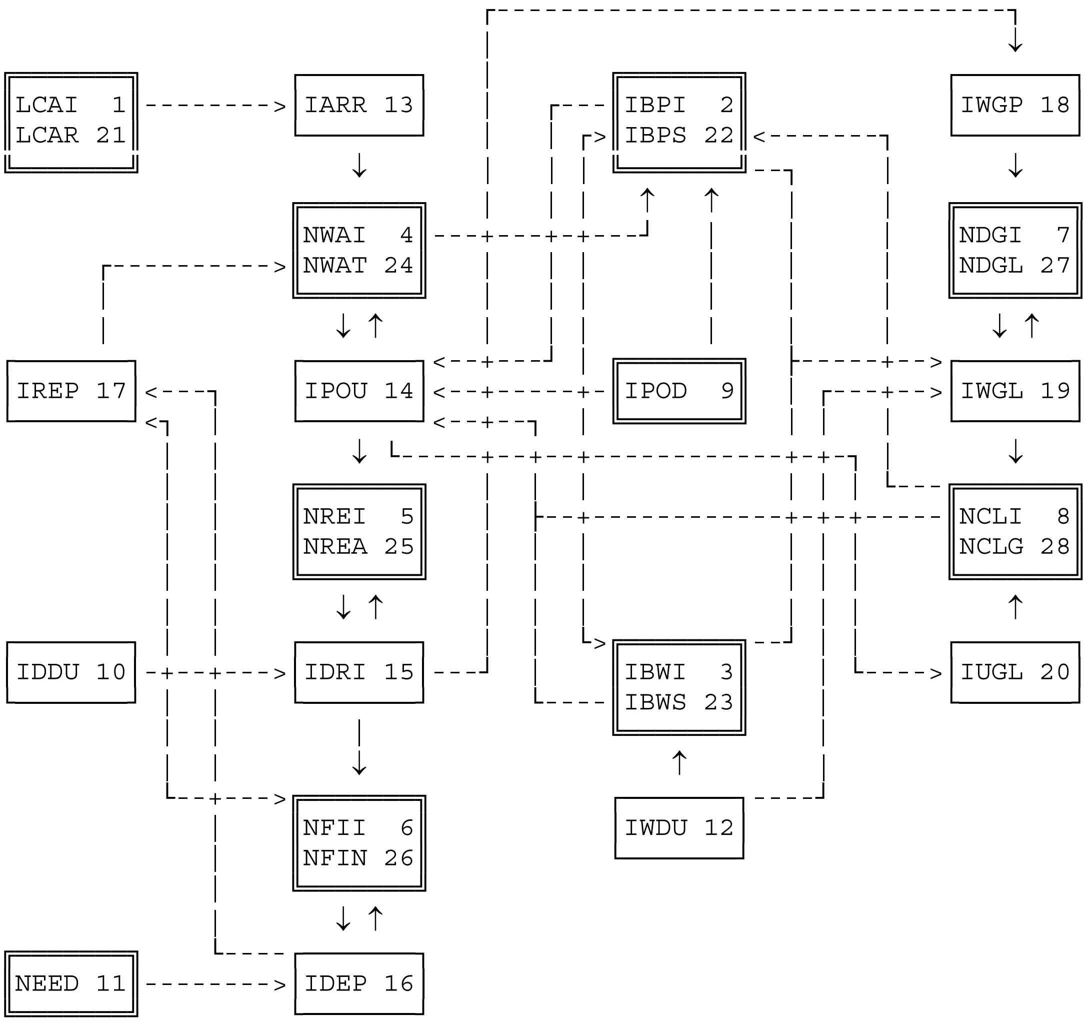
2.4.4.2 Selección de las variables relevantes.

La lista de variables que los métodos conocidos seleccionaron es la siguiente.

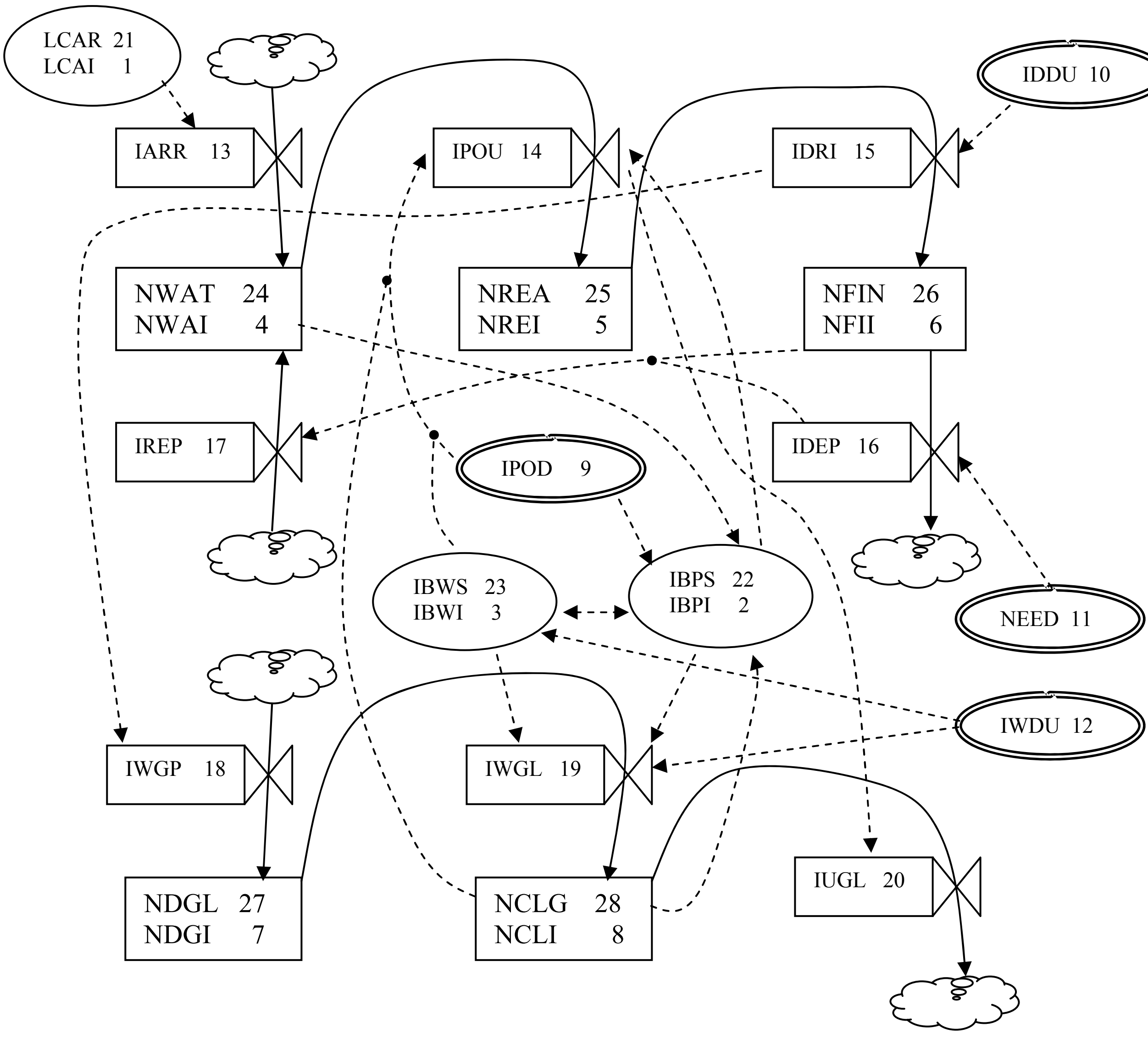
```
1  LCAI Minuto inicial en que llegó el último cliente.
2  IBPI Camarera inicialmente sirviendo o no (0=NO, 1=SÍ).
3  IBWI Camarera inicialmente lavando vasos o no (0 o 1).
4  NWAI Número inicial de clientes esperando ser servidos.
5  NREI Número inicial de clientes servidos.
6  NFII Número inicial de clientes que han terminado su bebida.
7  NDGI Número inicial de vasos sucios.
8  NCLI Número inicial de vasos limpios.
9  IPOD Duración de un servicio (minutos).
10 IDDU Duración de una bebida (minutos).
11 NEED Probabilidad de repetición de la bebida.
12 IWDU Duración de la operación de lavado de un vaso (min.).
13 IARR Llegada de un cliente (0 o 1 en un minuto dado).
14 IPOU Cliente siendo servido (0 o 1 en un minuto dado).
15 IDRI Número de clientes que han terminado su bebida.
16 IDEP Número de clientes saliendo del pub.
17 IREP Número de clientes que desean repetir.
18 IWGP Número de vasos vacios retirados en un minuto dado.
19 IWGL Número de vasos lavados en un minuto dado.
20 IUGL Número de vasos usados en un minuto dado.
21 LCAR Tiempo en que llegó el último cliente (minuto n°).
22 IBPS Camarera sirviendo o no (0 o 1).
23 IBWS Camarera lavando o no (0 o 1).
24 NWAT Número de clientes esperando ser servidos.
25 NREA Número de clientes servidos.
26 NFIN Número de clientes que han terminado su bebida.
27 NDGL Número de vasos sucios.
28 NCLG Número de vasos limpios.
```

2.4.4.3 Identificación de las relaciones de dependencia.

Los métodos conocidos conducen al grafo de conexiones (diagrama causal) que se presenta a continuación. Este gráfico deberá entenderse utilizando la lista de relaciones funcionales especificada en 2.4.4.4.



Y al diagrama de Forrester que aparece a continuación. Este gráfico también deberá entenderse utilizando la lista de relaciones funcionales especificada en 2.4.4.4.



2.4.4.4 Representación funcional de las relaciones.

La lista de ecuaciones y/o reglas que describen el comportamiento de las variables del modelo (contenido del fichero Gpub.txt) y que se deducen de la interpretación del planteamiento de objetivos y restricciones son las siguientes (las comentaremos después).

```
LCAR x=k-lcai:px=.1*exp(-.1*x):if rnd<=px then lcar=k
IARR x=k-lcai:px=.1*exp(-.1*x):if rnd<=px then iarr=1 else iarr=0
IWGP iwgp=idri
IREP irep=nfii-idep
IBPS if nwai>0 and ncli>0 and ibpi=0 and ibwi=0 then
       ibps=1
      else if ibpi>0 and ibpi<ipod then ibps=ibpi+1
      else if ibpi=ipod then ibps=0
      endif
```

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

```
IPOU if nwai>0 and ncli>0 and ibpi=0 and ibwi=0 then
        ipou=0
       else if ibpi>0 and ibpi<ipod then ipou=0
       else if ibpi=ipod then ipou=1
       endif
IBWS if ndgi>0 and ibwi=0 and ibpi=0 then
        ibws=1
       else if ibwi>0 and ibwi<iwdu then ibws=ibwi+1
       else if ibwi=iwdu then ibws=0
       endif
IWGL if ndgi>0 and ibwi=0 and ibpi=0 then
        iwgl=0
       else if ibwi>0 and ibwi<iwdu then iwgl=0
       else if ibwi=iwdu then iwgl=1
       endif
IDRI idri=0:for i1=1 to nrei:if rnd<iddu then idri=idri+1
       next
IUGL iugl=ipou
IDEP idep=0:for i1=1 to nfii:if rnd<need then idep=idep+1
       next
NWAI nwai=nwai+iarr+irep-ipou
NREA nrea=nrei+ipou-idri
NFIN nfin=nfii+idri-idep
NDGL ndgl=ndgi+iwgp-iwgl
NCLG nclg=ncli+iwgl-iugl
```

Las relaciones LCAR y IARR se interpretan como sigue: siendo k el contador de tiempo, siendo X el número de periodos transcurrido desde la llegada del última cliente, siendo PX la probabilidad de X, y siendo RND un número aleatorio uniformemente distribuido entre 0 y 1, si obtenemos un valor de RND menor que PX llega un cliente y entonces IARR=1 y LCAR=k.

La relación IWGP es evidente. La relación IREP establece que los clientes que terminan su bebida y no se van repiten la bebida.

Las relaciones IPOU e IBPS establecen que si hay clientes esperando ser servidos, hay vasos limpios, y la camarera está libre, entonces la camarera empieza a servir, pero: si la camarera ya estaba sirviendo y no había terminado, entonces continúa sirviendo, y si ya había terminado entonces ya hay un cliente servido más.

Las relaciones IBWS e IWGL establecen que si hay vasos sucios y la camarera está desocupada, entonces la camarera empieza a lavar vasos, pero: si estaba lavando vasos y no había terminado entonces continúa haciéndolo, y si ya había terminado entonces ya hay un vaso limpio más.

La relación IDRI cuenta el número de clientes que terminan su bebida en un periodo dado. Para cada cliente bebiendo obtiene un número aleatorio uniformemente distribuido entre 0 y 1. Si ese número es menor que la probabilidad de que ese cliente haya terminado su bebida, entonces ese cliente se va.

La relación IUGL es evidente. La relación IDEP cuenta el número de clientes que se van en un periodo determinado y es similar a la IDRI. Las relaciones que quedan son evidentes.

2.4.4.5 Programación para la computadora.

En el desarrollo de una sesión de trabajo con SIGEM correspondiente a este ejercicio podrán verse los programas producidos por SIGEM.

2.4.4.6 Diseño de experimentos.

Recordemos el planteamiento del problema: véase 2.4.4.1. En este caso tan simplificado de lo que es el funcionamiento de un bar, ¿Qué puede interesarnos averiguar? Obviamente, nos interesará averiguar si la camarera puede o no atender a su trabajo y si el negocio será o no interesante (número de clientes atendidos y porcentaje de ocupación del tiempo de la camarera). Otras cuestiones posteriores pueden ser las siguientes.

(a) En el caso de que las colas de clientes esperando ser atendidos o de vasos sucios fuesen excesivas, ¿Convendría introducir otras tecnologías para acelerar los procesos de lavado de vasos o de servicio a los clientes? Por ejemplo: una máquina de lavar vasos o el sistema de autoservicio a las mesas.

(b) Cuando la camarera inicia su trabajo, ¿Qué es más conveniente, comenzar lavando vasos o atendiendo a los clientes que esperan?

(c) Se supone que no es posible influir sobre la llegada de un número mayor de clientes pero, ¿Convendría introducir nuevas técnicas para que los clientes que acuden repitiesen más veces la consumición? Por ejemplo: una televisión, juegos, etc.

(d) ¿Convendría que el número total de vasos fuese mayor? o ¿Convendría lavarlos a otra hora?

Para poder contestar a la pregunta (a) se requieren los valores de IPOD y IWDU correspondientes a las distintas alternativas. Para contestar a la pregunta (c) se requieren los valores de IDDU y NEED de las alternativas correspondientes. La cuestión (b) se contesta dando los valores 1 y 0 ó bien 0 y 1 a las variables IBPI e IBWI en relación con los valores estimados de NDGI (número inicial de vasos sucios), NCLI (número inicial de vasos limpios) y, NWAI (número inicial de clientes esperando). Con relación a la cuestión (d) se requieren los valores de NDGI y NCLI alternativos.

Parece lógico pensar que no todos los días son iguales respecto a la afluencia de clientes, y tampoco lo son todas las horas del día. No obstante, el enunciado así parece considerarlo. La consecuencia para nosotros es que lo que se desea estudiar es una determinada hora de un determinado día. Simularemos, por tanto, el comportamiento del pub en una hora, es decir, en 60 minutos, puesto que los tiempos de las operaciones los tenemos en minutos.

Conviene hacer notar cuales son las variables de acción o de control del sistema (las que el usuario controlar), y cuáles son las variables esenciales (las que permiten evaluar el resultado de un ensayo). En este caso, son de acción: IBPI, IBWI, IPOD, IWDU, IDDU, NEED, NDGI, NCLI; y son esenciales: NWAT, NREA, NFIN, NDGL, NCLG.

Como consecuencia de los razonamientos anteriores proponemos los siguientes ensayos.

Ensayo 1: Situación actual. Daremos los valores promedio observados en la realidad a las variables de estado inicial del sistema, a IDDU y a NEED les daremos valores aleatorios distribuidos como indica el enunciado, y a IPOD y IWDU les daremos los valores conocidos de la situación actual. Por ejemplo: LCAI=1; IBPI=0; IBWI=1; NWAI=3; NREI=2; NFII=1; NDGI=5; NCLI=10; IPOD=1.1; IDDU=5,25/6,25/7,25/8,25; NEED=1,25/2,25/3,25/4,25; IWDU=0.25.

Ensayo 2: ¿Empezar lavando o sirviendo? Para contestar esta pregunta se requiere repetir el ensayo 1 con los valores IBPI=1; IBWI=0.

Ensayo 3: ¿Introducir una máquina de lavar vasos? Probaríamos con los mismos datos pero con IWDU=0.10.

Ensayo 4: ¿Autoservicio? Cambiaríamos el valor de IPOD dándole el valor 0.30.

Ensayo 5: ¿Televisión y/o juegos? Probaríamos las correspondientes tablas de valores de IDDU y NEED. Por ejemplo: IDDU=6,25/7,25/8,25/9,25; NEED=3,25/4,25/5,25/6,26.

Ensayo 6: Seleccionar la mejor combinación de alternativas. En un caso sencillo como este tenemos solamente 16 combinaciones de alternativas y no es mucho trabajo probarlas todas. En situaciones reales el número de combinaciones suele ser de miles, quizás millones, y suele ser necesario recurrir al muestreo. Es decir, si no tenemos tiempo o medios de cálculo para evaluar todas las posibilidades, seleccionemos una muestra aleatoria representativa del total y quedémonos con la mejor solución dentro de esa muestra. Se demuestra con relativa facilidad que una muestra aleatoria de tamaño n=60 permite capturar al menos a una solución situada dentro del p=5% de las mejores con una probabilidad del P=95%, cualquiera que sea el número total de posibilidades. Para p=1% y P=99% se obtiene un tamaño de muestra n=460. La relación utilizada para el cálculo es: $P=1-(1-p)^n$, que una vez meditada resulta evidente. No obstante, este tipo de enfoque es más adecuado para modelos de tipo determinista que para modelos de tipo estocástico como es el caso que nos ocupa. Téngase en cuenta que cada ensayo debe repetirse un cierto número de veces con valores al azar de IDDU, NEED, LCAR y IARR, y tomar como valores de las variables esenciales los promedios correspondientes. Entonces cabe que nos preguntemos: dada una diferencia entre los valores promedio de una variable esencial correspondientes a dos ensayos distintos ¿Es significativa? Es decir, ¿se debe a que realmente hay diferencia entre los dos ensayos? o ¿se debe al efecto del azar? Para dilucidar cuestiones como esta hay que recurrir al Análisis de Varianza y a las técnicas de diseño de experimentos apropiadas.

2.5 DISCUSIÓN

En este punto trataremos de efectuar algunas comparaciones, tanto en lo referente a la metodología propuesta como al ejemplo estudiado, con otras alternativas conocidas. Las comparaciones que haremos no son las únicas posibles ni siquiera las más significativas, pueden considerarse con elegidas al azar y las realizamos con el único fin de ilustrar el tipo de diferencias que pueden existir y el modo de establecer comparaciones.

2.5.1 Comparación entre eLSE y SIGEM.

El ejemplo del pub puede encontrarse también en un artículo de Crookes et al. (1986), enfocado desde el entorno eLSE elaborado por la universidad de Lancaster (U.K.).

Comienza esta metodología elaborando, a partir de la descripción del problema, el diagrama llamado de "ciclo de actividades" o de actividades y colas, o de actividades simplemente. La idea que lo rige es que en el sistema real existen "entidades" que circulan y que se detienen. Cuando se detienen lo hacen en una "cola" a la espera de que se ejecute sobre ellas o con ellas una "actividad". Las actividades están relacionadas con determinados "sucesos". Así pues, es necesario distinguir los siguientes tipos de elementos.

- Entidades. En el ejemplo del pub: clientes, vasos, camarera.

- Colas. Son estados pasivos de las entidades. Son colas los estados de espera de los clientes para ser atendidos por la camarera, y los de los vasos para ser lavados.

- Actividades. Su iniciación depende de que existan entidades disponibles en las respectivas colas. Por ejemplo: la actividad servir una bebida solo podrá iniciarse cuando existan clientes esperando, vasos limpios y la camarera desocupada.

- Sucesos. Son cambios de estado del modelo que ocurren en un instante del tiempo. Son de destacar los sucesos que corresponden al fin de las actividades ("sucesos finales") y los que se corresponden con el inicio de las mismas ("sucesos condicionales").

El paso siguiente de esta metodología es la programación del modelo siguiendo el llamado "método de las tres fases" que consiste en, previa inicialización:

Fase A. Comprobar el tiempo de finalización de las actividades en progreso. Encontrar el menor y avanzar el reloj hasta ese tiempo.

Fase B. Ejecutar los sucesos finales identificados en la fase A, es decir, mover las entidades apropiadas dentro de las respectivas colas.

Fase C. Intentar ejecutar todos los sucesos condicionales en turno y ejecutar aquellos cuyas condiciones se cumplen. Repetir el intento hasta que ya no sea posible ejecutar ninguno más de estos sucesos, es decir, hasta que ya no puedan iniciarse más actividades.

Finalmente, pasar otra vez a la fase A, o elaborar resultados y terminar.

La programación la realiza el usuario en PASCAL ayudándose de la librería de eLSE y de un programa marco o esquemático que se le proporciona. La librería de eLSE contiene cinco unidades denominadas: (1) Entidades, (2) Muestreo, (3) Histogramas, (4) Colas, (5) Pantallas. Cada una de estas cinco unidades contiene: declaraciones, procedimientos y funciones, que tienen que ver con algún aspecto del programa que se construye. El programa marco lo rellena el usuario programando, con ayuda de la librería, los sucesos condicionales y finales de las actividades, la inicialización y la producción de informes.

Respecto a la comparación entre esta metodología y la que va asociada con SIGEM diremos:

- SIGEM no necesita clasificar los elementos más que en variables de estado (las que necesitan un valor inicial) y de no estado, mientras que eLSE necesita identificar entidades, colas, actividades y sucesos. El tipo de diagrama que exige SIGEM es de puras conexiones tipo influencia/dependencia mientras que el de eLSE representa movimientos de entidades entre colas a través de actividades. El análisis del sistema es, en principio, más sencillo en SIGEM.

- El entorno eLSE solo puede aplicarse a un sistema que pueda acoplarse a un análisis de sus elementos como el que se ha especificado. Es decir, eLSE solo es aplicable para construir programas de Simulación de Sucesos Discretos (DEVS), mientras que SIGEM tiene un espectro de posibilidades mucho más amplio.

- La parte esencial de eLSE es una librería de declaraciones, procedimientos y funciones que el usuario utiliza para construir su programa, facilitándosele así la tarea. El usuario necesita una cierta dosis de entrenamiento en programación. En cambio, SIGEM es un generador de código fuente de proyectos completos, que reduce la intervención del usuario a contestar preguntas del tipo sí ó no y a escribir las ecuaciones y/o tablas y/o reglas que determinan las relaciones por las que cada variable se obtiene a partir de otras. Por consiguiente, SIGEM requiere, en principio, menos entrenamiento en programación y menos trabajo que eLSE.

2.5.2 Comparación entre MATLAB/SIMULINK y SIGEM.

Las diferencias esenciales que hemos encontrado entre estas dos aplicaciones informáticas son las siguientes.

– SIGEM actúa como un experto que dialoga con el usuario y produce unos programas en lenguaje fuente como resultado. Estos programas representan un prototipo del modelo que se está elaborando. MATLAB/SIMULINK es un entorno en el cual el usuario puede encontrar una caja de herramientas que le ayudan a definir, identificar, validar, simular, conectar subsistemas, etc. MATLAB no produce ningún programa independiente.

– En MATLAB el usuario debe conocer un cierto conjunto de comandos y necesita entrenamiento. Con SIGEM no se necesita aprender comandos y el entrenamiento específico requerido es muy pequeño.

– MATLAB trata mayormente con sistemas lineales invariantes con el tiempo, es decir, $x'=Ax+B$, $y'=Cx+D$, siendo x el vector espacio de estado, u el vector de entradas, y es el vector de salidas, y A, B, C, D, son matrices que MATLAB identifica a partir de los datos de campo que se le suministran. En cambio, SIGEM es una herramienta de tipo más general. Puede ayudar al usuario a construir modelos tanto cuantitativos como cualitativos, tanto modelos dinámicos como estáticos, tanto estocásticos como deterministas, tanto continuos como discretos, tanto lineales como no lineales, tanto invariantes con el tiempo como con relaciones internas ligadas al tiempo. Permite al usuario introducir en el modelo relaciones expresadas por medio de tablas, de ecuaciones, de reglas o combinaciones de ecuaciones y reglas. Permite al usuario utilizar en el modelo variables literales, numéricas, deterministas, estocásticas, simples, o indexadas, y entradas constantes o/y variando con el tiempo.

– El acoplamiento de subsistemas en un suprasistema único se puede realizar con ambos paquetes, pero la descomposición automática de un sistema en subsistemas ensamblables de tamaño elegido por el usuario solamente la realiza SIGEM.

– Respecto a sistemas no lineales y a diagramas, MATLAB obliga a la construcción en pantalla de un diagrama de tal complejidad que, como el sistema no sea muy sencillo resulta prácticamente inabordable. Téngase en cuenta que para realizar una simple suma el diagrama necesita 3 iconos. En cambio SIGEM no exige diagrama alguno en pantalla. Los diagramas de SIGEM son ayudas opcionales, muy recomendables para el modelizador, pero pueden ser construidos a mano y presentados en limpio únicamente cuando se trata de explicar el modelo a otras personas.

2.5.3 Comparación entre STELA y/o VENSIM y SIGEM

– STELA y VENSIM son dos entornos. Es decir, los modelos con ellos se producen no pueden salir de ellos (sin pagar “royalties”, se entiende). En cambio, SIGEM permite crear aplicaciones totalmente independientes, únicamente sometidas al lenguaje de programación (Visual Basic) y a una hoja de cálculo para cuestiones estéticas.

– STELA y VENSIM obligan al usuario a manejar un interface gráfico y a dibujar el diagrama de Forrester en pantalla (lo cual en modelos complejos es realmente un problema). En cambio SIGEM no obliga a dibujar ningún diagrama, usa un interface de dialogo (más eficiente, según Zhang et al. (1990)) y permite al usuario trabajar con ficheros de texto (donde es mucho más fácil corregir errores).

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

– STELA y VENSIM están más restringidos (tienen menos posibilidades) que SIGEM en cuanto al tipo de modelos y al tipo de variables a manejar en los mismos.

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

REFERENCIAS

Balci, O., (1986), "Requirements for model development techniques”, Computers and Operations Research 13, N.1, pp 53-67.

Bollen, K.A., 1989, Structural equations with latent variables. John Wiley & Sons. New York
Bunge, M., 1972, Theory and Reality. M. Bunge, ed. Montreal.

Caselles, A., (1984), "A method to compare theories in the light of the General Systems Theory", in: R. Trappl (ed.), Cybernetics and Systems 2. Elsevier S.P.B.V. (North Holland), Amsterdam.

Caselles, A., (1988), "SIGEM: A realistic models generator expert system", in: R. Trappl (ed.), Cybernetics and Systems'88. Kluwer A.P., Dordrecht.

Caselles, A., (1991), "A problem structuring method", in: M.C. Jackson, R.L. Flood, R.B. Blackman, G.L. Mansell and S.V.E. Probert (ed.), Systems Thinking in Europe, Plenum P.C., London.

Caselles, A., (1992a), Simulation of Large Scale Stochastic Systems. In Cybernetics and Systems'92, R. Trappl (Ed.). World Scientific, pp. 221-228.

Caselles, A., (1992b) "Structure and Behavior in General Systems Theory". Cybernetics and Systems: An International Journal, 23 pp. 549-560.

Caselles, A., (1993a) "Systems Decomposition and coupling". Cybernetics and Systems: An International journal" (in press).

Caselles, A., (1994a) "Improvements in the Systems Based Program Generator SIGEM", Cybernetics and Systems: An International journal", 25:81-103.

Caselles, A., (1994b) "Goal-Seeking Systems”, R. Trappl (ed.) Cybernetics and Systems Research'94. World Scientific Publishing Corp. Singapore. pp. 87-94. ISBN: 981-02-1936-9. ISBN 981-02-1761-7 (set).

Caselles, A., (1995) “Systems Autonomy and Learning from Experience”, Advances in Systems Science and Applications. Inauguration Issue. pp. 97-102. ISSN 1078-6236.

Caselles, A., (1996) “Building Intelligent Systems from General systems Theory.” R. Trappl (ed.) Cybernetics and Systems Research'96. Austrian Society for Cybernetic Studies. Vienna. pp. 49-54. ISBN 3 85206 133 4

Caselles, A., (1998) “REGINT: A Tool for Discovery by Complex Function Fitting.” R. Trappl (ed.) Cybernetics and Systems'98. Austrian Society for Cybernetic Studies. Vienna. pp. 787-792. ISBN 3 85206 139 3

Caselles, A., Ferrer, L., Martínez de Lejarza, I., Pla, R., Temre, R. (1999) “Control del desempleo por Simulación.” Editorial: Universitat de València. ISBN: 84-370-4167-8. 300 páginas más un CDROM con programas.

Caselles, A., Temre, R., Martínez de Lejarza, I. (2000), “A Systems Dynamics Model for Unemployment Control”, R. Trappl (Ed.) Cybernetics and Systems'00. Austrian Society for Cybernetic Studies. Vienna. pp. 498-503, ISBN 3 85206 151 2.

Caselles, A. y Romero P.D., 2004 “Aplicación de la dinámica de sistemas al control de la accesibilidad a la vivienda”. Revista Española de Sistemas Vol. 3 Nº 1 pp. 21-66.

Crookes, J.G., D.W. Balmer, S.T. Checa and R.J. Paul, (1986), "A three phase simulation modeling system written in Pascal", Journal of Operations Research Society. 6, 603-618.

Davies, R. and R. O'Keefe, (1989), Simulation Modeling with Pascal. Prentice Hall. New York.

Forrester, J., (1961), Industrial Dynamics. M.I.T. Press.

Forrester, J., (1966), Principles of Systems. M.I.T. Press.

Forrester, J., (1970), Urban Dynamics. M.I.T. Press.

Gelovany, V.A., (1985), "A man-machine simulation system for global development processes", in: J.M. Gvishiani (Ed.), Systems Research II Pergamon Press, London.

Gorokhov, G.V.., (1985), "Development of Systems Engineering Theory", in: J.M. Gvishiani (ed.), Systems Research II, Pergamon Press, London.

Hackstaff, L.H., 1966, Systems of formal Logics. Reidel, ed., Dordrecht.

Klir, G.J., (1985), Architecture of Systems Problem Solving. Plenum Press. New York. Lakatos, I., 1971, In memory of Rudolf Karnap. Reidel ed., Dordrecht.

Linstone H.A., and Turoff, M., 1975, The Delphi Method: Techniques and Applications. Addison Wesley. Reading. Mass.

Mathewson, S.C., (1989), "The implementation of Simulation languages". In M. Pidd (ed), Computer Modelling and Simulation. John Wiley & Sons Ltd., Chichester.

Melése, J., 1976, La Gestion par les Systémes. Dunod. Paris

Meyer, J.A., S. des Clers and Chahuneau, F.(1979), "La Simulation Numerique des systemes Complexes: object et Techniques", in: Actes du Colloque Elaboration et Justification des Modeles. Maloine S.A., Paris.

Micó, J.C. and Caselles, A. (1998), “Space-Time Simulation Models for Social Systems”, R. Trappl (ed.) Cybernetics and Systems'98. Austrian Society for Cybernetic Studies. Vienna. pp. 486-491, ISBN 3 85206 139 3

Micó, J.C., Caselles, A., Ferrer, L., Soler, D. (2002) “The Multidimensional Approach in Systems Dynamics, Modeling Urban Systems”. Journal of Applied Systems Studies, Vol. 3, No. 3, pp. 644-655.

Miller, D., (1976). Verisimilitude Redeflated. The British Journal for Philosophy of Science 27(4):363-380.

Miller, J.G., (1978). Living Systems. McGraw-Hill. New York.

Morecroft, J.D.W., (1982), A critical Review of diagramming tools for conceptualizing feedback system models. Dynamics, vol. 8, part 1, pp 20-29.

Popper, K.R., 1972, Objective Knowledge. The Clarenton Pr. Oxford.

Popper, K.R., 1976, A note on Verisimilitude. The british Journal for Philosophy of Science. Sage, A.P., (1977), Methodology for Large Scale Systems, McGraw Hill. N.Y.

Simonot, F., LeDoeuf, R., Haddad, S., Benkhoris, M.F., (1990), Rev. Gen. Electr. (France) Vol 1, pp 5-8.

Solé, R.V. y Manrubia, S.C., 1996, Orden y caos en sistemas complejos. Universidad Politénica de Cataluña. Barcelona.

Standbridge, C.R., (1985), Performing simulation Projects with extended simulation system (1ESS)", Simulation, 45, 283.

Von Bertalanffy, L., 1972, History and Situation of General Systems theory. In Klir, G.J. (ed.), J. Wiley & Sons Inc.

Wartofsky, M.W. 1968, Conceptual Foundations of Scientific Thought. An Introduction to Philosophy of Science. Mac Millan ed.

National Journal. 23:549-560.

Yang, Z., 1989, New model of General Systems Theory. Cybernetics and Systems: An International Journal, 20: 67-76.

Zeigler. B.P., (1984), Multifaceted Modeling and Discrete Event Simulation, Academic Press, London.

Zeigler, B.P., (1987), "Hierarchical, Modular, Discrete Event Modeling in an Object Oriented Environment, Simulation 47:5.

Zeigler, B.P., (1989), "DEVS Representation of Dynamical Systems: Event-Based Intelligent Control". Proceedings of the IEEE, Vol.77, No.l, pp. 72-80.

Zeigler, B.P., (1990), "Object Oriented Simulation with Hierarchical, Modular, Models". Academic Press. London.

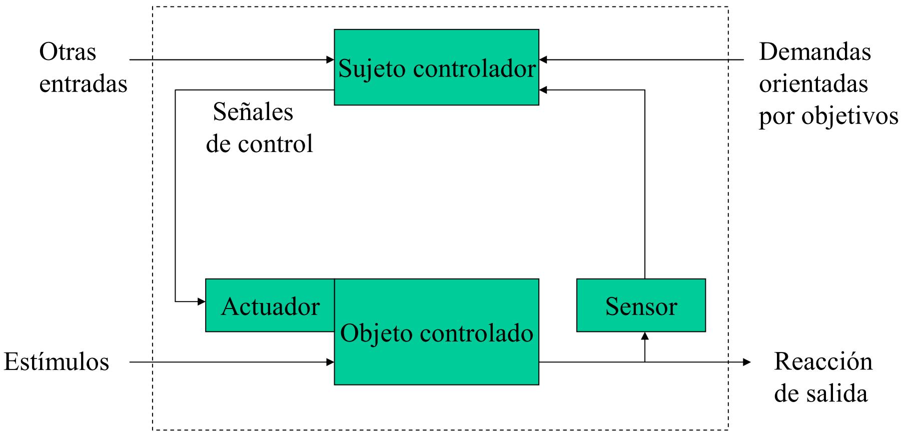
Zhang, S.H., Schroer, B.J., Messimer, S.L., and Tseng, F.T., (1990), "Software Engineering and Simulation". Third Inc. Sof. For Strat. Synth. Conf. Proc., pp. 33-42. University of Alabama.

Apéndice 1

Sistemas cibernéticos, sistemas con objetivos y sistemas vivos

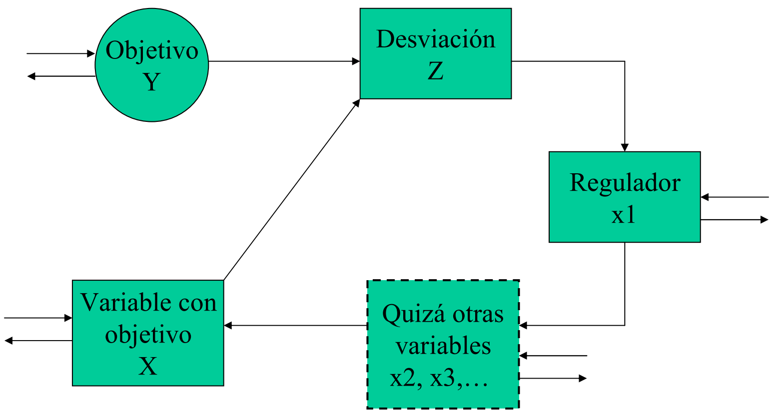
Cibernética: ciencia y tecnología de la comunicación y el control en los sistemas vivientes y en las máquinas.

Diagrama de un sistema autocontrolado



Antonio Caselles. Universitat de València. España 2

Diagrama de una variable con objetivo y otras variables conexas



Antonio Caselles. Universitat de València. España 3

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Sobre la Teoría de los sistemas vivos de Miller.

Miller (1978) elabora toda una teoría de los sistemas vivos cuya parte más interesante para nosotros se resume a continuación. Se trata de un esquema general o plantilla en la que encajar un posible modelo de un sistema viviente, desde una célula hasta el mundo entero. No quiere esto decir que obligatoriamente haya que ajustarse a este esquema cuando se modeliza un sistema viviente. Que esto se haga o no dependerá de los objetivos para los que se construya el modelo. En todo caso, es una ayuda para no olvidar algún aspecto importante.

Sistema viviente.

Es un sistema abierto y con objetivos compuesto de subsistemas que procesan entradas, salidas y circulaciones, de varias formas de materia, energía e información.

Se agrupan o agregan en siete niveles de integración o suprasistemas y se desagregan en diecinueve subsistemas esenciales para la vida.

En cada uno de los subsistemas hay que distinguir tres aspectos fundamentales:

a) **estructura** (órganos o entes, parte física y sus relaciones de influencia o dependencia);
b) **comportamiento** o función (normas, restricciones, etc.);
c) **control** (objetivos, decidores, normas de decisión).

Los 19 subsistemas esenciales para la vida.

1. **Reproductor.** Subsistema capaz de producir sistemas semejantes a aquel en el que se encuentra. Es esencial para la especie pero no para el individuo.
2. **Frontera.** Está en el perímetro del sistema. Mantiene juntos a los componentes del sistema. Permite o no determinados intercambios de materia, energía e información.
3. **Ingestor.** Introduce materia y/o energía en el sistema a través de la frontera desde el entorno.
4. **Distribuidor.** Reparte las entradas al sistema entre sus componentes y las salidas de los subsistemas entre sus componentes.
5. **Convertidor.** Convierte ciertas entradas del exterior en formas más útiles al sistema.
6. **Productor.** Forma asociaciones estables (por periodos significativos) entre las entradas de materia-energía al sistema o entre las salidas del convertidor, facilitando energía para el movimiento o constituyendo salidas para el suprasistema.
7. **Almacenador.** Guarda materia-energía por diferentes periodos de tiempo.
8. **Excretor.** Saca materia-energía del sistema como productos o desperdicios.
9. **Motor.** Mueve el sistema o sus partes en relación con el entorno, o mueve partes del entorno con relación a él.
10. **Sostén.** Mantiene la adecuada disposición espacial entre los componentes del sistema con el fin de evitar que interactúen con peso o se amontonen estorbándose.
11. **Traductor de entradas.** Introduce en el sistema a los portadores de información y los convierte en otras formas de materia-energía más adecuadas para el sistema.
12. **Traductor interno.** Recibe indicadores de alteraciones significativas en otros componentes del sistema y las convierte en otras formas de materia-energía mejor transmisibles.

13. **Canal y red.** Subsistema compuesto por una simple ruta o red de ellas por las cuales circulan los indicadores hacia cualquier parte del sistema.
14. **Descodificador.** Es el subsistema que convierte el código de la información que entra o atraviesa el sistema en código privado del sistema.
15. **Asociador.** Asocia informaciones relacionadas (el primer paso del aprendizaje).
16. **Memoria.** Almacena información durante diversos periodos de tiempo (el segundo paso del aprendizaje).
17. **Decisor.** Recibe información de todas las partes del sistema y les devuelve información de control.
18. **Codificador.** Traduce código privado de otro subsistema a código público para otros sistemas del entorno.
19. **Traductor de salidas.** Convierte indicadores internos del sistema en otras formas de materia-energía que pueden ser enviadas a través de canales del entorno.

Los 7 suprasistemas.

Los 19 subsistemas esenciales para la vida se mantienen y se van complicando en siete niveles de integración que son los siguientes:

1. **Célula.**
2. **Órgano.**
3. **Organismo.**
4. **Grupo.**
5. **Organización.**
6. **Sociedad.**
7. **Sistema supranacional de sociedades o de organizaciones.**

Los sistemas con un nivel de agregación menor que 1, es decir, los sistemas no vivos, podrían agregarse como sigue:

Partículas → átomos → moléculas → cristales y orgánulos (virus por ejemplo)

Los sistemas con nivel de agregación mayor que 7 incluyen sistemas vivos y no vivos, por ejemplo:

Ecosistemas → planetas → sistemas solares → galaxias → … → universo

Al ascender en el nivel de agregación aparecen propiedades nuevas en el sistema que no estaban en ninguno de los subsistemas agregados. A este fenómeno se le conoce con el nombre de “**emergencia**”. Por ejemplo, al agregar hidrógeno y oxígeno se forma agua, cuya molécula tiene unas propiedades que no tienen ni el hidrógeno ni el oxígeno.

Apéndice 2

Nociones de Visual Basic 6

Visual Basic 6.0 (en adelante VB) es un lenguaje de programación de tipo general que resulta relativamente fácil de aprender. Aunque los programas elaborados con él resulten más lentos que los elaborados con el lenguaje C o el C++, para los casos en que el volumen de cálculo no es demasiado grande, este lenguaje resulta ventajoso.

Un programa contiene una entrada de datos, un algoritmo que los procesa y una salida de resultados. Los programas almacenan y procesan los datos (numéricos o no) en “variables”. Una variable es un nombre que designa una zona de la memoria del ordenador donde se almacena un dato (variable escalar) o una lista de datos (variable, vectorial, matricial, etc.). A cada variable se le da un nombre (generalmente corto, de unos pocos caracteres). Un algoritmo es un conjunto de operaciones ordenadas que producen un resultado. Estas “operaciones” pueden ser de varias clases: de control, de repetición, etc. Una operación consta de una o de varias frases o “sentencias” que van separadas por el signo “:” o por un salto de carro. Se pueden introducir comentarios, útiles al programador y que no forman parte del algoritmo, precediendolos por la palabra REM o por una comilla simple “ ’ ”. Un programa escrito en un lenguaje de alto nivel como el VB puede ser “interpretado” o “compilado”. La interpretación la hace VB trabajando con el programa que nosotros hemos escrito o “programa fuente” (*.bas). La compilación (que también la puede hacer VB) consiste en la traducción del programa fuente a un lenguaje que entiende solo el ordenador o “lenguaje de máquina” produciendo un programa “ejecutable” (*.exe) desde el sistema operativo sin necesidad de la presencia del VB.

Nombres de las variables

El nombre de una variable (o de una constante) tiene que comenzar siempre por una letra y puede tener una longitud hasta 255 caracteres. No se admiten espacios o caracteres en blanco, ni puntos (.), ni otros caracteres especiales.

Los caracteres pueden ser letras, dígitos, el carácter de subrayado (_) y los caracteres %, &, #, !, @, y $. El nombre de una variable no puede ser una palabra reservada del lenguaje (For, If, Loop, Next, Val, Hide, Caption, And, etc.). Para saber cuáles son las palabras reservadas en VB puede utilizarse el Help de dicho programa, buscando la referencia Reserved Words. De ordinario las palabras reservadas del lenguaje aparecen de color azul en el editor de código, lo que hace más fácil saber si una palabra es reservada o no.

A diferencia de otros lenguajes de programación, VB no distingue entre minúsculas y mayúsculas. La declaración del tipo (entera, real, alfanumérica, etc.) de una variable o la primera vez que se utiliza determinan cómo se escribe en el resto del programa. El tipo de una variable se declara de manera explícita o se determina por la primera letra de su nombre de acuerdo con unas reglas determinadas.

Tipos de variables

Las variables pueden ser:

1. Escalares
2. Vectoriales
3. Matriciales

Y también:

a. Numéricas: enteras, reales de simple precisión, de doble precisión, etc.
b. Cadenas de caracteres (alfanuméricas)

Tipos de Operadores

Los operadores sirven para combinar las variables en una fórmula. La siguiente tabla representa los distintos tipos de operadores de VB y su función.

Tipos de operador	Operación que realiza	Signo que lo representa
Aritméticos	Exponenciación	^
	Cambio de signo	-
	Multiplicación y división	* , /
	División entera	\
	Resto de una división entera	**Mod**
	Suma y resta	+ , -
Concatenación	Concatenar o enlazar cadenas de caracteres	& , +
Asignación	Asigna un valor a una variable	=
Relacionales	Es igual a	=
	Es distinto de	<>
	Es menor que , es menor o igual que	< , <=
	Es mayor que , es mayor o igual que	> , >=
Otras comparaciones	Comparar dos expresiones de caracteres	**Like**
	Comparar dos referencias a objetos	**Is**
Lógicos	Negación	**Not**
	And	**And**
	Or inclusivo	**Or**
	Or exclusivo	**Xor**
	Equivalencia (opuesto a Xor)	**Eqv**
	Implicación (*False* si el primer operando es *True* y el segundo operando es *False*)	**Imp**

Sentencias de control

Las sentencias de control o estructuras de control, permiten tomar decisiones y realizar un proceso repetidas veces. Son las bifurcaciones y los bucles. VB dispone, entre otras, de las siguientes estructuras de control:

- Condicionales
  - If ... Then ... Else ...
  - Select Case
- Bucles de repetición
  - For ... Next
  - Do ... Loop
  - While … Wend

La estructura If ... Then ... Else ...

Esta estructura permite ejecutar condicionalmente una o más sentencias y puede escribirse de dos formas. La primera ocupa sólo una línea y tiene la forma siguiente:

If condicion **Then** sentencia1 [**Else** sentencia2] *(el corchete significa que es opcional)*

La segunda es más general y se muestra a continuación:

If condicion **Then**
sentencia(s)
[Else
sentencia(s)**]** *(el corchete significa que es opcional)*
End If

Si el valor de *condicion* es *True (verdadero)*, se ejecutan las sentencias que están a continuación de *Then*, y si el valor de *condicion* es *False (falso)*, se ejecutan las sentencias que están a continuación de *Else*.

Para indicar que se quiere ejecutar uno de varios conjuntos de sentencias dependientes cada uno de ellos de una condición, escribiremos:

If condicion 1 **Then**
Sentencias 1 *(para cuando se cumple la condición 1)*
ElseIf condición 2 **Then**
Sentencias 2 *(para cuando se cumple la condición 2)*
ElseIf condicion … **Then**
Sentencias …
Else
Sentencias n *(para el caso en que no se cumpla ninguna de las condiciones)*
End If

La estructura Select Case

Esta estructura permite ejecutar una de entre varias acciones (conjuntos de sentencias) en función del valor de una expresión. Es una alternativa a If ... Then ... ElseIf útil cuando se compara la misma expresión con diferentes valores posibles. Su forma general es la siguiente:

Select Case *expresión*
Case *H1*
[sentencias1]
Case *H 2*
[sentencias2]
…………………
Case Else
Sentencias para cuando no estamos en ninguno de los casos previstos
End Select

Dónde, según que el valor de *expresión* coincida con *H1*, *H2*, ... se ejecutarán los distintos conjuntos de sentencias. *H1*, *H2*, ... , pueden tomar de las formas siguientes:

1. *expresion*
2. *expresion* **To** *expresion*
3. **Is** operador-relacional *expresion*

4. combinación de las formas anteriores separadas por comas

Bucles de repetición

Un bucle repite la ejecución de un conjunto de sentencias mientras una condición dada sea cierta, o hasta que una condición dada sea cierta. La condición puede ser verificada antes o después de ejecutarse el conjunto de sentencias. Sus posibles formas son las siguientes:

Bucle For … Next

```
For i=1 To n [step p]   (Se ejecutarán las Sentencias mientras el contador i esté entre 1 y n)
  Sentencias            (El contador i avanza de 1 en 1, o de p en  p si se especifica step p)
Next
```

Bucle Do … Loop

```
Do [While/Until condicion]     (Se ejecutarán las Sentencias mientras (While) o hasta
   [sentencias]                  (Until) que la condición se cumpla.)
   [Exit Do]                     (Exit Do permite salir del bucle sin terminarlo.)
   [sentencias]
Loop
```

```
Do                              (Con este formato la condición se comprueba al final.)
   [sentencias]
   [Exit Do]
   [sentencias]
Loop [While/Until condicion]
```

Bucle While ... Wend

```
While condicion                 (Se ejecutarán las Sentencias mientras se cumpla la condición.)
   [sentencias]
Wend
```

Funciones

Un gran número de funciones están pre-programadas dentro del VB. Por ejemplo:

Función	En VB	Función	En VB
Valor absoluto	Abs(x)	Nº aleatorio entre 0 y 1	Rnd
Arco tangente	Atn(x)	Seno y coseno	Sin(x) , Cos(x)
Exponencial (e^x)	Exp(x)	Tangente	Tan(x)
Parte entera	Int(x), Fix(x)	Raíz cuadrada	Sqr(x)
Logaritmo neperiano	Log(x)	Signo (1, 0, -1)	Sgn(x)
Redondeo a ndec decimales	Round(x, ndec)		

Las funciones trigonométricas de VB utilizan radianes para medir los ángulos.

Ejemplos:

Ejercicio con bucles:

1. Calcular la media aritmética dada una tabla de frecuencias.

   Suponemos que X_i y N_i están en la memoria como vectores.

i	X_i	N_i
1	3	4
2	5	7
3	9	18
4	12	6
5	20	4

```
M = 0: For i = 1 To 5: M = M + Xi(i) * Ni(i) : Next
N = 0 : For i = 1 To 5 : N = N + Ni(i) : Next
M = M / N
```

2. Calcular probabilidades con la distribución de Poisson.
   $P=(\lambda^r/r!)\cdot e^{-\lambda}$
   Supuestos conocidos r y λ, a quien llamaremos L, primero calcularemos r! y la llamaremos RF, y después P.

```
RF= 1 : For i = 2 To r : RF = RF * i : Next
P = L ^ r / RF * exp(-L)
```

Ejercicios con Condicionales

Dado un vector V de 200 números enteros, como por ejemplo
V = (1, 3, 8, 14, 27, 84, 86, 125, 143,…),

1. Escribir los pares en otro vector V1.
2. Escribir los pares en V1 y los impares en V2.
3. Escribir los pares en V1, los múltiplos de 3 en V2, los de 5 en V3, y el resto en V4.
4. Análogamente los múltiplos de 4, los múltiplos de 4 + 1, los múltiplos de 4 + 2, etc., pero escribiendo los resultados en una matriz de nombre Mul en lugar de en vectores de diferente nombre.

El caso 1 consiste en un condicional dentro de un bucle. Como con los impares no hay que hacer nada, el algoritmo quedaría de esta manera:

```
for i = 1 to 200                    '(caso 1)
    j = 1
    If V(i) Mod 2 = 0 Then
        V1(j) = V(i): j = j+1
    Endif
next
```

El caso 2 incluye lo que hay que hacer con los impares.

```
For i = 1 To 200                    '(caso 2)
    j = 1 : k = 1
    If V(1) Mod 2 = 0 then
        V1(j) = V(i): j = j+1
```


© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

```
        Else
         V2(k) = V(i) : k = k + 1
      Endif
   Next
```

El caso 3 se resolvería con If, ElseIf, ElseIf, etc.

```
For i = 1 To 200                    '(caso 3)
   j=1: k=1: m=1: n=1
   If V(i) Mod 2 = 0 Then
         V1(j) = V(i): j = j+1   '(para cuando el número es par)
   ElseIf  V(i) Mod 3 = 0 Then
         V2(k) = V(i): k = k+1  '(para cuando es múltiplo de 3)
   ElseIf V(i) Mod 5 = 0 Then
         V3(m) = V(i): m = m+1  '(para cuando es múltiplo de 5)
   Else
          V4(n) = V(i): n= n+1  '(para cuando no es múltiplo ni de 2, ni de 3 ni de 5))
   End If
Next
```

El caso 4 podría resolverse como el caso 3 pero lo haremos con Select Case.

```
For i = 1 To 200                    '(caso 4)
   k1=1: k2=1: k3=1: k4=1
   Select Case V(i) Mod 4
     Case 1
       Mul(1,k1)=V(i): k1=k1+1
     Case 2
       Mul(2,k2)=V(i): k2=k2+1
     Case 3
       Mul(3,k3)=V(i): k3=k3+1
     Case Else
       Mul(4,k4)=V(i): k4=k4+1
   End Select
Next
```

Para más ejemplos y casos posibles se recomienda consultar la ayuda de Visual Basic 6.

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Apéndice 3

Métodos aplicables ante la escasez de datos históricos

PROSPECTIVA

INTRODUCCION

El objetivo de la prospectiva es anticipar el futuro de forma fiable. La Prospectiva trata de aplicar el método científico dentro de su metodología. Se basa en que **el futuro se construye** y, si sabemos quiénes y cómo, podremos simular el proceso y anticipar los detalles.

En realidad Prospectiva y Estrategia son casi lo mismo, varían en el punto donde se hace énfasis. La Prospectiva hace énfasis en la anticipación de acontecimientos y la Estrategia en quienes y como preparan estos acontecimientos. Es obvio que se necesita un modelo del comportamiento del sistema con el que se trabaja, por consiguiente podría considerarse como integrada en la Teoría de Sistemas en lo que hace referencia a construcción de sistemas nuevos o a la preparación de los datos (valores de las variables de entrada) para efectuar simulaciones.

Tradicionalmente, (desde los años cuarenta en que apareció) ha tenido un carácter independiente considerándose como la ciencia (?) que estudia los métodos para anticipar o pre-ver los futuros posibles y el camino que conduce a ellos en cada situación determinada. Hoy día, como hemos apuntado antes, quedaría reducida a algunos métodos para preparar datos para simular con ellos y totalmente absorbida por la Teoría de Sistemas.

LOS METODOS DE LA PROSPECTIVA

En realidad son métodos para construir modelos y simular con ellos. Se decía que tienen una parte descriptiva, una parte predictiva y una parte normativa. Efectivamente, primero hay que describir el sistema (construir el modelo), después utilizarlo para hacer predicciones y por último sacar consecuencias sobre lo que hay que hacer para lograr el futuro apetecido. Todo ello, corno puede verse, forma parte de la metodología sistémica general.

Teníamos métodos "sintéticos" (cualitativos y de tipo "caja negra"); métodos "semi-analíticos" (con cuantificaciones basadas en opiniones de expertos y descripciones poco detalladas) y métodos "analíticos" (que usan descripciones detalladas y cuantificaciones basadas en datos observados). Ahora se habla de metodologías sistémicas de tipo "soft" y de tipo "hard" con sentido análogo. Por consiguiente, queda claro que tenemos que construir un modelo (o modelos) mental o computarizado, cualitativo, cuantitativo o "semi"; y después simular con él (o ellos); para sacar consecuencias que nos conduzcan a la elaboración de un programa de actividades interventoras sobre el sistema distribuidas a lo largo del tiempo.

Entre los métodos prospectivos destacamos los siguientes:

- Métodos sintéticos:

  Brainstorming, Brainwriting y Delphi.

- Métodos semi-analíticos
  - Construcción de escenarios
  - Cross-impact
    - Con sucesos y/o con tendencias.
    - Monoperiodo o multiperiodo.
    - Con impactos estimados o con impactos obtenidos por regresión.
- Métodos analíticos:
  - Series temporales: Regresión simple o múltiple, lineal e no lineal.
  - Medias móviles.
  - Auto regresión (ARIMA ó técnica de Box-Jenkins).
  - Arboles de decisión.
  - Simulación con modelos dinámicos: utilización de modelos cuantitativos con muchas variables interrelacionadas, conectados con la construcción de escenarios y técnicas de tipo estocástico.
  -

¿Qué no debemos olvidar al utilizar estos métodos?

(a) El futuro deseado nos hace determinar el presente, teniendo en cuenta nuestros conocimientos. El futuro no es único y cierto.

(b) Los inconvenientes tradicionales de la previsión son:

- El efecto anuncio: incitativo o disuasivo
- La insuficiencia de información y la inexactitud de los datos
- La inestabilidad, la parcialidad y la evolución de los modelos
- La exclusión de variables no cuantificables
- La validación solamente con el pasado
- La elaboración excesiva de los modelos (charlatanería matemática)
- Los errores de interpretación

BRAINSTORMING, BRAINWRITING, DELPHI.

Son perfeccionamientos del grupo de discusión tradicional que tratan de corregir sus defectos y además de adaptarse a la separación espacial de los miembros del grupo. Recordemos los inconvenientes del grupo de discusión:

(a) La excesiva influencia de algunos miembros (prestigio, oratoria, etc.)

(b) Presión hacia la conformidad (mayoría, cansancio, temor,...)

(c) Los temas ajenos o irrelevantes (noticias del día, interés personal,...)

(d) Los factores psicológicos (susceptibilidades, ansias de relevancia,...)

Obviamente un grupo de discusión tradicional cuenta con la experiencia de sus miembros, lo cual en lenguaje sistémico serían modelos mentales de subsistemas del sistema objeto de estudio. Modelos que serían de tipo cualitativo y muy poco detallados, normalmente. La mayoría de las veces serían modelos de caja negra aunque a veces podrían tratar de determinarse explícitamente las repercusiones en cadena de determinados sucesos y/o acciones:

El **brainstorming** se caracteriza por:

- Elementos: moderador, secretario, pizarra, grupo de expertos.
- Procedimiento:
  - · 1ª Fase: el moderador presenta objetivos y método de trabajo. El grupo aporta ideas (en dos o tres palabras y permaneciendo en silencio). El secretario apunta en la pizarra.
  - · 2ª Fase: cada idea de la lista (una tras otra) es evaluada con una tabla de pros y contras, seleccionándose la mejor opción según el balance de los mismos.

El **brainwriting** es lo mismo pero por escrito. Caben variantes intermedias.

El **Delph**i es lo mismo pero por correo. Normalmente el proceso se lleva a cabo con cuatro cuestionarios que elabora y analiza el equipo director, cada uno basándose en las respuestas al anterior y ofreciendo además el balance de respuestas de los demás miembros del grupo. Obsérvese que con este método:

- Se obtienen respuestas simultáneas y anónimas, evitándose con ello los inconvenientes (a) y (b).
- Se permite la interacción entre los miembros y el retorno de las conclusiones parciales, evitando el diálogo irrelevante (c) y los problemas derivados del contacto entre las personas (d).

EL METODO DE LOS ESCENARIOS

Tiene su origen en Herman Khan y Norbert Wiener que lo utilizaron en el informe del Hudson Institute de 1967, Lo llamaron "scenario writing method".

Este método intenta establecer una sucesión lógica de acontecimientos (escenario) con el fin de mostrar cómo a partir de una situación dada o actual es posible evolucionar hacia una situación futura (16). En principio este método no pretende prever el futuro sino mostrar, a modo de simulación experimental, como puede una realidad ser posible a partir de una situación dada, en un contexto dado. Cuenta pues con unas hipótesis de partida. El mismo Kahn avisa centra el riesgo de pensar que tales hipótesis de partida son lo suficientemente correctas como para pensar que los escenarios que se construyen apoyados en ellas van a tener una parte substancial de realidad. Así pues sus inventores presentan este método sólo como "un modo de hacer salir el pensamiento de allí donde se cuece" ilustrando y dramatizando las posibilidades sobre las que aquel se concentra. Además obliga al analista a ocuparse de detalles y aspectos dinámicos que podrían muy bien dejar pasar si se limitasen a hacer consideraciones abstractas.

Una aplicación que perfeccionó determinados aspectos del método fue el estudio Francia 2000 del grupo DATAR, (órgano de la Administración francesa) publicado en 1970. Dejando aparte las hipótesis de partida de este estudio, algunas evidentemente falsas, consideramos qué representó un avance considerable en la puesta a punto del método (17).

El método que sigue Francia-2000 consiste, en síntesis, en lo siguiente:

1) Determinar las cuestiones a estudiar.- Considerando a Francia corno un sistema, se trata de determinar los elementos del sistema. Esto lo hace por sucesivas divisiones en bloques (sociedad urbana, sociedad rural, sociedad agrícola y sociedad industrial) y sub-bloques (fuerzas de producción, modos de producción, relaciones de producción, instituciones y espacio acondicionado). Y por fin, en cada sub-bloque distingue una serie de elementos, cuestiones a estudiar o "componentes".

2) Determinar las relaciones entre las componentes.-Se trata de determinar cómo repercute sobre cada una de ellas lo que les pueda ocurrir a las demás y cómo pueden intervenir en tal mecanismo las estrategias de los "actores", (entes que ejercen algún control sobre el sistema). El modo de llegar a ello es el estudio del pasado de las componentes y de su evolución hasta el presente a través de sucesivos análisis sincrónicos (en unas fechas determinadas) y diacrónicos (entre cada tina de esas fechas y la siguiente).

Durante la serie de análisis sincrónicos y diacrónicos del pasado, cuyo objetivo es captar los mecanismos que rigen el funcionamiento del sistema, creando los modelos mentales correspondientes, al tiempo que una base de partida de cara al futuro, procede tratar de detectar:

- Tensiones. Diferencias entre lo existente y lo deseable. Pueden aparecer entre una componente o entre dos o más de ellas. Son puntos de conflicto que hacen necesaria la intervención de los mecanismos reguladores del sistema. Es procedente, cuando una tensión ha sido detectada, investigar cómo ha sido corregida, para de este modo descubrir los mecanismos reguladores y por tanto los actores.
- Tendencias pesadas. El estado de cada componente ha evolucionado con el tiempo. Seguro que ha tenido pequeñas oscilaciones de corta duración. Pero por encima y a pesar de estas pequeñas oscilaciones ha seguido una tendencia a la que por su gran inercia se llama "pesada". Cuando una tendencia pesada ha sido detectada procede tratar de explicar a qué obedece.
- Mutaciones. Se llama así a la aparición o desaparición de componentes o rupturas de la tendencia pesada que las anima. Se las debe encontrar una explicación.
- Gérmenes de mutación. De la explicación atribuida a determinadas mutaciones observadas debe obtenerse el criterio necesario para detectar lo que puede dar lugar a una mutación en el futuro, si se dan determinadas circunstancias.

3) Como consecuencia de lo anterior se debe de disponer, de una especie de tabla de doble entrada cuyos encabezamientos sean las componentes estudiadas, en su estado actual, y cuyas casillas sean los, impactos, expresados de forma verbal, que unas componentes tienen sobre las otras.

Ahora procede hacer la proyección hacia el futuro. Esta proyección puede hacerse de modo "tendencial" de tal manera que con sucesivos análisis sincrónicos y diacrónicos queda dibujada la tendencia de la evolución del estado de las componentes. Y un análisis, sincrónico sería el estado de las componentes en una fecha Y un análisis diacrónico sería el proceso de acumulación de impactos de todas las componentes sobre cada una de las demás de tal manera que, de resultas de esta acumulación de impactos quede determinado el estado de cada componente al final del período que se está considerando.

4) Esta imagen "tendencial" debe servir de base para aplicar en ella políticas determinadas en los puntos en los que sea necesaria una regulación (tensiones, tendencias pesadas mutaciones y gérmenes de mutación). Aplicando una determinada estrategia, entendida esta come una serie de decisiones o elecciones que tienden al logro de un objetivo, al escenario tendencial se obtiene un escenario "semi-contratado". Con varias estrategias alternativas se obtendrían varios escenarios semi-contrastados el mejor de les cuales se tomaría como escenario contrastado con la política (objetivo) seguida. Esto se hace así porque el escenario ideal contrastado con

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

determinada política puede no ser alcanzable.

El método parece prometedor si se piensa en temar en cuenta además lo que puede suceder en el universo: exterior al sistema que se considera. Esto no parece muy difícil en principio. Bastaría para ello identificar les elementos del universo exterior que pueden incluir sobre el sistema y hacer hipótesis sobre ellos. Cada combinación entre estas hipótesis daría origen a un escenario tendencial. Y si se desea un mayor detalle en lo que respecta a este Universo exterior se le puede aplicar alguna versión del cross-impact. Nosotros hemos tratado de llevar adelante esta metodología aplicándola a la Comunidad Valenciana y nuestras conclusiones son:

1) Los sucesivos análisis sincrónicos y diacrónicos a los que se alude, si se quieren hacer con carácter exhaustivo, son de una laboriosidad tal que resulta desbordante.
2) Cuando les impactos entre componentes no se cuantifican sino que se describen verbalmente, siendo además el número de componentes de una cierta cuantía (más de seis), la acumulación de los mismos con ayuda exclusiva de la mente resulta prácticamente imposible. Y si esto falla, todo lo que viene a continuación se queda en el plano teórico por irrealizable.

Está claro que puede no hacerse de modo tan exigente, pero entonces el método se convierte en un puro juego especulativo que solo tiene el valor que le asignaron sus inventores en un principio, o poco más, a pesar de su mayor laboriosidad.

En realidad, el método puede llevarse a buen término con ayuda de un ordenador, pero en este caso estaríamos hablando del cross-impact multiperíodo con tendencias, que se describe más adelante. Así pues, lo que hace el cross-impact multiperíodo con tendencias es en realidad una construcción de escenarios, solo que con posibilidades de optimización más claras que en la metodología del Francia-2000. Recuérdese que la optimización aquí se basa en la construcción de escenarios semi-contrastados por aplicación de estrategias determinadas sobre un determinado escenario tendencial, para posteriormente seleccionar "el mejor" entre ellos. En el cros-impact los reguladores y las variables objetivo actúan igual que las demás tendencias, gracias a lo cual el ensayo de "políticas" alternativas se hace de modo más claro y rápido.

Una visión ligeramente distinta del método de los escenarios la da Lesourne (1979). Según Lesourne un escenario es una combinación de hipótesis sobre valores futuros de las varíales de entrada a un modelo, así como los correspondientes valores de las variables de salida. Describe una posible realidad futura y está destinado a iluminar la acción del presente.

Los escenarios se construyen:

1. Construyendo primero el modelo.
2. Identificando las variables de entrada cuyos valores no se

pueden asignar más que como hipótesis.

3. Estableciendo unas pocas hipótesis para cada variable.
4. Dando una probabilidad subjetiva a cada hipótesis.
5. Calculando las probabilidades de las diferentes combinaciones de hipótesis.
6. Corriendo el modelo para el conjunto de las combinaciones de hipótesis más probables (las que, entre todas, abarquen el 90 por 100 de la probabilidad, por ejemplo).

En un principio, se intentó trabajar exclusivamente con modelos mentales ilustrados con descripciones verbales y numéricas de la realidad actual y pasada. La experiencia demuestra que los modelos mentales corren con dificultad si se manejan más de tres o cuatro variables interrelacionadas. No obstante, la construcción de modelos mentales, parte esencial del primitivo método de los escenarios (H. Kahn y el Hudson Institute, y el grupo DATAR francés), puede ser muy útil para adquirir un criterio para evaluar los resultados de los modelos cualitativos y/o cuantitativos computerizados; así como para adquirir el conocimiento suficiente para desarrollar un modelo computerizado.

La adquisición de la "base" para poder construir el modelo se realiza estudiando el pasado y la situación actual de manera:

(a) detallada y en profundidad en el plano cualitativo y cuantitativo

(b) global (económica, tecnológica, política, ecológica...)

(c) dinámica (poniendo en evidencia las tendencias pasadas y los hechos portadores de futuro.

(d) explicativa de los mecanismos de evolución del sistema.

Con ello se debe haber logrado:

- delimitar el sistema o los elementos pertinentes;
- estructurar el sistema y eventualmente desglosarlo en subsistemas;

- explicar la evolución pasada y el estado actual;
- poner en evidencia los factores de evolución o de estabilidad;
- localizar a los "actores" y sus proyectos en un cuadro estratégico;
- cada actor ha sido relacionado con unos objetivos estratégicos;
- se han evaluado las posibles alianzas y conflictos;
- se han identificado los cuestiones clave para el futuro.

Un modo muy adecuado para establecer las hipótesis que darán lugar a los escenarios es identificar las "dimensiones" o aspectos, del problema y recurrir a análisis morfológico (un árbol de combinaciones posibles). Veámoslo con un ejemplo. El estudio "Interfuturos" (Lesourne, 1979) de prospectiva mundial distingue cuatro dimensiones en el problema:

- Relaciones entre países desarrollados, con dos opciones: gestión colegiada o fragmentación parcial entre los polos.
- Dinámica interna de las sociedades desarrolladas, con tres opciones: consenso (crecimiento fuerte), cambios de valores rápidos y crecimiento moderado y conflictos entre grupos sociales y crecimiento moderado.
- Evolución de las productividades relativas, con dos opciones: convergencia y divergencia
- Relaciones Norte-Sur y entre países en vías de desarrollo, con tres opciones: amplio incremento de intercambios, acentuación de las divisiones, y fragmentación del Sur por regiones y conexión con los países desarrollados.

Combinando las opciones de cada dimensión salen 36 escenarios posibles. Interfuturos centra su análisis en algunos de estos escenarios, los más probables ("tendenciales") o los que interesan a algunos de los actores ("contrastados").

Cuando las posibilidades de evolución del problema estudiado se caracterizan por el cumplimiento o no de *n* hipótesis o "sucesos" fundamentales se obtienen *2n* escenarios posibles. Asignando una probabilidad simple a cada suceso y estudiando el efecto o impacto que tendría la ocurrencia de cada suceso sobre las probabilidades de los demás se puede llegar a dar una probabilidad a cada escenario y seleccionar los más-probables. El método que se utiliza para ello es el de impactos cruzados (cross-impact).

EL METODO DE LOS IMPACTOS CRUZADOS

Este método surgió cuando en los años 60 se trató de incorporar a los métodos de previsión existentes la posibilidad de que la ocurrencia de un suceso pudiera afectar a la de los que todavía estaban por ocurrir. No obstante, también se pueden estudiar además de "sucesos", "tendencias", es decir, variables cuyo valor evoluciona con el tiempo, y ambas cosas a la vez, si el problema así lo requiere. Asimismo, cabe la posibilidad de dividir el periodo total de tiempo en el que se hace la previsión en varios periodos más cortos y localizar la ocurrencia de los sucesos en uno de ellos. Esto da lugar a las variantes "monoperiodo" y "multiperiodo" del método. El efecto o "impacto" de la ocurrencia de un suceso o de la variación de una tendencia sobre otro u otra se suele determinar por consulta a expertos pero también cabe utilizar el coeficiente de regresión si se dispone de datos numéricos o de una encuesta los suficientemente amplia. Tenemos, por tanto, las siguientes variantes del método:

Cross-impact

- Con sucesos
  Con tendencias
  Mixto
- Monoperiodo
  Multiperiodo
- Con impactos estimados por expertos
  Con impactos calculados por regresión
- Y otras clasificaciones todavía

Veremos dos casos: el monoperiodo con sucesos y el multiperiodo con tendencias que son los más sencillos.

Cross-impact monoperiodo con sucesos

Consta de las siguientes fases:

1. Determinar los factores que influyen positivamente sobre aquello que nos interesa. Debe resultar que son sucesos (pueden tener lugar o no y son eventos puntuales) todos ellos a tener lugar o no en un periodo único. Esto puede hacerse con un Brainstorming o similar.

2. Determinar el orden de sucesión. Se supone que este orden es único o que es el más probable.

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

3. Asignar, preguntando a un grupo de expertos y promediando las respuestas, una probabilidad de ocurrencia a cada uno de los sucesos.

4. Determinar la matriz de impactos cruzados. Esto se logra también promediando las respuestas de los expertos a las siguientes preguntas:

- ¿Considera Ud. que la ocurrencia del suceso A influiría sobre la probabilidad del suceso P?
- Caso de influir lo haría ¿positiva o negativamente?
- Y ¿cuánto? Muy poco, poco, regular, mucho o muchísimo. Con ello el impacto I_{AB} quedaría así ponderado entre -5 y +5.

5. Calcular las probabilidades condicionadas simples a partir de las probabilidades simples y de los impactos cruzados. Esto hay varias formas de hacerlo que pueden verse en un artículo de Caselles (1986). Una de ellas es la siguiente :

$$P_{i/j} = P_i + I_{ij} (P_i - d) /5 \qquad \text{cuando } I_{ij} < 0$$

siendo d la mayor de las cotas inferiores de $P_{i/j}$

$$P_{i/j} = P_i + I_{ij} (u - P_i) /5 \qquad \text{cuando } I_{ij} >= 5$$

siendo u la menor de las cotas superiores de $P_{i/j}$

Respecto de las cotas superiores $P_{i/j} <= P_i/ P_j$ o bien 1

según se deduce de $P_i= P_{i/j} P_j + P_{i/-j} P_{-j}$

Y respecto de las cotas inferiores $P_{i/j} \geq 1 + (P_i-1)) / P_j$ o bien 0 según se deduce de $P_{i \vee j} =P_i+P_i-P_{i/j}\cdot P_j$

Las probabilidades condicionadas a la "no ocurrencia" se calculan con la relación:

$P_i= P_{i/j} P_j+ P_{i/-j} P_{-j}$

6. Calcular las probabilidades condicionadas múltiples a partir de las probabilidades condicionadas simples y de las probabilidades simples. También hay varios modos de hacer esto. Nosotros (Caselles, 1986) proponemos el siguiente.

$$P_{i/jk} = P_i \cdot P_{i/j} / P_i \cdot P_{i/k} /P_i$$

7. Calcular las probabilidades de los escenarios a partir de las probabilidades simples y de las probabilidades condicionadas.

$$P_{ijklmn} = P_{i/jklmn} \cdot P_{jklmn} = P_{i/jklmn}\cdot P_{j/klmn}\cdot P_{klmn} = \ldots =$$
$$P_{i/jklmn} \cdot P_{j/klmn} \cdot P_{k/lmn} \cdot P_{l/mn} \cdot P_{m/n} \cdot P_n$$

Obsérvese que el orden de sucesión es: primero sucede "n" o no, después "m" después "l", etc.

8. Ordenar de mayor a menor las probabilidades de los escenarios, y estudiar lo que ocurre y no ocurre en los escenarios más probables. De este modo, comparando las frecuencias de los sucesos en los escenarios más probables con las probabilidades simples estimadas inicialmente podremos saber si el efecto de las interacciones entre los sucesos beneficia o perjudica la ocurrencia de cada uno de ellos.

9. Efectuar un análisis de sensibilidad. Este análisis permite determinar cuáles son los sucesos más influyentes en el problema y por tanto los sucesos que habría que tratar de provocar o de bloquear en la medida de lo posible. Para ello hay que calcular la matriz de sensibilidades de unos sucesos respecto de otros, con los totales al margen. Sensibilidad de! suceso "i" respecto del suceso "j" es:

$$S_{ij} = (\Delta P_i/P_i) \;/\; (\Delta P_j/P_j) = ((P_{i/j}-P_i)/P_i) \;/\; ((1-P_j)/P_j)$$

Cross-impact multiperiodo con tendencias.

Consta de las siguientes fases.

1. Determinar los factores que influyen sobre aquello que nos interesa. Debe resultar que son todos "tendencias" (variables cuyo valor evoluciona con el tiempo). Esto puede hacerse con un Brainstorming o similar.
2. Asignar por procedimientos objetivos (estadísticas) o subjetivos (expertos) un valor inicial a cada tendencia.
3. Estimar un valor máximo absoluto y un valor máximo alcanzable durante la prospección para cada tendencia. También un mínimo absoluto y un mínimo alcanzable.
4. Estimar análogamente (con expertos y promedios) los impactos cruzados entre las tendencias. En el caso que nos ocupa (tendencias) diríamos a los expertos: Considere Ud. que la tendencia A ha llegado a su valor máximo alcanzable previsto. ¿Cómo afecta esto a la tendencia B? Le favorece, le perjudica o le es indiferente. En caso de favorecerle o perjudicarle ¿Cuánto? Muy poco, poco, regular, mucho, o muchísimo.
5. Calcular los valores de las tendencias para el período siguiente acumulando a los del período anterior los que se derivan de los impactos de las demás tendencias.

$$T_i(t+1) = T_i(t) + \sum (I_{ij}/5)\cdot(TM_i-T_i(t))\cdot(T_j(t)-T_j(t-1))/(TM_j-T_j(t-1))$$

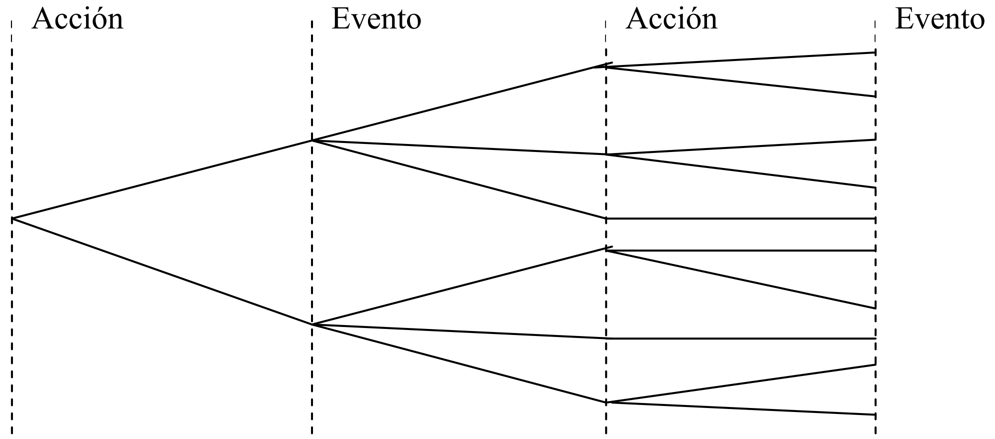
Fórmula válida para impactos positivos, $I_{ij} \geq 0$, e incrementos positivos de la tendencia impactante. TM representa, pues, el respectivo valor máximo alcanzable. Para impactos negativos sustituiríamos $(TM_i-T_i(t))$ por $(T_i(t)-Tm_i)$ siendo Tm_i el valor mínimo alcanzable. Y para disminuciones de la tendencia impactante sustituiríamos $(TM_j-T_j(t-1))$ por $(T_j(t-1)-Tm_j)$.

ÁRBOLES DE DECISION

El concepto de grafos o árboles de relevancia, pertinencia, decisión o confianza (nombres todos ellos utilizados para indicar prácticamente lo mismo) no es nuevo. Parecen ser Churchman y sus colaboradores (30) quienes primero proponen su utilización en contextos industriales generales.

Son antecedentes importantes en su utilización el PPBS ("Planing Programing budgeting System") del Ministerio de Defensa U.S.A. que data de 1961, (3) y el sistema PATTERN ("Planning Assistance Through Technical Evaluation of Relevance Number") de la empresa Honeywell, de 1965 (32). Jantsch (33) y Grof 1 (34) describen esta metodología con ejemplos y bastante detalle. En síntesis consiste en lo siguiente.

1) Tatar de acomodar el planteamiento del problema a una estructura de este estilo:



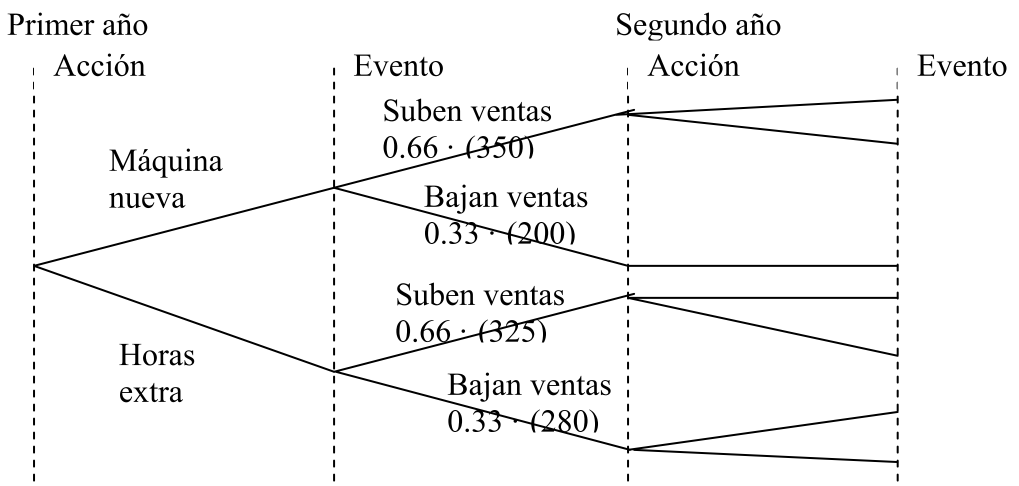
2) Tratar de seleccionar el mejor camino a través de este árbol. Para ello es necesario:

- Dar un nombre a cada acción y a cada evento posible posterior a cada acción.
- Realizar una asignación de ganancias (o pérdidas, o riesgos, etc.) subsecuentes a cada evento y acción predecesora.
- Asignar probabilidades a los eventos.
- Fijar uno o más criterios que permitan el recorrido del camino buscado. Por ejemplo, flujo de fondos esperado máximo, esperanza de ganancia máxima, etc.
- Quizá disponer un cuadro de los utilizados en el análisis multicriterio, para cada decisión, y proceder desde el final hacia el principio en el sentido del tiempo.

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Este proceso, cuando el problema no es muy complejo puede seguirse a mano, en otro caso, será necesario construir un programa de ordenador adecuado al mismo.

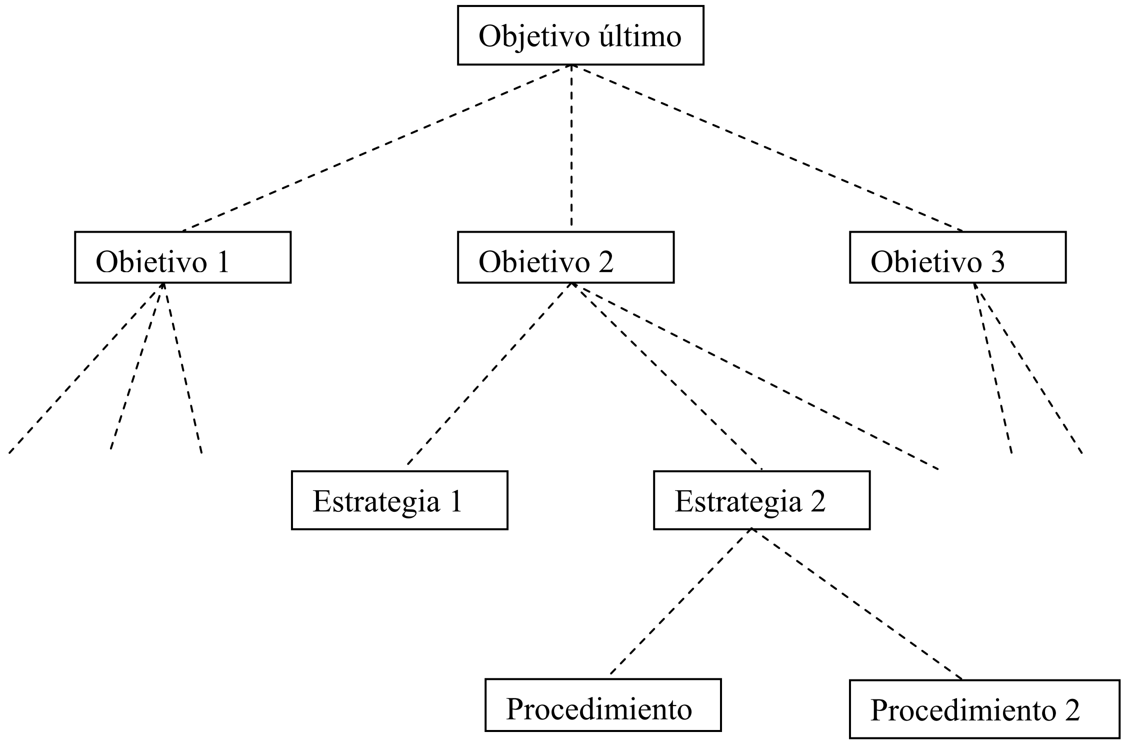
Ejemplo: En una empresa manufacturera, a la vista del buen estado del mercado, se duda entre comprar una máquina nueva o hacer horas extra. El resultado dependerá de si el mercado continúa como hasta ahora o no. La probabilidad de que continúe es de 2/3. El árbol, para el primer año, quedaría así:



Y, la esperanza de ganancia sería:
para máquina nueva: 0.667·350+0.333·200=300
y para horas extra: 0.667·325+0.333·280=310
Por consiguiente, interesa hacer horas extra. Para el segundo año y siguientes continúa el proceso.

El análisis multicriterio (35) es una de las técnicas que se estudian en la Investigación Operativa, y permite la selección entre varias opciones según varios criterios considerados simultáneamente. No es la única posibilidad con tal fin, en los árboles de decisión. Cabe utilizar otras técnicas de optimización como son la Programación Matemática (lineal, Cuadrática, etc.) y la Programación Dinámica. La Programación Dinámica (36) es un método de optimización que resuelve a la vez problemas interdependientes situados en diferentes etapas. Las decisiones tomadas en una etapa se convierten en las condiciones que gobiernan la siguiente. De este modo puede llegarse a la minimización de los costes de un amplio proyecto que se desarrolle en etapas sucesivas.

La descripción de los árboles de decisión que hemos dado corresponde a su versión más amplia en posibilidades y en complejidad matemática. No obstante, puede quedarse en un procedimiento para examinar ordenadamente las diferentes posibilidades y los factores y circunstancias que afectan a cada una de ellas, y en relación con los diferentes objetivos. Tal es el "Método Morfológico" propuesto por Zwicky (37). En esta línea cabe plantear un árbol genérico que podría ser el siguiente:



 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

No obstante, cada caso particular tendrá su árbol particular que será el que mejor se adapte al estudio que se está haciendo.

Respecto a la posible aplicación de este método a la Ordenación del Territorio, podría iniciarse con un árbol como este:

Calidad de la Vida

Sanidad Educación Renta familiar Medio ambiente …..

Nuestra opinión es que el método tiene muchas posibilidades. Y el hecho de que se presente como esencialmente enfocado a la selección entre distintas soluciones para determinados problemas concretos perfectamente localizados en el grafo en relación con otros, lo hace más claro y asimilable para el usuario, aunque su aparato matemático puede llegar a ser tan complicado como el que más.

Ahora bien, cuando la forma de árbol es difícil de conseguir, puesto que las elecciones son múltiples y se observan fenómenos de retroalimentación, la idea se hace difícilmente aplicable. Y esto es lo que suele ocurrir en los sistemas socio-económicos donde las interrelaciones son intrincadas y no se pueden identificar con facilidad las repercusiones de una determinada decisión. No obstante, tal vez en algún enfoque parcial e indicativo pueda tener aplicación.

BIBLIOGRAFIA

Jantsch, E., 1967, "La Previsión Technologique", OCDE. Paris.

Sage, A.P., 1977, "Methodology for Large Scale Systems", Mc Graw Hill.

Godet, M., 1993 "De la anticipación a la acción". Marcombo. Barcelona.

Fontela, E., 1980,"España en la Becada de los 80", Instituto Nacional de Prospectiva. Madrid.

Linstone, H.A., y Turoff M., 1975 “The Delphi Method". Addison- Wessley. Reading (Mass).

DATAR. 1970 "France-2000". Ministère d’amenagement du Territoire. París.

Groff, G.K., 1974, "Modelos de decisión". Ateneo. Buenos Aires.

Diputación Foral de Navarra, 1975, "Navarra-2000".

Caselles, A., 1978 "Un método para la gestión de empresas agrícolas", Tesis doctoral. ETSI. Agrónomos. Valencia.

Lesourne, S., 1979, "Interfutures", OCDE. París.

Caselles, A. 1986, "An empirical comparison of cross-impact models for forecasting sales". Internacional Journal of Forecasting. 2(1986) 295-303.

Apéndice 4

Métodos numéricos útiles en los modelos dinámicos

En muchas ocasiones nos encontramos con que tenemos una tabla de valores de dos o más variables interrelacionadas y nos gustaría encontrar una ecuación que nos permitiera calcular una de ellas en función de todas o parte de las demás. En estas condiciones, lo primero que debemos tener claro es si existe incertidumbre (ruido o inexactitud en la medida) en los datos o si deseamos considerarlos como exactos. En el primer caso el método a utilizar será el ajuste por mínimos cuadrados (o regresión lineal, simple o múltiple) y en el segundo caso será la interpolación (polinómica, generalmente).

Por otra parte, los modelos dinámicos suelen contener ecuaciones diferenciales o ecuaciones en diferencias finitas y el proceso de simulación con estos modelos implica la integración de estas ecuaciones. Es necesario, por tanto, tener claro el procedimiento de integración de este tipo de funciones.

Destacamos, pues tres métodos numéricos: ajuste por mínimos cuadrados, interpolación, e integración de ecuaciones diferenciales ordinarias.

Ajuste por mínimos cuadrados

El método de Mínimos Cuadrados

- Dada una tabla de *n* valores (x,y), se trata de obtener una función y=f(x) con parámetros: $a_0 , a_1 , a_2 ,..., a_{p-1}$ de tal manera que:

$$S = \sum_{i=1}^{n} \left[f(x_i) - y_i \right]^2 \rightarrow \min$$

$$lo \ \ cual \ \ implica \ \ :$$

$$\frac{\partial S}{\partial a_j} = 0 \qquad para \qquad j = 0, 1, 2, ..., \ p-1$$

Antonio Caselles. Universitat de València. España 64

Ajuste de una recta

$$y = f(x) \quad \rightarrow \quad y = a_0 + a_1 x$$

$$S = (a_0 + a_1 x_1 - y_1)^2 + (a_0 + a_1 x_2 - y_2)^2 + ... + (a_0 + a_1 x_n - y_n)^2 \rightarrow \min$$

$$\frac{\partial S}{\partial a_0} = 2[(a_0 + a_1 x_1 - y_1) + (a_0 + a_1 x_2 - y_2) + ... + (a_0 + a_1 x_n - y_n)] \rightarrow = 0$$

$$\frac{\partial S}{\partial a_1} = 2[(a_0 + a_1 x_1 - y_1)x_1 + (a_0 + a_1 x_2 - y_2)x_2 + ... + (a_0 + a_1 x_n - y_n)x_n] \rightarrow = 0$$

$$na_0 + a_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i = 0$$

Ajuste de una parábola

$$y = f(x) \quad \rightarrow \quad y = a_0 + a_1 x + a_2 x^2$$

$$S = \sum_{i=1}^{n} (a_0 + a_1 x_i + a_2 x_i^2 - y_i)^2 \rightarrow \quad \min$$

$$\frac{\partial S}{\partial a_0} = 0 \quad ; \quad \frac{\partial S}{\partial a_1} = 0 \quad ; \quad \frac{\partial S}{\partial a_2} = 0$$

$$na_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Ejercicios

- 1. **Ajustar una recta a los datos**:

x:	1	3	4	6	8	9	11	14
y:	1	2	4	4	5	7	8	9

- 2. **Ajustar una parábola a los datos:**

x:	-5	-4	-3	-2	-1	0	1	2	3	4	5
y:	23	31	39	50	62	76	92	105	122	131	151

- 3. **Ajustar la función** $P{\cdot}V^{\gamma} = C$ **a los datos:**

x:	54.3	61.8	72.4	88.7	118.6	194.0
y:	61.2	49.5	37.4	28.4	19.2	10.1

Antonio Caselles. Universitat de València. España 67

Interpolación

Método de Lagrange

- Definición de polinomio interpolador de grado ≤n

$$f(x)\ \textit{definida}\quad y\ \textit{continua}\quad \textit{en}\ [a,b] \subset R$$
$$\textit{siendo}\quad (x_0, f(x_0)),\ (x_1, f(x_1)),\ \ldots,\ (x_n, f(x_n)),\ n+1\ \textit{puntos}$$
$$\textit{Polinomio}\quad \text{int}\ \textit{erpolador}\ : P_n(x) \quad \Leftrightarrow \quad P_n(x_i) = f(x_i)\ ;\ i = 0,1,\ldots,\ n$$

- Fórmula de lagrange: $P_n(x) = \sum_{k=0}^{n} f(x_k) \cdot L_k(x)$

ya que

$$L_k(x_j) = 1 \quad si \quad j = k$$
$$= 0 \quad si \quad j \neq k$$

$$L_k(x) = \prod_{i=0\,;\,i\neq k}^{n} \frac{x - x_i}{x_k - x_i}$$

Antonio Caselles. Universitat de València. España 47

Método de Newton

- **Fórmula de Newton:**

$$P_n(x) = y_0 + (x - x_0)y_{(x_0,x_1)} + (x - x_0)(x - x_1)y_{(x_0,x_1,x_2)} + \dots$$
$$+ (x - x_0)\cdots(x - x_{n-1})y_{(x_0,x_1,\dots,x_n)}$$

$$y_{(x_0,x_1)} = \frac{y_1 - y_0}{x_1 - x_0} \quad \cdots \quad y_{(x_{k-1},x_k)} = \frac{y_k - y_{k-1}}{x_k - x_{k-1}}$$

$$y_{(x_0,x_1,x_2)} = \frac{y_{(x_1,x_2)} - y_{(x_0,x_1)}}{x_2 - x_0} \quad \cdots \quad y_{(x_{k-1},x_k,x_{k+1})} = \frac{y_{(x_k,x_{k+1})} - y_{(x_{k-1},x_k)}}{x_{k+1} - x_{k-1}}$$

Etc.

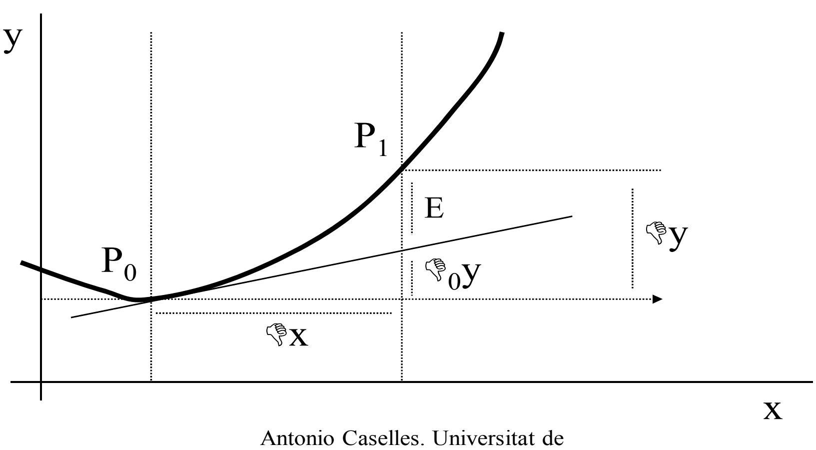
Integración numérica de ecuaciones diferenciales ordinarias

Procedimiento general

Dada y'=f(x,y) con $P_0(x_0, y_0)$

Obtener $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, ... , $P_n(x_n, y_n)$

Es decir, obtener su integral por puntos con error local = E.



Métodos usuales

- **Método de Euler.** $E \leq (\Delta x)^2$

$$x_1 = x_0 + \Delta x \quad ; \quad y_1 = y_0 + \Delta_0 y$$

$$\Delta x \ \textit{arbitrario y cte} \quad ; \quad \Delta_0 y = f(x_0, y_0)\cdot \Delta x$$

- **Método de Runge.** $E \leq (\Delta x)^3$

$$x_1 = x_0 + \Delta x \quad ; \quad y_1 = y_0 + \Delta_1 y$$

$$\Delta x \ \textit{arbitrario y cte} \ ; \ \Delta_1 y = f(x_0 + \Delta x / 2, y_0 + \Delta_0 y / 2)\cdot \Delta x$$

Métodos usuales

- **Método de Runge-Simpson.** $E \leq (\Delta x)^4$

Δx *arbitrario y cte*

$$x_1 = x_0 + \Delta x \quad ; \quad y_1 = y_0 + \frac{1}{6}\Delta_0 y + \frac{2}{3}\Delta_1 y + \frac{1}{6}\Delta_{II} y$$

$$\Delta_0 y = f(x_0, y_0)\cdot \Delta x$$

$$\Delta_1 y = f(x_0 + \Delta x / 2 \ , y_0 + \Delta_0 y / 2)\cdot \Delta x$$

$$\Delta_I y = f(x_0 + \Delta x \ , y_0 + \Delta_0 y)\cdot \Delta x$$

$$\Delta_{II} y = f(x_0 + \Delta x \ , y_0 + \Delta_I y / 2)\cdot \Delta x$$

Métodos usuales

- **Método de Kutta.** $E \le (\Delta x)^5$

$$\Delta x \ _{arbitrario \ \ y \, cte} \quad ; \quad x_1 = x_0 + \Delta x$$

$$y_1 = y_0 + \frac{1}{6}\Delta_0 y + \frac{1}{3}\Delta_1 y + \frac{1}{3}\Delta_2 y + \frac{1}{6}\Delta_3 y$$

$$\Delta_0 y = f(x_0, y_0)\cdot\Delta x$$

$$\Delta_1 y = f(x_0 + \Delta x / 2 \ , y_0 + \Delta_0 y / 2)\cdot\Delta x$$

$$\Delta_2 y = f(x_0 + \Delta x / 2 \ , y_0 + \Delta_1 y / 2)\cdot\Delta x$$

$$\Delta_3 y = f(x_0 + \Delta x \ , y_0 + \Delta_2 y)\cdot\Delta x$$

Ejercicios

1. $y' = x + y \quad con \quad P_0(0,1) \ ; \quad \Delta x = 0.1$
2. $y' = x \cdot y \quad con \quad P_0(1,2) \ ; \quad \Delta x = 0.1$
3. $y'' = x + y \quad con \quad P_0(0,1) \ ; \quad \Delta x = 0.1 \ ; \quad y'_0 = 1$
4. $y'' = x \cdot y \quad con \quad P_0(1,2) \ ; \quad \Delta x = 0.1 \ ; \quad y'_0 = 1$
5. $y' = -0.01 \cdot y \quad con \quad P_0(0,1000) \ ; \quad \Delta x = 0.1$
6. $y' = 0.1 \cdot y^2 - x \ y \quad con \quad P_0(0,1) \ ; \quad \Delta x = 0.1$
7. $y'' + \frac{y'}{x} + y = 0 \quad con \quad P_0(0,1) \ ; \Delta x = 0.2 \ ; y'_0 = 0$
8. $y''' = y \quad con \quad P_0(1,1) \ ; \Delta x = 0.1 \ ; y'_0 = 1 \ ; y''_0 = 2$

Ejercicios

Método de "reducción canónica": $y''=9y$ *se transforma en* $\left.\begin{array}{l} y'=z \\ z'=9y \end{array}\right\}$

$$\left.\begin{array}{l} y'=z \\ z'=9y \end{array}\right\} \quad con \quad P_0=(1,1,1); \quad \Delta x = 0'1; \quad Método \quad de \ Euler$$

$$x_0 = 1; \quad y_0 = 1; \quad z_0 = 1;$$

$$\Delta_0 y = f(x_0, y_0, z_0)\cdot\Delta x = 1\cdot 0'1 = 0'1$$

$$\Delta_0 z = g(x_0, y_0, z_0)\cdot\Delta x = (9\cdot 1)\cdot 0'1 = 0'9$$

$$\left.\begin{array}{l} x_1 = x_0 + \Delta x = 1 + 0'1 = 1'1 \\ y_1 = y_0 + \Delta_0 y = 1 + 0'1 = 1'1 \\ z_1 = z_0 + \Delta_0 z = 1 + 0'9 = 1'9 \end{array}\right\} con \ lo \ que \ P_1 = (1'1,1'1,1'9) \quad etc.$$

Ejercicios

$$y'''=y \quad con \quad P_0(1,1); \ \ y'_0 = 1; \ y''_0 = 2; \ \Delta x = 0'1$$

$$\left.\begin{array}{l} y'=z \\ z'=u \\ u'=y \end{array}\right\} \quad con \quad P_0=(1,1,1,2); \quad \Delta x = 0.1; \quad Método \quad de \ Runge$$

$$x_0 = 1; \quad y_0 = 1; \quad z_0 = 1; \quad u_0 = 2$$

$$\left.\begin{array}{l} \Delta_0 y = z_0\cdot\Delta x = 1\cdot 0'1 = 0'1 \\ \Delta_0 z = u_0\cdot\Delta x = 2\cdot 0'1 = 0'2 \\ \Delta_0 u = y_0\cdot\Delta x = 1\cdot 0'1 = 0'1 \end{array}\right\} \quad \left.\begin{array}{l} \Delta_1 y = (z_0 + \Delta_0 z/2)\cdot\Delta x = (1 + 0'2/2)\cdot 0'1 = 0'11 \\ \Delta_1 z = (u_0 + \Delta_0 u/2)\cdot\Delta x = (2 + 0'1/2)\cdot 0'1 = 0'205 \\ \Delta_1 u = (y_0 + \Delta_0 y/2)\cdot\Delta x = (1 + 0'1/2)\cdot 0'1 = 0'105 \end{array}\right\}$$

$$\left.\begin{array}{l} x_1 = x_0 + \Delta x = 1 + 0'1 = 1'1 \\ y_1 = y_0 + \Delta_1 y = 1 + 0'11 = 1'11 \\ z_1 = z_0 + \Delta_1 z = 1 + 0'205 = 1'205 \\ u_1 = u_0 + \Delta_1 u = 2 + 0'105 = 2'105 \end{array}\right\} con \ lo \ que \ P_1 = (1'1,1'11,1'205,2'105) \quad etc.$$

Apéndice 5

Conceptos y métodos estadísticos útiles en modelos dinámicos

1. MUESTREO DE ATRIBUTOS. DISTRIBUCIÓN BINOMIAL

1.1. CONCEPTOS BÁSICOS.

- Población.- Es el colectivo objeto de estudio. Por ejemplo: el contenido en trigo de un silo, o el número de piezas fabricadas en un día por una máquina determinada.

- Muestra.- Es una porción tomada de la población. Debe ser "representativa", es decir, debe representar fielmente a la población. Esto se consigue tomándola "al azar", es decir, por procedimientos aleatorios. Más adelante veremos esto con más detalle.

- Estimación o inferencia.- Consiste en deducir características de la población a partir de otras análogas en la muestra. Por ejemplo, el promedio de piezas defectuosas.
 . Estimación por punto.- Consiste en dar una cifra para el parámetro o característica que se está estimando. Por ejemplo: 2% de defectuosos.
 . Estimación por intervalo.- Consiste en dar unos valores máximo y mínimo entre los que se puede encontrar el parámetro que se estima con una frecuencia o probabilidad del 95% o del 99%, u otra. Por ejemplo: el número de defectuosos se encontrará entre el 1,2% y el 2,7% en el 95% de las muestras que tomemos. Esto significa que solamente en 5 de cada 100 muestras que tomemos saldrá un número de defectuosos menor del 1,2% o mayor del 2,7%.

- Atributo.- Característica no susceptible de ser medida. Por ejemplo: la cara o cruz de una moneda.

- Muestreo de atributos.- Registro del número de individuos que poseen el atributo que se estudia. Por ejemplo: Número de caras que salen al lanzar diez veces una moneda.

- Nueva definición de "Población".- Aquel conjunto del cual cada uno de sus miembros tiene una probabilidad conocida de salir en la muestra. Para que un conjunto, por ejemplo, un montón de trigo, pueda ser considerado una población se debe de cumplir una de las siguientes condiciones:
 . Debe homogeneizarse mezclándolo adecuadamente. De no ser así, no existe garantía de que la parte de abajo sea como la de arriba o la exterior como la interior.
 . Debe encontrarse un procedimiento para tomar la muestra, que rebase el control de quien muestrea. Por ejemplo, un procedimiento mecánico consistente en hacer pasar todo el trigo del montón por un tubo que tuviese un agujero, de tal manera que al final de la operación, por el agujero en cuestión, haya salido la porción considerada como muestra.

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

- Muestreo estratificado al azar.- Consiste en tomar una muestra de cada una de las clases de individuos que componen la población de modo que el tamaño de cada muestra sea proporcional al de la clase de la que procede.
Por ejemplo: se desea estimar la edad media de la población laboral de una ciudad. Para ello se consulta al 1% de la misma, pero se tiene la precaución de que la parte que corresponde a digamos "fontaneros", sea el 1% de los fontaneros y la que corresponde a "pintores" sea el 1% de los pintores, etc..

- Tamaño de la muestra.- La precisión con que se estima un parámetro depende del tamaño de la muestra que se toma. Por ejemplo: con una muestra pequeña de piezas fabricadas saldría un número de defectuosos comprendido entre 0% y 20% con una probabilidad del 95%. En cambio con una muestra mayor saldría un número de defectuosos comprendido entre 10,3% y 12,7% con el mismo 95% de probabilidad. Existen tablas que relacionan el tamaño de la muestra con la amplitud del intervalo.

- Nivel de significación.- Así se llama al aludido 95% del ejemplo anterior. Y representa la proporción de muestras en la que se cumplirá presumiblemente la estimación que se hace.

- Tabla de números aleatorios.- Es un instrumento que permite sustituir a un bombo con bolas numeradas, dado que ha sido construida con ayuda del mismo (o equivalente). Consiste en diez mil cifras comprendidas entre el cero, 0, y el nueve, 9, dispuestas una a continuación de otra en filas y columnas constituyendo una tabla.

- Muestreo aleatorio a "al azar".- Consiste en tomar la muestra de modo que se esté seguro de que todos y cada uno de los individuos de la población tienen la misma probabilidad de salir en la muestra.

Un modo frecuente de tomar muestras aleatorias es numerando a todos los individuos de la población y sacando números de una tabla de números aleatorios y, modernamente, apretando la tecla RAN de la calculadora de bolsillo.

Tanto en un caso como en el otro se toman números de magnitud equivalente a los que han sido asignados a los individuos de la población. Por ejemplo, si la población tiene. 10.000 individuos numerados del 0000 al 9999, está claro que hacen falta cuatro cifras para representar a un individuo. Y así, cuando al apretar la función RAN aparece en la pantalla 0.027651004, tomaremos las cuatro primeras cifras (o cualesquiera otras, pero siempre del mismo modo, para no introducir elecciones personales). Téngase presente que la función RAN da números aleatorios comprendidos entre cero y uno. Por eso aparece el punto decimal.

1.2. PRUEBAS DE HIPÓTESIS CON LA DISTRIBUCIÓN CHI-CUADRADO.

1.2.1. El problema a resolver.

Se tiene una población con varias clases de individuos y una teoría sobre en qué proporción se encuentra cada clase. Se trata de determinar qué probabilidad tiene esa teoría de ser cierta en base al estudio de una muestra. 0 dicho de otro modo y más rigurosamente, determinar que probabilidad tendría esa muestra en caso de ser cierta la teoría.

Si tal probabilidad es alta, será lógico pensar que la teoría es buena y digna de aceptarse.

Por ejemplo: se lanza una moneda al aire cien veces y salen 39 caras y 61 cruces. La teoría que se tiene dice que deberían haber salido 50 y 50 en el caso ideal. Cada vez que se lance 100 veces una moneda saldrá una proporción distinta de caras y cruces. Pero ¿cuándo cabe decir que una determinada proporción es anormal? Si se dispone de un criterio para ello se estará en condiciones de afirmar con ciertas garantías que la moneda está trucada o que no, a la vista de una muestra determinada.

1.2.2. χ^2, un índice de dispersión.

Intuitivamente, χ^2 es una medida de la "distancia" que separa a la nuestra obtenida de la muestra ideal que se ajustase exactamente a las proporciones teóricas.

$$\chi^2 = \frac{(\text{Frecuencia real} - \text{Frecuencia teórica})^2}{\text{Frecuencia teórica}}$$

Fórmula que referida al ejemplo de la moneda se aplica como sigue:

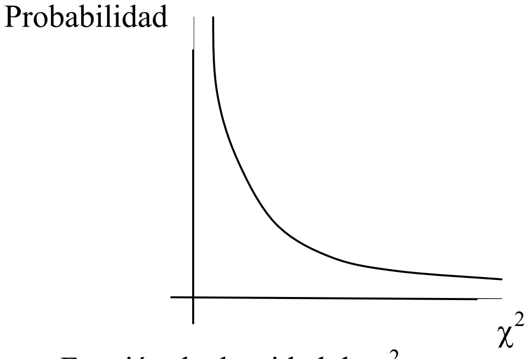
$$\chi^2 = \frac{(39-50)^2}{50} + \frac{(61-50)^2}{50} = 4.84$$

Obsérvese que χ^2 crecerá a medida que sean mayores las diferencias "real - teórico" de las frecuencias de las diferentes clases de individuos. Estas diferencias están al cuadrado para que salgan siempre positivas, y se dividen por la frecuencia teórica para darles un valor relativo, independiente del tamaño de la muestra.

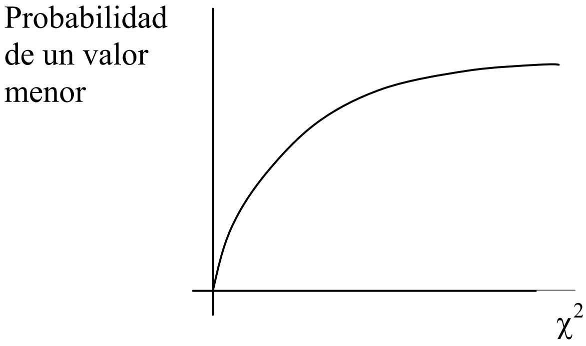
Cada muestra que se obtenga tendrá su χ^2, de tal modo que si de una población que tenga proporciones reales 50% - 50% de dos clases de individuos se sacan muchas muestras y se calcula el χ^2 de cada una de ellas y se representa en una tabla las veces que se repite cada χ^2 **se** habrá construido un instrumento de gran utilidad, dado que en la misma estarán en mayor proporción los χ^2 mas "normales" y en menor proporción los mas "raros". Y a la vista de esa tabla podremos catalogar al 4,84 como "normal" o como "raro", concluyendo en este último caso que la proporción teórica probablemente no sea 50% - 50%.

1.2.3. Manejo de las tablas χ^2.

Las tablas de la distribución χ^2 que se utilizan en la práctica son las acumulativas. En 1.2.2. se ha descrito someramente la tabla de "densidad de probabilidad" que representa la frecuencia o probabilidad con que aparece cada valor de χ^2. En las tablas acumulativas se representa la "función de distribución", es decir, la frecuencia o probabilidad de un valor menor que el que se considera. Y en algunas, las más prácticas, se representa la frecuencia o probabilidad de un valor mayor que el que se considera.



Función de densidad de χ^2



Función de distribución de χ^2

De este modo, la tabla para dos clases de individuos dice así: Probabilidad de un valor mayor

.900	.750	.500	.250	.100	.050	.025	.010	.005
.02	.10	.45	1.31	2.71	3.84	5.02	6.83	7.88

Buscamos el 4.84 que nos ha salido y vemos (interpolando) que corresponde a una "probabilidad de un valor mayor" de 0,028. Esto significa que si sacamos un número muy alto de muestras (de 100 tiradas) solamente el 2,8% de ellas tendrá un X2 mayor que 4,84. Así de "raro" es, por consiguiente, ese 4,84 y la muestra de donde procede. Vamos a precisar ahora dos conceptos que hemos utilizado pero no definido todavía.

1°) Nivel de significación.- Con la moneda del ejemplo anterior hemos probado una hipótesis: la proporción teórica del 50% - 50% de caras y cruces, y nos ha salido que NO, al nivel de significación del 97,2%.

97,2 es 100 - 2,8

Nivel de significación es, por consiguiente el porcentaje de veces que, caso de ser cierta la hipótesis que se prueba, saldría un X2 menor que el obtenido en la muestra.

2°) Grados de libertad.- Es el número de clases de individuos que presenta la población menos uno. En el caso de la moneda tenemos un grado de libertad: dos clases de individuos (cara y cruz) menos uno. Significa que basta con que conozcamos una de las proporciones (la de las caras o la de las cruces), la otra se deduce por diferencia, no es necesario conocerla en principio, viene determinada en función de la otra.

2. MUESTREO DE POBLACIONES CON DISTRIBUCIÓN NORMAL.

Los datos que consideramos ahora ya no son atributos sino características susceptibles de ser medidas. Por ejemplo: estaturas de personas, diámetros de tornillos, etc.

La población objeto de estudio será pues una población de personas, cada una con su estatura, o de tornillos, cada una con su diámetro etc. Esta población tendrá una distribución de frecuencias (veces que se repite cada estatura o diámetro). Y esta distribución podrá ser o no la distribución normal o campana de Gauss. El procedimiento para comprobar si es o no una distribución normal es el explicado en el capítulo anterior: el recurso a X2 considerando el número **de clases** de individuos igual al número de intervalos fijados en la distribución que se está estudiando. Cuando está claro que estamos ante una distribución normal, lo primero es estimar su media **μ** y su desviación típica **σ**.

2.1. ESTIMADORES "POR PUNTO" DE μ Y DE σ

Tomamos una muestra al azar de la población objeto de estudio, y calculamos la media muestral X, la amplitud: (Xmáximo - Xmínimo) y la desviación típica muestral, S.
La media poblacional /u- se estima por la x, pero la cr se estima, bien por la s o bien por una fracción de la amplitud dada por unas tablas. La s es un estimador mejor que la fracción de la amplitud y si se dispone de una calculadora que la calcule automáticamente, es sin duda el estimador más adecuado.

$$\bar{x} = \frac{\sum_{1}^{n} X_i}{n} \qquad s_{n-1} = \sqrt{\frac{\sum_{1}^{n} (X_i - \bar{x})^2}{n-1}}$$

Xi son los datos.
n es el tamaño de la muestra.
Obsérvese que en el cálculo de s se divide por n - 1. Ello se debe a que se puede demostrar que haciéndolo así se estima mejor σ que dividiendo por n.

2.2. ESTIMACIÓN POR INTERVALO DE μ Y DE σ

2.2.1. La distribución "t de Student".

Esta distribución cumple una misión análoga a la X2 vista en el capítulo anterior. El parámetro "t" referido a una muestra concreta trata de medir la "distancia" entre la muestra que nos ha salido y la muestra media, para así evaluar la "rareza" o "normalidad" de la muestra obtenida.

Se define así :

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad s_{\bar{x}} = s / \sqrt{n}$$

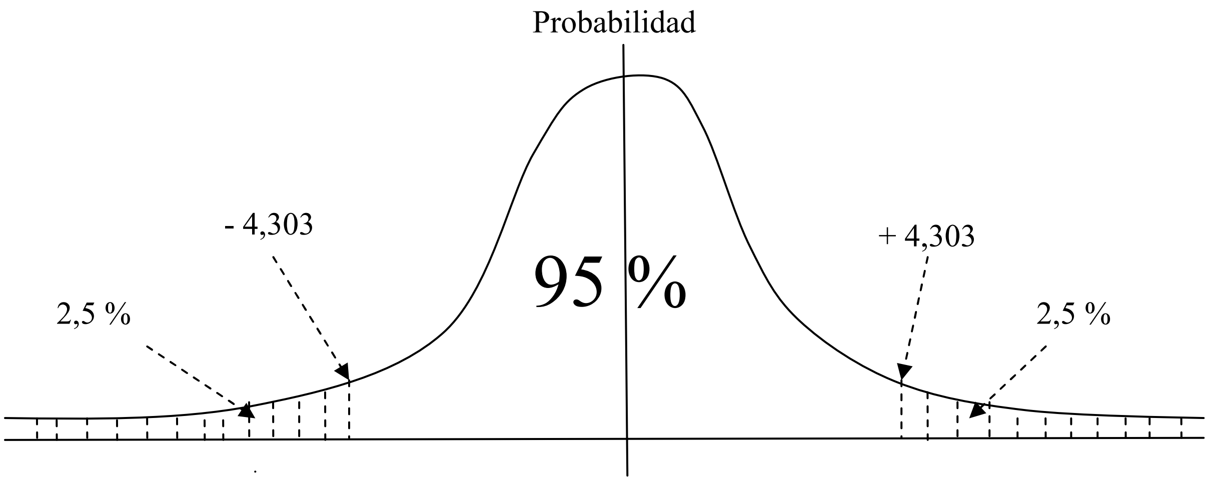
Vemos que "t" es la diferencia entre la media muestral y la media poblacional medida en unidades de $s_{\bar{x}}$. Se puede demostrar que $s / \sqrt{n}$ es la desviación típica de las medias muestrales.

Aclaremos algo esto por si acaso. Imaginemos que sacamos una gran cantidad de muestras. Cada una tiene su media x y su desviación típica s. Pero ¿cuál es la desviación típica de todas estas x? Pues precisamente $s / \sqrt{n}$. Análogamente cada muestra tiene su "t", pero ¿cómo se distribuyen las frecuencias de los valores de t? Pues de una manera muy concreta y reflejada en unas tablas.

De este modo comparando la "t" de nuestra muestra con la de las tablas podremos saber si nuestra muestra es o no "rara", dada las condiciones supuestas.

¿A qué condiciones nos referimos? Veámoslo. Primero la condición de normalidad en la distribución objeto de estudio. Después al valor de μ. Las tablas se hacen para una distribución normal con μ y σ conocidas. En cambio nosotros no conocemos ni μ ni σ. Tenemos que hacer hipótesis sobre ellas. Y estas hipótesis pueden quedar invalidadas por la "rareza" de t.

La curva de la función de densidad de la distribución de "t" es una campana parecida a la normal.



Las tablas de la t suelen ser como sigue:

Grados De Libertad	Probabilidad de un valor mayor ignorando signos .500	.400	.200	.100	.050	.025	.010	.005
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657	
2	.816	1.061	1.886	2.290	4.303	6.205	9.925	14.089

Grados de libertad son aquí el tamaño de la muestra menos uno.

$$g.l. = n - 1$$

¿Cómo se interpreta esto? Veámoslo. Las tablas dicen que, en las condiciones de la hipótesis, y con una muestra de tres mediciones, el 95% de las veces que se saque la muestra saldrá una "t" comprendida entre -4,303 y + 4.303. Esa es la interpretación de: grados de libertad = 2, y probabilidad de un valor mayor = .050, ignorando signos.

2.2.2 Estimación por intervalo de μ

¿Es posible utilizar estos conceptos para estimar por intervalo a μ? Efectivamente. Obsérvese que si el 95% de las muestras deben dar una t comprendida entre -t.05 y + t.05, podremos escribir:

$$
\begin{aligned}
& -t_{0.05} \leq \frac{\bar{x} - \mu}{s_{\bar{x}}} \leq +t_{0.05} \\
& -t_{0.05}\ s_{\bar{x}} \leq \bar{x} - \mu \leq +t_{0.05}\ s_{\bar{x}} \\
& -\bar{x} - t_{0.05}\ s_{\bar{x}} \leq -\mu \leq -\bar{x} + t_{0.05}\ s_{\bar{x}} \\
& \bar{x} + t_{0.05}\ s_{\bar{x}} \geq \mu \geq \bar{x} - t_{0.05}\ s_{\bar{x}} \\
& \bar{x} - t_{0.05}\ s_{\bar{x}} \leq \mu \leq \bar{x} + t_{0.05}\ s_{\bar{x}}
\end{aligned}
$$

que es una estimación por intervalo de μ al nivel de significación del 95%.

Ejemplo: Medimos tres alumnos elegidos al azar en una clase y obtenemos: 1,68, 1,71, y 1,74 ¿Entre qué valores podemos afirmar que se encontrará la estatura media de la clase con una probabilidad del 95% ?

Calculamos $\bar{x} = 1.71$ $s = 0.03$

$1.71 - 4.303 * 0.03/\sqrt{3} <= \mu <= 1.71 + 4.303 * 0.03 / \sqrt{3}$

$$1.635 <= \mu <= 1.785$$

2.2.3 Pruebas de hipótesis sobre μ·

Sigamos con el ejemplo anterior. Podemos plantearnos ahora la siguiente pregunta: si la media de la clase fuese 1,80 ¿Qué probabilidad tendría la muestra obtenida?

Calculamos $$t = \frac{1.71 - 1.80}{0.03/\sqrt{3}} = -5.196$$

Vamos a las tablas con g,1.=2 y 5.196 corresponde a una probabilidad de un valor mayor de 0.038. La muestra obtenida tiene una probabilidad de 0,019 por ser t negativa (conocemos el signo). Podemos decir que es una muestra "rara" pues solamente un 3,8% de las muestras darían un t mayor +5,196 o menor que -5,196. Debemos rechazar la hipótesis de **μ** = 1,80 al nivel de significación del 95% pues 3,8% es menor que 5%. En cambio al nivel de significación del 99% debemos aceptarla pues 3,8% > 1%.

Esta idea sirve de base para el diseño de experimentos que tienen por finalidad probar la existencia de diferencias entre tratamientos dados a una determinada población en estudio.

2.2.3.1. Formación de pares.

Consideremos el siguiente problema: pintar un coche con la pintura A cuesta 800 €. y con la pintura B cuesta 600 € Parece ser que la pintura A tiene mayor duración pero deseamos asegurarnos de ello. Por otra parte, y en el supuesto de que la tenga, deseamos saber si esta diferencia de duración compensa la diferencia de precio.

Para contestar a estas cuestiones diseñamos el siguiente experimento: pintar 10 coches seleccionados al azar, cada uno de ellos mitad de una pintura y mitad de la otra. Dejarlos rodar y contar el número de meses transcurridos hasta que un determinado criterio indique que la pintura ya está estropeada. El hecho de pintar cada coche con las dos pinturas (mitad izquierda con la A y mitad derecha con la B) tiene por objeto que no se enmascaren los resultados por factores extraños como pueden ser: el que un coche sea de una ciudad y otro de otra con diferente clima, el haber sido utilizado más o menos etc.

Los resultados figuran en la tabla siguiente:

	Duración en meses		
Coche nº	Pintura A	Pintura B	Diferencia
1	55	48	7
2	59	61	-2
3	60	49	11
4	63	51	12
5	75	63	12
6	65	61	4
7	57	53	4
8	59	57	2
9	58	55	3
10	63	51	12

$\bar{x}_A = 61.4$

$\bar{x}_B = 54.9$

$\bar{x}_{dif} = 6.5$ s = 5.039

Vemos que la diferencia media de las duraciones es de 9.0 meses y su desviación típica de 5.637, ambos valores estimados en base a la muestra de 10 coches.

Ahora planteamos la siguiente cuestión: en el caso de que no hubiese diferencia entre ambas pinturas, ¿Que probabilidad tendría la muestra obtenida, con su $\bar{x}_{dif}$ = 9.0 meses a favor de la pintura A?

Y para contestar calculamos "t"

$$t = \frac{\bar{x}_{dif} - \mu}{s/\sqrt{n}} = \frac{6.5 - 0}{5.039/\sqrt{10}} = 4.079$$

y vamos a las tablas, entrando por la fila de 9 grados de libertad (10 menos 1). Vemos lo siguiente:

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Grados de libertad	Probabilidad de un valor mayor ignorando signos							
	.400	.200	.100	.050	.025	.010	.005	.001
9	.883	1.383	1.833	2.262	2.634	3.169	3.690	4.781

El valor 4.079 corresponde a una probabilidad de la muestra comprendida entre 0,005 y 0,001, exactamente p=0,00643, es decir un 6,43 por mil. En otras palabras : si no existiese diferencia entre las dos pinturas, una muestra como esta (con su t = 4.079) saldría unas seis veces de cada mil que hiciésemos el experimento. Como conclusión rechazamos la hipótesis de que no hay diferencia entre la pintura A y la B sabiendo que nos equivocaremos seis veces de cada mil.

La segunda cuestión planteada es si la diferencia de duración compensa la diferencia de coste. Esta diferencia (800 € - 600 € = 200 €) representa 1/3 del coste total de B. Como la duración media de B es de $\bar{x}_B$ = 54,9 meses, para ser más rentable la pintura A, debería durar 54,9 meses + 1/3 * 54,9 que son 54,9 + 18,3 = 73,2 meses.

La diferencia mínima para que se gane dinero pintando con A es de 18,3 meses, y la muestra nos ha dado una diferencia media de 6,5 meses. Ahora nos preguntamos : si existiese esa diferencia de 18,3 entre A y B ¿Que probabilidad tendría la muestra que hemos obtenido?. Calculamos:

$$t = \frac{\bar{x}_{dif} - \mu}{s/\sqrt{n}} = \frac{6.5 - 18.3}{5.039/\sqrt{10}} = -7.405$$

Vamos a las tablas (9 grados de libertad) y vemos que este valor de t se sale de la tabla, es decir, tiene una probabilidad inferior al uno por mil. Como conclusión, rechazamos la hipótesis de que n= 18.3, en otras palabras, que si realmente existiese esa diferencia, la muestra obtenida sería improbabilísima. Dicho de otro modo todavía: que es prácticamente seguro que no se gana dinero pintando con A, a pesar de que dura más que B.

2.2.3.2. Formación de grupos.

Consideremos el siguiente problema: somos una gran empresa de transportes y necesitamos saber que motor tiene mayor duración, el motor A o el motor B. En este caso no podemos formar pares como en el caso de la pintura, pues no se puede equipar a un vehículo con los dos motores a la vez. La solución es, de un lote homogéneo (vehículos idénticos) de, por ejemplo, 20 camiones, dotar a 10 de ellos seleccionados al azar del motor A y a los otros 10 del motor B. Dejar rodar a los vehículos hasta que un criterio técnico determine que sus motores están fuera de servicio, y tomar nota de la duración. Los resultados son los de la tabla que sigue:

Motor	Duración en miles de kilómetros										Media	Desv.Tip.
A	351	347	358	321	322	361	356	318	316	337	338.7	18.037
B	305	291	307	301	286	339	351	299	301	331	313.9	22.093

El análisis de estos resultados también podemos efectuarlo con ayuda de la "t" de Student. La "t" es una media muestral menos una media poblacional dividido por la desviación típica de esa media muestral (recuérdese 2.2.1). Pues bien, aquí vamos a considerar como media muestral $\bar{x}_A$ - $\bar{x}_B$, es decir la diferencia de medias. La media poblacional será μ_A - μ_B. Y el denominador será la desviación típica de la diferencia de medias.

Para ayudar a entender esto imaginemos dos urnas A y B cada una con su μ. Sacamos, con una mano en cada urna, muestras de 10 bolas en cada mano. Calculamos $\bar{x}_A$ - $s^2{}_A$ de las diez de la derecha, y $\bar{x}_B$ - $s^2{}_B$ de las diez de la izquierda. Restamos $\bar{x}_A$ - $\bar{x}_B$ y anotamos ese_ valor. Repetimos la operación 1000 veces. Los valores $\bar{x}_A$ - $\bar{x}_B$ tienen a su vez una media y una desviación típica. Esa desviación típica es lo que hemos llamado desviación típica de la diferencia de medias, $s_{\bar{x}_A-\bar{x}_B}$, y se puede demostrar que se estima por el valor:

$$s_{\bar{x}_A-\bar{x}_B} = \frac{\sqrt{s_A^2 - s_B^2}}{\sqrt{n}} \qquad\qquad s_{\bar{x}_A-\bar{x}_B} = \sqrt{\frac{s_A^2}{n_1} + \frac{s_B^2}{n_2}}$$

con lo que la fórmula para t queda:

$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{s_A^2 + s_B^2} / \sqrt{n}}$$

Y aplicándola a este caso concreto:

$$t = \frac{(338.7 - 313.9) - (0)}{\sqrt{(18.037)^2 + (22.093)^2} / \sqrt{10}} = 2.750$$

Considerar μ_A - μ_B = 0 equivale a probar la hipótesis de que ambos motores duran lo mismo. Es decir, nos preguntamos: si ambos motores, el A y el B, durasen en realidad lo mismo, ¿qué probabilidad tendría la muestra que hemos obtenido? ¿sería una muestra rara? ¿o no?.

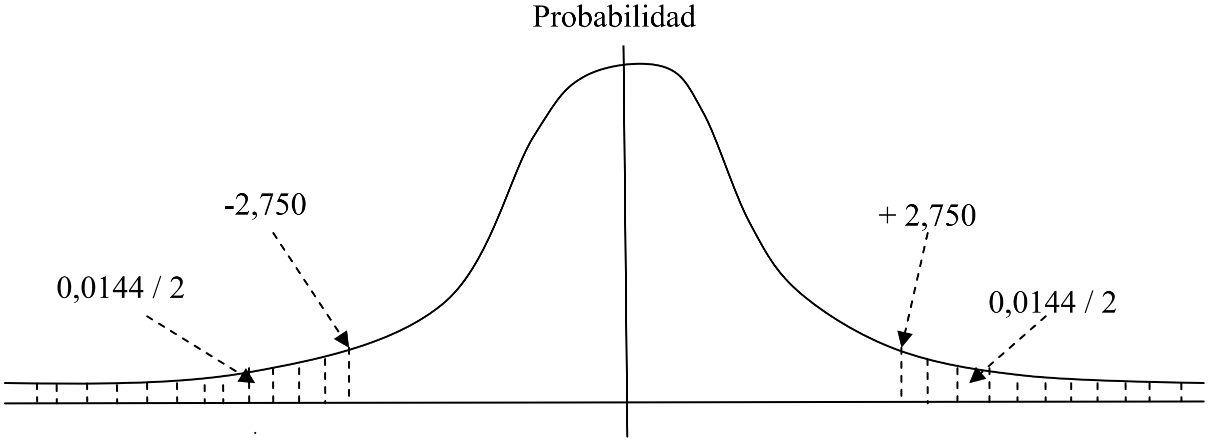
Para contestar nos vamos a la tabla de "t" con 18 grados de libertad (9 + 9), y vemos lo siguiente :

Grados de libertad	Probabilidad de un valor mayor ignorando signos							
	.400	.200	.100	.050	.025	.010	.005	.001
18	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922

El valor 2.750 corresponde a una probabilidad comprendida entre 0.025 y 0.010, que es exactamente p=0.0144, es decir, un 1,44 por ciento.

Esto nos obliga a rechazar la hipótesis $\mu_A - \mu_B = 0$ con lo que diremos que ambos motores tienen una duración distinta y nos equivocaremos 1,44 veces de cada cien. Es decir lo diremos con un grado de seguridad del 98,56%.

Si por razones técnicas tuviésemos la completa seguridad de que el motor A jamás puede durar menos que el B, la hipótesis alternativa a $\mu_A - \mu_B = 0$ ya no podría ser $\mu_A - \mu_B \neq 0$, sino $\mu_A - \mu_B > 0$ con lo que el 0.0144, qué representa la suma de las dos áreas rayadas de la curva "t",(recuérdese que dice "probabilidad de un valor mayor, ignorando signos", es decir, probabilidad de un valor mayor que +2.750 o menor que -2.750) ya no representaría la probabilidad de la muestra obtenida, sino su doble.



En estas condiciones, la muestra obtenida todavía resulta, pues, más improbable, y aumenta la seguridad de la afirmación " el motor A dura más que el B", llevándola al 99,325%. Grado de seguridad aquí, es lo mismo que nivel de significación. Aunque cuando se habla de nivel de significación se suele referir uno al 95% o al 99% sin decimales. Así si el "grado de seguridad" sale mayor que 95% se dice que la prueba es significativa al 95% y si es mayor que 99% análogamente.

2.2.4. Pruebas de hipótesis sobre σ.

2.2.4.1 Muestras de poblaciones normales.

Se puede demostrar que el valor $\dfrac{\sum_1^n (x_i - \bar{x})^2}{\sigma^2}$ obtenido a partir de una población con distribución normal, se distribuye en el muestreo según una X2 con n-1 grados de libertad. Esto permite hacer pruebas de hipótesis sobre σ y también estimarla por intervalo. Veámoslo.

$$\chi^2_{0.975} \leq \frac{\sum_1^n (x_i - \bar{x})^2}{\sigma^2} \leq \chi^2_{0.025}$$

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

$$\frac{\sum_{1}^{n}(x_i-\bar{x})^2}{\chi^2_{0.025}} \leq \sigma^2 \leq \frac{\sum_{1}^{n}(x_i-\bar{x})^2}{\chi^2_{0.975}}$$

Se sospecha que una dieta B para cerdos que se pretende introducir en sustitución de otra A, va a dar lugar a una mayor dispersión en los pesos finales de los cerdos. Como quiera que esto sería un serio inconveniente para B, se pretende comprobarlo mediante una prueba estadística.

El valor de σ para los pesos finales con la dieta A es de 25. Se desea saber si se ha modificado o no con el cambio de dieta. Para ello se toman 20 cerdos seleccionados al azar de entre los que han sido criados con la dieta B y ya han alcanzado el peso final y se pesan. Se calcula el peso medio x y la suma de los cuadrados de las diferencias entre cada uno de los pesos finales y su promedio. El resultado resulta ser 21679. Calculamos:

$$\chi^2 = \frac{21679}{(25)^2} = 34.686$$

Vamos a la tabla de X2 con 19 grados de libertad y vemos:

Grados de libertad	Probabilidad de un valor mayor						
	.995	.990	...	.025	.010	.005	.001
19	6.84	7.63	...	32.85	36.19		

Con lo que el valor 34,686 corresponde a una probabilidad comprendida entre 0.025 y 0.010, es decir, aproximadamente un dos por ciento. Ello nos obliga a considerar que σ no puede ser 25, con lo que diremos que la desviación típica se ha modificado y nos equivocaremos dos veces de cada cien.

2.2.4.2. Caso de formación de grupos.

En el caso de que exista la sospecha de que el tratamiento dado a cada grupo puede afectar a la varianza es necesario comprobarlo, pues el método expuesto en 2.2.3.2 solo es válido cuando la varianza es la misma.

Para ello se recurre a una nueva distribución: la "F" de Snedecor. Se define así:

$$F = \frac{s^2 \text{ muestral mayor}}{s^2 \text{ muestral menor}}$$

En el caso de los motores de 2.2.3.2 sería:

$$F = \frac{(22.093)^2}{(18.037)^2} = 1.500$$

Vamos a la tabla de F y vemos que para el 5% (ambos extremos):

Grados de Libertad para La s^2 menor	Grados de libertad para la s^2 mayor			
	2	4	...	9
9	...		...	4.03

Esto significa que, si la varianza de la población correspondiente al numerador fuese la misma que la de la población correspondiente al denominador, solo un 5% de las veces que se obtuviesen dos muestras, una de cada población, y se dividiesen sus varianzas (la mayor por la menor) saldría un cociente mayor que 4,03. Como en el caso de los motores F= 1,500 es menor que 4,03 concluimos que es un valor bastante probable y por consiguiente las distribuciones de las duraciones de ambos motores tienen la misma varianza.

3. REGRESION LINEAL

3.1. Objetivos

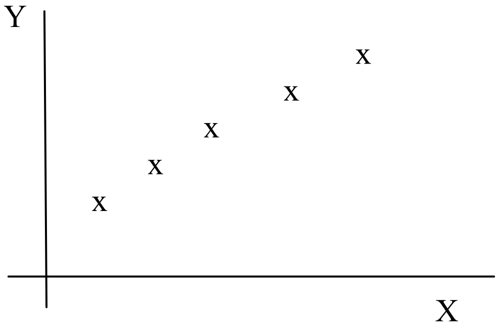
Se pretende estudiar la relación que puede existir entre dos variables aleatorias X e Y. Concretamente:

- Si existe dependencia entre X e Y.
- Como depende X de Y.
- Con qué fiabilidad se puede predecir Y a partir de X

3.2. Dependencia entre dos variables

En muchos problemas de la vida real surge esta cuestión. Por ejemplo: ¿El peso corporal de los adultos depende de la edad? ¿La velocidad de una reacción química depende de la temperatura?, etc. Normalmente se dispone de un conjunto de observaciones simultáneas que corresponden a una muestra aleatoria de la población que se pretende investigar. Por ejemplo:

ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008



X Edad	Y Peso corporal medio (Kg)
30	65
40	72
50	75
60	78
70	83

La idea que surge en este momento es: parece que cuando la edad aumenta el peso también lo hace y se podría ajustar una recta a esos puntos por el método de los mínimos cuadrados.

Se puede demostrar que la recta obtenida por ese procedimiento tiene por ecuación:

$$\hat{Y} - \bar{y} = b\,(X - \bar{x}) \qquad \text{siendo} \qquad b = \frac{\sum(Xi - x)(Yi - y)}{\sum(X1 - X)^2}$$

Es decir, es la recta que pasa por el punto P(x,y) y tiene por pendiente b. $\hat{Y}$ es el valor estimado con la recta de regresión para la variable dependiente Y cuando la variable independiente vale X. Las variables x e y son los valores medios de X e Y respectivamente. La variable b es lo que se llama "coeficiente de regresión" y se suele escribir: $b=(\sum xy)/(\sum x^2)$, donde x e y (minúsculas) son las diferencias respecto de los respectivos promedios, es decir: $x_i = X_i - \bar{x}$; $y_i = Y_i - \bar{y}$.

Para el cálculo manual del coeficiente de regresión y su estimación por intervalo habría que rellenar la siguiente tabla:

	X_i	Y_i	$X_i - \bar{x}$ x	$Y_i - \bar{y}$ y	x^2	xy	$\hat{Y}$	$d_{y.x}$	$d^2_{y.x}$
	30	65	-20	-9,6	400	192	66,2	-1,2	1,44
	40	72	-10	-2,6	100	26	70,4	1,6	2,56
	50	75	0	0,4	0	0	74,6	0,4	0,16
	60	78	10	3,4	100	34	78,8	-0,8	0,64
	70	83	20	8,4	400	168	83	0	0
Suma	250	373	0	0	1000	420			4,8
Media	50	74,6							

Donde:

$$b = \frac{420}{1000} = 0.42 \quad ; \quad \hat{Y} - 74.6 = 0.42\,(X - 50) \quad ; \quad \hat{Y} = 0.42\,X + 53.67$$

Ahora surge la pregunta: ¿Podría esta aparente relación entre X e Y deberse al azar?, es decir, si tomáramos otra muestra ¿Podrían razonablemente salir resultados opuestos? El problema se resuelve con una estimación por intervalo o una prueba de hipótesis sobre b. Para ello se utiliza la "t" de Student, que en este caso sería:

$$t = \frac{b - \beta}{s_b}$$

donde b es el coeficiente de regresión calculado a partir de la muestra experimental, β es el valor real correspondiente (que no puede ser conocido), y s_b es la desviación típica de b en el muestreo (la de las diferentes b que se obtendrían con diferentes muestras). Se puede demostrar que:

$$s_b = \frac{s_{y.x}}{\sqrt{\sum x^2}} \quad ; \quad s_{y.x}^2 = \frac{\sum d_{y.x}^2}{n-2} \quad ; \quad d_{y.x} = Y - \hat{Y}$$

Los grados de libertad de la t así calculada son n-2 porque en su cálculo intervienen dos promedios que son $\bar{y}$ y b. En el ejemplo que nos ocupa tendríamos:

$$s_b = \frac{1.26}{\sqrt{1000}} = 0.04 \quad ; \quad s_{y.x}^2 = \frac{4.8}{5-2} = 1.6 \quad ; \quad t = \frac{0.42 - 0}{0.04} = 10.5$$

Este valor de t es para la hipótesis de que no hay ninguna dependencia de la Y respecto de la X. Contrastado con t = 3.182 que dan las tablas para 3 grados de libertad y 5% de significación vemos que es improbabilísimo que β = 0. La estimación por intervalo de b, con este nivel de significación (5 %) quedaría:

$b - t \cdot s_b < \beta <= b + t \cdot s_b$

0.42 - 3.182 * 0.04 <= β <= 0.42 + 3.182 * 0.04
0.293 <= β <= 0.547

con una probabilidad del 95%.

3.3. PREDICCION DE LA Y A PARTIR DE LA X

El problema que queda por resolver es el de la estimación por intervalo de Y a partir de X. Hemos visto que Y nos daba una estimación por punto de Y utilizando la recta de regresión. Podemos estimar por intervalo dos cosas:

a) El valor medio de Y para un determinado valor de X
b) El valor de Y para un X tomado al azar.

Ambas se estiman por intervalo con ayuda de la "t" de Student pero con denominadores diferentes.

(a) $\quad t = \dfrac{\bar{y}_x - \bar{\mu}_x}{s_{\bar{y}_x}}$ (b) $\quad t = \dfrac{y_x - \mu_x}{s_{y_x}}$

Se puede demostrar que

$$s_{\bar{y}_x} = s_{y.x}\sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2}} \qquad\qquad s_{y_x} = s_{y.x}\sqrt{1 + \frac{1}{n} + \frac{x^2}{\sum x^2}}$$

con lo que:

(a) $\quad \bar{y}_x - t \cdot s_{\bar{y}_x} \leq \bar{\mu}_x \leq \bar{y}_x + t \cdot s_{\bar{y}_x}$

(b) $\quad y_x - t \cdot s_{y_x} \leq \mu_x \leq y_x + t \cdot s_{y_x}$

La t se toma con (n-2) grados de libertad.

Veamos como quedaría esto en un caso práctico. Tenemos una industria textil y queremos estudiar la calidad de los talleres de confección autónomos. Para ello nos interesa determinar, primero si el porcentaje de prendas defectuosas depende del número de prendas que produce cada taller y, en caso de que así sea, poder estimar por intervalo el porcentaje de prendas defectuosas para los talleres que producen 30000 prendas. Para el experimento, seleccionamos al azar 12 talleres, averiguamos su número producción diaria y su % de defectuosas. Los datos figuran en la tabla que sigue, así como los cálculos para llegar a la recta de regresión.

Taller Nº	Miles Prendas X_i	% defec. Y_i	$X_i - \bar{x}$ x	$Y_i - \bar{y}$ y	x^2	xy	$\hat{Y}$	$d_{y.x}$	$d^2_{y.x}$
1	7	60	-12,58	16,33	158,26	-205,43	52,71	7,29	53,19
2	8	56	-11,58	12,33	134,10	-142,78	51,99	4,01	16,09
3	10	51	-9,58	7,33	91,78	-70,22	50,55	0,45	0,20
4	20	49	0,42	5,33	0,18	2,24	43,37	5,63	31,73
5	15	48	-4,58	4,33	20,98	-19,83	46,96	1,04	1,08
6	16	46	-3,58	2,33	12,82	-8,34	46,24	-0,24	0,06
7	19	41	-0,58	-2,67	0,34	1,55	44,09	-3,09	9,52
8	25	40	5,42	-3,67	29,38	-19,89	39,78	0,22	0,05
9	20	37	0,42	-6,67	0,18	-2,80	43,37	-6,37	40,54
10	25	35	5,42	-8,67	29,38	-46,99	39,78	-4,78	22,80
11	30	32	10,42	-11,67	108,58	-121,60	36,18	-4,18	17,50
12	40	29	20,42	-14,67	416,98	-299,56	29,00	0,00	0,00
Suma	235	524	0,04	-0,04	1002,92	-933,67	524,00	0,00	192,77
Media	19,58	43,67	0,00	0,00	83,58	-77,81	43,67	0,00	16,06

$$b = \frac{-933,67}{1002,92} = -0,718 \qquad\qquad s^2_{y.x} = \frac{192,77}{12 - 2} = 19,77$$

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

$\hat{Y}$ = -0,718 * X + 57,736

$$s_{\bar{y}_x} = \sqrt{19.77}\ \sqrt{1/12 + x^2/1002.92} = \sqrt{1.6475 + 0.0197 \cdot x^2}$$

Y para los talleres con 30000 prendas: X = 30, x = 30-19,58 = 10,42

$\hat{Y}$ = -0,718 * 30 + 57,736= 36,18 %

$$t \cdot s_{\bar{y}_x} = 2.228 \cdot \sqrt{1.6475 + 0.0197 \cdot 10.42^2} = 4.34\ \%$$

$$36.18 - 4.34 \leq \bar{\mu} \leq 36.18 + 4.34$$

$$31.84\ \% \leq \bar{\mu} \leq 40.52\ \%$$

Este será el porcentaje medio de prendas defectuosas de los talleres con 30000 prendas. Si se deseara estimar el porcentaje de prendas defectuosas de un taller escogido al azar que resultase producir 30000 prendas habría que aplicar la relación (b).

4. CORRELACION

La idea básica de la correlación es similar a la de la regresión, diferenciándose esencialmente en que la correlación persigue llegar a determinar si cuando una variable evoluciona la otra también lo hace o no, mientras que la regresión lo que busca es explicar la evolución de una variable en función de la otra. Podría decirse que en la regresión hay implícita una búsqueda de la causalidad mientras que en la correlación se presupone que ambas variables evolucionan independientemente pero con cierto paralelismo. No obstante, regresión y correlación pueden estudiarse juntas y aplicarse al mismo fenómeno, según veremos, porque la teoría lo permite y la práctica lo aconseja en ocasiones.
Por ejemplo, tomados al marido y a la mujer de cada una de 11 familias elegidas al azar en una población se midió la estatura de ambos. A la vista de la tabla de datos parecía que ambas variables crecían con cierto paralelismo. Es fácil darse cuenta que una de las variables no puede ser en ningún caso causa de la otra. Ante un caso como este nos formulamos dos preguntas: ¿Cómo medir el grado de paralelismo? ¿Podría este aparente paralelismo haberse producido por casualidad, y salir lo contrario en otra muestra? Veamos como contestar a estas preguntas.

Familia número	1	2	3	4	5	6	7	8	9	10	11
Marido X1	170	169	165	166	171	172	174	172	171	165	156
Mujer X2	168	163	161	160	166	163	167	164	161	160	160

Para medir el grado de correlación entre dos variables X1 y X2 se utiliza el llamado "coeficiente de correlación" que se define así:

$$r = \frac{\sum x_1 \cdot x_2}{\sqrt{(\sum x_1^2)\,(\sum x_2^2)}}$$

Obsérvese que, cuando no hay una relación de tipo causa-efecto clara entre dos variables (y en todos los casos) se podría ajustar una recta de regresión que nos diera x_2 en función de x_l o bien una recta de regresión que nos diera x_l en función de x_2.

Los respectivos coeficientes de regresión serían:

$$b_{21} = \frac{\sum x_1 \cdot x_2}{\sum x_1^2} \qquad b_{12} = \frac{\sum x_1 \cdot x_2}{\sum x_2^2}$$

de donde resulta que el coeficiente de correlación es la media geométrica de los dos coeficientes de regresión, es decir:

$$r = \sqrt{b_{21} \cdot b_{12}}$$

En el caso que nos ocupa, compruébese que:

n=11; $\bar{x}_1 = 168.27$; $\bar{x}_2 = 163.00$; $\sum x_1^2 = 256.27$; $\sum x_2^2 = 89.33$; $\sum x_1 \cdot x_2 = 98.53$

$$r = \frac{98.53}{\sqrt{256.27 \cdot 89.33}} = 0.660$$

Se puede demostrar que el coeficiente de correlación varía entre -1 y +1 siendo estos dos valores los de la máxima correlación (positiva: x_l y x_2 crecen o decrecen a la vez, o negativa: cuando x_l crece x_2 decrece o viceversa), y correspondiendo el cero a los datos sin ninguna correlación. En el ejemplo que nos ocupa r = 0.660 es una correlación positiva y de tipo medio. No obstante surge la pregunta: ¿Podría haber salido así casualmente? Para contestar debemos aprender a hacer pruebas de hipótesis y estimaciones por intervalo de *r*.

Para ello se puede utilizar la "z" de Fisher. Según demostró Fisher el parámetro *z* obtenido partir de *r* por la relación:

z = (1/2)*[L(1+r)-L(1-r)] (L = logaritmo neperiano) (1)

se distribuye casi normalmente, con varianza = 1/(n-3). Habida cuenta que la Normal es equivalente a la "t" con infinitos grados de libertad, podemos hacer estimaciones por intervalo y pruebas de hipótesis sobre z como en otras ocasiones. Veámoslo.

Para buscar la equivalencia entre la *r* y la *z* se suele usar una tabla pero también puede usarse la fórmula anterior a falta de tabla. Para *r* = 0.660 sale un valor de *z* = 0.793, y puesto que n = 11, la varianza será: $s^2 = 1/8$ y la desviación típica $s = 1/(2\sqrt{2})$. Como la "t" con

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

infinitos grados de libertad (o la normal), para un grado de confianza del 95%, es 1.96, una estimación por intervalo con una confianza del 95% de z sería:

$$0.793 - 1.96 * 1/(2\sqrt{2}) \leq z \leq 0.793 + 1.96 * 1/(2\sqrt{2})$$

$$0.100 \leq z \leq 1.486$$

Despejando r de (1) tenemos: $$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Transformando ahora z en r tenemos: $0.099 \leq r \leq 0.903$ intervalo que no incluye al cero. Por consiguiente podemos rechazar la hipótesis de que $r = 0$ a ese nivel de significación.

Para probar otras hipótesis (si se tiene alguna razón para ello) puede usarse la "t" de Student como en otras ocasiones.

5. ANALISIS DE VARIANZA

En 2.2.4.2 vimos como comprobar que dos muestras podían ser consideradas como sacadas de la misma población normal. Imaginemos que tenemos más de dos muestras y que a cada muestra se le ha aplicado un tratamiento distinto, algo parecido al ejemplo de los motores de 2.2.3.2 pero con más de dos motores. El problema ahora es si los tratamientos producen efectos similares o no, y en caso de producirlos, entre qué tratamientos existe diferencia y cuanta.

Para resolverlo vamos a utilizar el siguiente ejemplo:
se trata de determinar si cuatro laboratorios: A, B, C, y D, producen diferentes resultados al analizar la pureza de determinado producto químico o no. Para ello se eligen 20 frascos al azar de la cadena de producción y se forman también al azar cuatro lotes de cinco frascos cada uno que se envían a los diferentes laboratorios. Los resultados figuran en la tabla siguiente.

	a = 4 tratamientos o lotes				
% de materia activa	A	B	C	D	
m = 5 repeticiones o individuos por lote	49	38	20	26	
	35	38	42	32	
	55	29	26	37	
	31	50	48	44	
	44	54	48	30	
$\sum X$	214	209	184	169	
$\bar{x}$	42,8	41,8	36,8	33,8	38,8
$\sum X^2$	9548	9145	7448	5905	
$(\sum X)^2 / m$	9159,2	8736,2	6771,2	5712,2	
$\sum x^2$	388,8	408,8	676,8	192,8	
s^2	97,2	102,2	169,2	48,2	

Obsérvese que $\sum x^2 = \sum X^2 - (\sum X)^2 / m$

Esta relación se demuestra sin más que desarrollar $\sum x^2 = \sum (X - \bar{x})^2 = ...$

Obsérvese también que dentro del mismo lote las diferencias entre unos resultados y otros solamente pueden ser debidas al azar. Por otra parte, cuando promediamos los resultados de un determinado lote lo que hacemos es eliminar el efecto del azar.

Supongamos ahora que no hay ninguna diferencia entre los cuatro laboratorios. Los datos obtenidos, tal como están dispuestos nos permiten hacer al menos dos estimaciones de la varianza poblacional:

1) Promediando las varianzas de los lotes (con ello tenemos una estimación mejor del efecto del azar):

$s^2 = [97.2 + 102.2 + 169.2 + 48.2]/4 = 104.2$

2) Estimando la varianza de las medias de lotes y considerando que es una estimación de s^2/n (recuérdese el denominador de la "t" de Student), podemos despejar s^2 y tenemos otra estimación de la varianza poblacional (al promediar se elimina el efecto del azar y queda solo el efecto de los tratamientos):

$$s_{\bar{x}} = s / \sqrt{n} \qquad ; \qquad s_{\bar{x}}^2 = s^2 / n \qquad ; \qquad s^2 = s_{\bar{x}}^2 \cdot n$$

$$s_{\bar{x}}^2 = [42.8-38.8)^2+(41.8-38.8)^2+(36.8-38.8)^2+(33.8-38.8)^2]/(4-1) = 18$$

$$s^2 = 18 * 5 = 90$$

Ahora, para saber si el efecto de los tratamientos difiere del efecto del azar aplicamos la prueba de la F de Snedecor como en 2.2.4.2, comparando la estimación a partir de las medias de lotes (que no toma en cuenta las variabilidad debida a diferencias entre individuos) con la estimación obtenida promediando las varianzas de los lotes (que toma en cuenta únicamente la variabilidad debida a diferencias entre individuos dentro del mismo lote, es decir, el azar).

$$F = \frac{104.2}{90} = 1.158,$$

valor mucho menor que el que da la tabla de la F para 16 g.l. de la varianza estimada mayor y 3 g.l. para la varianza estimada menor y un nivel de significación del 95%, que es F = 8.69, lo cual nos indica que, efectivamente, no hay diferencia entre los cuatro laboratorios en cuanto a calidad de los análisis.

5.1. Detección de diferencias significativas entre tratamientos

En el ejemplo anterior no había diferencia entre los tratamientos (los laboratorios) pero, en el caso de que la prueba F hubiese dado positivo ¿Como saber entre qué laboratorios está la diferencia? Para contestar a esta pregunta vamos a presentar un nuevo experimento.

Un agricultor tiene el siguiente problema: duda entre cuatro clases de insecticida para combatir una determinada plaga. Si su grado de eficacia fuese diferente el agricultor se inclinaría por el más eficaz. Para averiguarlo se rocían seis parcelas seleccionadas al azar con cada uno de los insecticidas y se determina el porcentaje de plantas libres de plaga. Los resultados del experimento aparecen en la tabla siguiente.

% de materia activa	a = 4 tratamientos o lotes				
	A	B	C	D	
m = 6 repeticiones o individuos por lote	73	95	91	73	
	66	93	80	65	
	74	85	75	50	
	58	75	65	62	
	96	80	74	53	
$\sum X$	432	503	462	370	
$\bar{x}$	72	83,83	77,00	61,67	73,625
$\sum X^2$	31966	42549	35936	23196	
$(\sum X)^2 / m$	31104	42168,17	35574,00	22816,67	
$\sum x^2$	862	380,83	362,00	379,33	
s^2	172,4	76,17	72,40	75,87	

1) Promediando las varianzas de los lotes (con ello tenemos una estimación del efecto del azar):

$$s^2 = [172.4 + 76.17 + 72.40 + 75.87]/4 = 99.21$$

2) Estimando la varianza de las medias de lotes y considerando que es una estimación de s^2/n, podemos despejar s^2 y tenemos otra estimación de la varianza poblacional que solo incluye el efecto de los tratamientos):

$$s_{\bar{x}}^2 = [72-73.625)^2+(83.83-73.625)^2+(77.00-73.625)^2+(61.67-73.625)^2]/(4-1) = 87.08$$

$$s^2 = 87.08 * 6 = 522.49$$

3) El paso siguiente es calcular:

$$F = \frac{\text{varianza estimada mayor}}{\text{varianza estimada menor}} = \frac{522.49}{99.21} = 5.27$$

y compararlo con el que dan las tablas para 3 y 20 g.l. y para el grado de confianza que se desee (99% por ejemplo), que es 4.94. Este valor 4.94 sería el máximo admisible por efecto del azar con ese grado de confianza, si no hubiese diferencia entre los insecticidas. Como en nuestro experimento ha salido mayor, deducimos que sí que hay diferencia entre los insecticidas en cuanto al grado de eficacia.

Y ahora viene la respuesta a la pregunta que formulamos al principio de esta sección ¿Entre qué insecticidas está la diferencia? o ¿Qué insecticida es el que interesa usar?

Para responder utilizaremos la prueba de Tukey, que consiste en calcular con ayuda de unas tablas y una fórmula la diferencia máxima admisible por efecto del azar entre las medias de lotes y comparar las que hemos detectado en nuestro experimento con ese valor máximo. Las que resulten mayores serán las diferencias significativas. Las otras podrían deberse al azar. Veámoslo.

Insecticida	$\bar{x}$	$\bar{x}$ - 61.67	$\bar{x}$ - 72	$\bar{x}$ - 77
B	83.83	22.16	11.83	6.83
C	77	16.33	5	
A	72	10.33		
D	61.67			

Estas son las diferencias detectadas por el experimento, ordenadas de mayor a menor. Tukey demostró que la diferencia máxima aceptable con un grado de confianza del 95% es:

$$D = Q * s_{\bar{x}}$$

donde Q es obtenida de una tabla, la cual para 4 tratamientos (insecticidas) y 20 grados de libertad (del efecto azar) dice que Q=3.93. En cuanto a la desviación típica de las medias muestrales, (caso de no haber diferencia entre tratamientos), sabemos que es: $s_{\bar{x}} = s/\sqrt{n}$, que en nuestro caso sería:

$s_{\bar{x}} = \sqrt{99.21}/\sqrt{6} = 4.07$, con lo que, D = (3.96)*(4.07) = 16.1

Teniendo en cuenta que 99.21 es la varianza del error y 6 son las repeticiones. Como solo hay una diferencia detectada que supere esa cantidad (el 22.16) deducimos que solo es significativa la diferencia entre los insecticidas B y D. Las otras pueden ser debidas al azar. Como conclusión recomendaríamos usar el insecticida B y no usar el D a pesar de que sus diferencias con el D y el A no sean significativas.

Apéndice 6

Conceptos básicos de la Teoría General de Sistemas

– **Sistema**: conjunto interrelacionado de elementos.

– **Conjunto de elementos**: E={a, b, c, d, …, n}

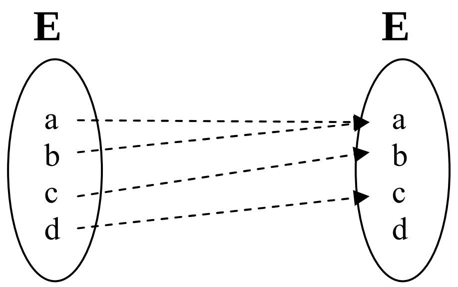
– **Manera de definir conjuntos:**

  - Por extensión (uno a uno todos los alumnos de una clase)
  - Por compresión (estudiantes de ….)

– **Producto cartesiano de dos conjuntos A y B**: conjunto de todas las parejas compuestas por un elemento de A y otro de B. Se representa por A□B. Si A y B coinciden tendremos:

E□E= {(a,a),(a,b),(a,c),…(n,n)}

– **Aplicación entre A y B**: Es un subconjunto del producto cartesiano A□B tal que todos los elementos del conjunto de partida A deben tener una imagen en el conjunto de llegada B. Ejemplo: siendo E={a, b, c, d}, R={(a,a), (b,a), (c,b), (d,c)} es una aplicación de E en E, se representa como R: ETE, y también se escribe: c = R(d).



ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Estructura: Sistema de Conexiones

- **Sistema general**: $S=(E,R)$; $\varnothing \in R \subseteq P[\cup_{n\in Ord} E^n]$
- **Conjunto objeto:** E
- **Conjunto de relaciones:** R
- **Relación estructural:** Considera los elementos de E como singulares.
- **Relación comportamental**: Considera los elementos de E como conjuntos.
- **Definición de sistema-estructura**: $S=(E,R)$; $R\subseteq(E \times E)$
- **Influencia directa de a sobre b**: $(a,b)\in R$
- **Influencia indirecta de a sobre b**:

  $\exists\, x_1, x_2, \dots x_n \in E$; $n \in Z^+ \rightarrow$

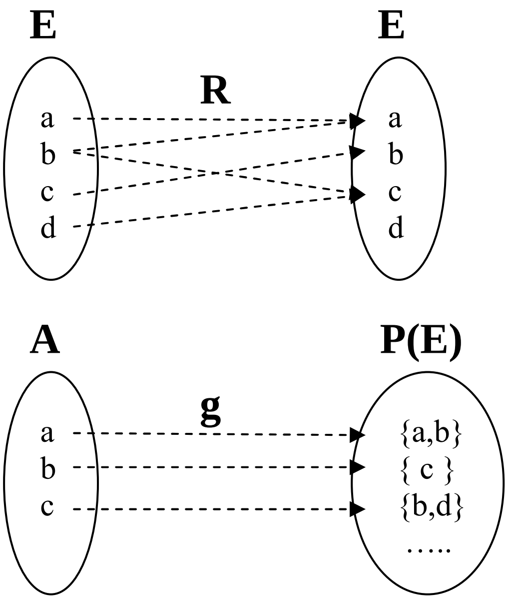
  $(a,x_1) \in R \wedge (x_1,x_2) \in R \wedge \dots \wedge (x_n,b)\in R$
- **Aplicación estructural**: g

  $A=\{a:(a \in E) \wedge (\exists\, x \in E) \wedge (x,a) \in R\}$

  $g:A \rightarrow P(E)$

  $(\forall a \in A)\; g(a)=\{x: (x \in E) \wedge (x,a) \in R\}$

Ejemplo de aplicación estructural:



- **Elemento de entrada**: x

  $(x \in E) \wedge [(\forall(a,b) \in R) \rightarrow (x \neq b)] \wedge [(\exists\ c \in E) \rightarrow (x,c) \in R]$
- **Elemento de salida estricta**: y

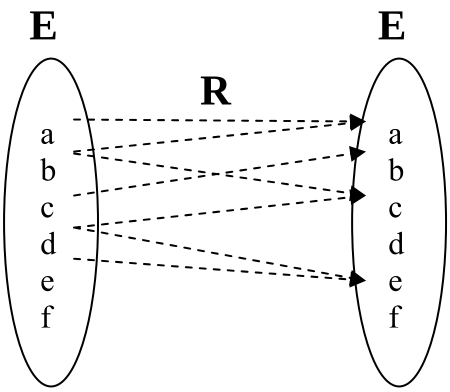
  $(y \in E) \wedge [(\forall(a,b) \in R) \rightarrow (y \neq a)] \wedge [(\exists\ c \in E) \rightarrow (c,y) \in R]$
- **Elemento aislado**: z

  $(z \in E) \wedge [(\forall(a,b) \in R) \rightarrow (z \neq a) \wedge (z \neq b)]$
- **Elemento de salida:** u ; $(u \in E) \wedge (\exists\ x \in E) \wedge (x,u) \in R$
- **Bucle**: A ; $i,j,k,n \in Z^+$

  $(A \subseteq R) \wedge (A=\{(a,x_1), (x_1,x_2),\ldots,(x_i, x_j), (x_j, x_k),\ldots, (x_n, a)\})$
- **Estructura jerárquica**: ausencia de bucles $\Rightarrow$ algoritmo
- **Elemento de nivel n**: x

  $(x=en.n) \Leftrightarrow [\exists\ (a,x) \in R] \wedge [a=en.(n-1)] \wedge [(\forall(b,x) \in R) \wedge (b=en.k) \rightarrow k<n]$

Ejemplo:



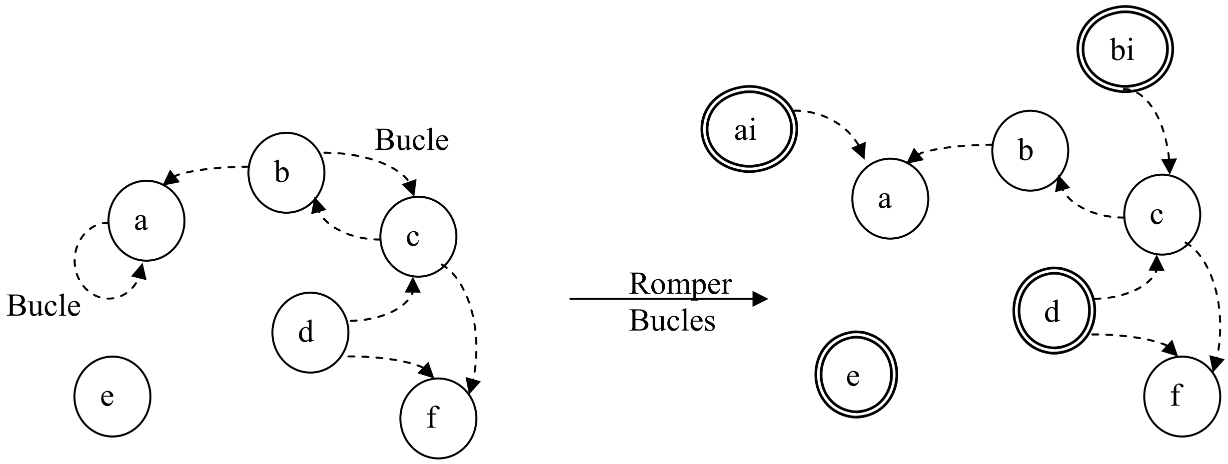
e: elemento aislado (no influye sobre ningún otro y ningún otro influye sobre él)

d: elemento de entrada (influye pero no es influido por otros)

f: elemento de salida estricta (no influye pero sí que es influido)

a, b, c: elementos de salida (influyen y son influidos)

Si hay bucles (retroalimentación), la estructura no es jerárquica. Para convertirla en jerárquica hay que romper los bucles introduciendo variables nuevas. Solo las estructuras jerárquicas dan lugar a algoritmos (conjuntos de operaciones ejecutables por un ordenador o similar).



Nivelación:

Nivel 1: variables de entrada o aisladas: **ai, bi, d, e**.

Nivel 2: variables que solo dependen de las de entrada: **c**.

Nivel 3: variables que dependen de las de nivel 1 y 2: **b, f**.

Nivel 4: variables que dependen de las de nivel 1, 2 y 3: **a**.

La nivelación es necesaria para determinar el orden de cálculo, dado que, solo cuando se conoce el valor de las de nivel inferior se pueden calcular las de nivel superior.

Sistema con Comportamiento

- **Definición** : S=(E,R) donde los elementos de E son conjuntos.
- **Variable**: símbolo x que representa un elemento no determinado de un conjunto D_x .
- **Valor de una variable x**: cualquier elemento de D_x .
- **Dominio de la variable x**: D_x
- **Relación de comportamiento global**: $R \subset D_{x_1} \times D_{x_2} \times \ldots \times D_{x_n}$
- **Sistema con comportamiento**: S=(E,R)

  $E=\{x_1, x_2, \ldots, x_n\}$; $R \subset D_{x_1} \times D_{x_2} \times \ldots \times D_{x_n}$
- **Caja Negra**: S=(E,R) ; $E=\{x_1, x_2, \ldots, x_n\}$

  $E=X \cup Y$; $X \cap Y=\varnothing$

  X={variables de entrada} ; Y={variables de salida}

  R= relación de comportamiento global

Ejemplo:

Dominios de las variables.

$d \in \{3, 2, 1, 4\}$
$e \in \{2, 4, 6, 5,1\}$
$c \in \{3, 2, 1,5\}$
$b \in \{2, 4, 6\}$
$f \in \{1, 10, 20, 30\}$
$a \in \{5, 15, 25\}$

Relación de comportamiento global.

d	e	c	b	f	a
3	1	2	2	1	5
1	2	3	2	20	5
4	6	5	4	20	25
2	4	1	6	30	15
4	5	1	6	10	15

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Sistema Normal

- **Sistema Normal**: $S=(E,R)$; $E=\{x_1,x_2,\ldots, x_n\}$; $R=(c,f)$

  $c\subseteq(E \times E)$; $f\subset D_{x_1} \times D_{x_2} \times \ldots \times D_{x_n}$
- **Comportamiento de una variable "y"**: By

  $y=g^{-1}(x_1,x_2,\ldots, x_k)$; $By \subset D_{x_1} \times D_{x_2} \times \ldots \times D_{x_k} \times D_y$

  $y=By\ (x_1,x_2,\ldots, x_k)$; $By : D_{x_1} \times D_{x_2} \times \ldots \times D_{x_k} \rightarrow D_y$
- **Sistema Normal con incertidumbre**: $S=(E,c,f)$

  $E=\{x_1,x_2,\ldots, x_n\} \wedge [(\exists\ x_i \in E) \rightarrow x_i=\text{variable aleatoria}]$
- **Sistema realista**: $S=(E,c,h)$; $E=\{x_1,x_2,\ldots, x_n\}$; $c\subseteq(E \times E)$

  $h=\{\text{comportamientos de las variables de salida}\}$

Un sistema normal tiene una relación de estructura (producto cartesiano E□E o diagrama causal) y una relación de comportamiento global (tabla de valores). Un sistema realista tiene una relación de estructura y el conjunto de los comportamientos de las variables de salida, que pueden venir dados en forma de tablas, ecuaciones, algoritmos, etc. Le llamamos realista porque es la forma que se suele presentar en la realidad cuando se construyen modelos matemáticos de sistemas reales.

Sistema Dinámico

- **Dirección de cambio**: Conjunto abstracto, D_i, linealmente ordenado con la relación de orden total $\leq$, $i \in Z^+$.
- **Trayectoria**: $(A, \leq)$; $A \subseteq T=D_1 \times D_2 \times \ldots \times D_n$; $n \in Z^+$
- **Sistema dinámico**: $S=(E,R,\Omega,\rho)$

  E= conjunto de variables

  R= relación de comportamiento global

  Ω= conjunto de todas las trayectorias sobre T

  $\rho \subset \Omega \times P(R) \wedge [\forall\ \rho_i \in \rho\ ] \wedge \rho_i =[(A, \leq),B] \Rightarrow \Phi_i :A\rightarrow B$
- **Variable de estado**: *y*

  y forma parte de un bucle.

  Para todo punto de cualquier trayectoria, el valor de *y* depende directa o indirectamente de *y* en un punto anterior.

  Normalmente $\exists y \in E \rightarrow y =$ *variable de estado*

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

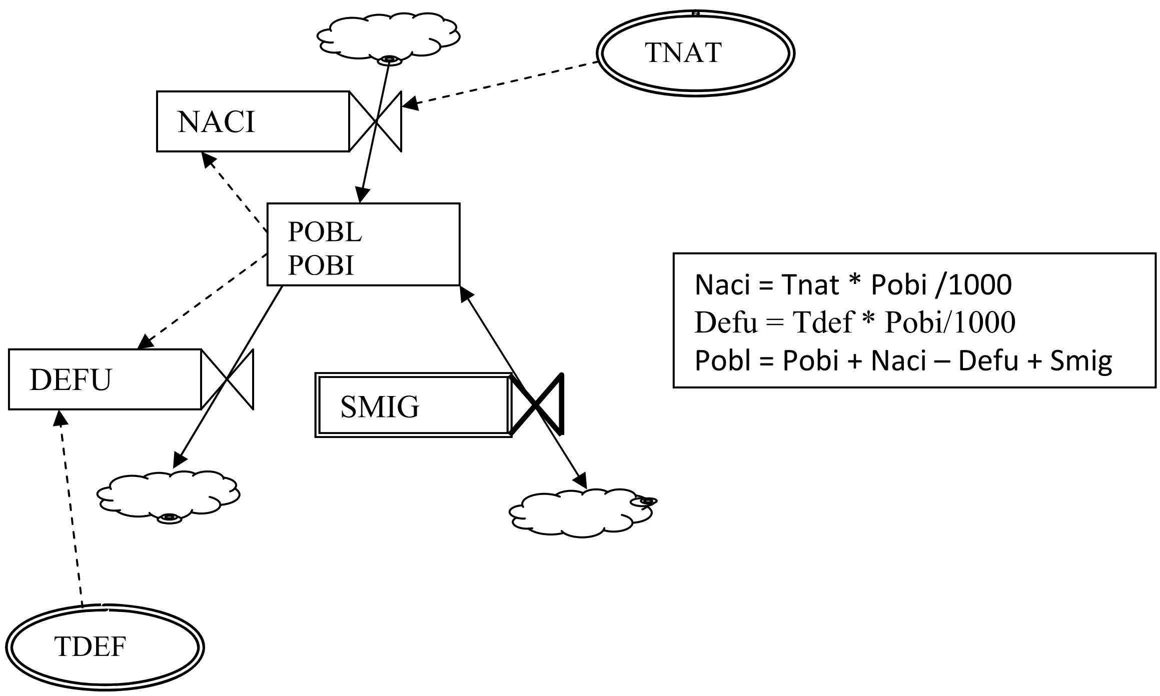
Con mucha frecuencia (casi siempre) nos encontraremos con sistemas dinámicos. La definición clásica de sistema dinámico es aquel que incluye al menos una variable que depende de un valor anterior de ella misma: $x_n = f(x_{n-1})$. Nosotros hemos generalizado esta definición. Así pues, consideramos sistema dinámico aquel en que los valores de las variables cambian con el tiempo, espacio, etc., o cualquier otra variable que cumpla unas determinadas condiciones y a la que llamamos dirección de cambio. Una dirección de cambio es un conjunto de elementos linealmente ordenados con una relación de orden total (siempre sabemos cuál está delante y cuál detrás dados cualesquiera dos puntos del conjunto).

Una trayectoria es un conjunto A con una relación de orden total $\leq$, siendo el conjunto A un subconjunto del producto cartesiano de los dominios de las direcciones de cambio definidas. Por ejemplo, una trayectoria en el espacio de cuatro dimensiones en el que vivimos (tres direcciones espaciales más el tiempo) sería el conjunto de puntos (sus coordenadas) que recorre una mosca en su vuelo unidos a la hora por la que pasa por cada punto.

Un sistema dinámico está compuesto por un conjunto de variables, una tabla de valores para las mismas, un conjunto de trayectorias en un espacio de n dimensiones y un modo (Φ_i) de asignar a cada punto de cada trayectoria un valor a cada variable. Lo que ocurre es que, generalmente, cuando se especifica ese modo de asignar, Φ_i , se utilizan variables de estado, es decir, variables que forman parte de un bucle, o bien, necesitan tener asociada una variable de memoria que guarde un valor anterior de ella misma (cuando se rompe el bucle).

Ejemplo:

Consideremos el modelo clásico de demografía elemental: Pobl (población a final de año) es una variable de estado y su valor inicial es Pobi (población a principio de año, o final del año anterior). El valor de Pobi es un dato para el primer año, y se actualiza cada año.



 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Apéndice 7

Ficheros de entrada y de salida de REGINT para el modelo PAREJAS descrito en 2.4.3

Téngase en cuenta que en estos ficheros que:

1. La anchura de cada columna en los ficheros de datos son 14 caracteres.
2. La variable dependiente es la correspondiente a la última columna.
3. El significado de las variables independientes es el siguiente:
   PARA Número de parados de larga duración
   PEMA Número de personas mayores de 65 años
   NORD Número de ordenadores
   COEN Consumo de energía
   GPRS Gastos de protección social
   PIBH Producto interior bruto por habitante
   YMED Ingreso medio por hogar
   GAST Gasto sanitario
   GMED Gasto medio por habitante
   CJUG Cantidades invertidas en juegos de azar
4. Las unidades de medida han sido ajustadas para que en la tabla no haya grandes diferencias en orden de magnitud.
5. Asumimos lo siguiente:
   - Dada una tabla de valores representando una función de una o varias variables, se pretende encontrar una ecuación del tipo Y=f(x0,x1,...,xn) que explique la tabla de valores de la mejor manera posible.
   - En estas condiciones, REGINT es capaz de encontrar una función del tipo Y=a+b1*T1+b2*T2+... donde, T1, T2, etc. (funciones transformadas) son funciones elementales de x0, x1, etc. (productos, cuadrados, coseno, exponencial, etc.).
   - REGINT además proporciona los datos necesarios para realizar estimaciones por intervalo y simulaciones con la ecuación obtenida, y ofrece la posibilidad de ajustar una función determinada por el usuario entre otros detalles interesantes (exhaustividad o toma de una muestra de funciones, listado total o no, transformadas de segundo orden, etc.)

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Contenido del fichero DDIVOR.DAT que utilizó REGINT para ajustar la función que calcula la tasa de divorcios TDIV.

PARA	PEMA	NORD	COEN	GPRS	PIBH	YMED	GAST	GMED	CJUG	TDIV
12.826	15.01	8.4	2.3278	1.84583	8.79810	11.3000	06.0503	14.47028	4.84	1.715
13.499	15.71	9.3	2.3525	2.08744	9.44455	12.4000	06.8221	15.57274	4.8991	1.854
17.912	15.73	11.5	2.3176	2.30074	9.74083	13.59759	07.3526	16.00624	4.9024	2.019
21.639	16.23	14.5	2.3766	2.32422	10.33214	14.50337	07.6741	16.55675	4.8689	2.101
21.047	16.59	17.2	2.4797	2.45400	11.35385	15.35286	08.4526	17.15058	4.7409	2.132
20.140	15.54	19.2	2.4807	2.57741	12.00266	16.03547	08.9574	17.41761	4.8250	2.254
18.979	15.68	21.9	2.6204	2.64859	12.73062	16.89524	09.3365	17.5500	5.1624	2.374
16.905	16.20	25.5	2.7863	2.74274	13.58169	18.29033	09.9408	17.73092	5.6310	2.430
13.519	16.75	27.9	2.9019	2.87247	14.52498	18.32642	10.6268	18.33034	5.9236	2.567
11.454	16.99	31.5	3.0190	3.06642	15.65320	21.45367	11.2870	19.86265	6.2411	2.564
07.593	17.11	33.3	3.0539	3.25309	16.71547	20.32878	12.0995	20.87892	6.3940	2.831
07.971	17.38	36.3	3.1162	3.48886	17.65033	21.36015	12.7981	21.32028	6.2579	3.072
08.211	17.42	38.9	3.1550	3.70580	18.62966	21.55100	14.6167	22.07157	6.3630	3.217
07.716	17.69	52.3	3.2375	3.93513	19.67842	22.41800	15.8759	23.34093	6.3916	3.499
05.534	18.41	54.9	3.2664	4.07581	20.86389	23.5000	17.000	25.08595	6.5291	3.7
05.534	18.41	54.9	3.2664	4.07581	20.86389	23.5000	17.000	25.08595	6.5291	3.7

Contenido del fichero DMATR.DAT que utilizó REGINT para ajustar la función que calcula la tasa de matrimonios TMAT.

PARA	PEMA	NORD	COEN	GPRS	PIBH	YMED	GAST	GMED	TMAT
12.826	15.01	8.4	23.278	18.4583	8.79810	11.3000	6.0503	14.47028	5.59
13.499	15.71	9.3	23.525	20.8744	9.44455	12.4000	6.8221	15.57274	5.16
17.912	15.73	11.5	23.176	23.0074	9.74083	13.59759	7.3526	16.00624	5.10
21.639	16.23	14.5	23.766	23.2422	10.33214	14.50337	7.6741	16.55675	5.11
21.047	16.59	17.2	24.797	24.5400	11.35385	15.35286	8.4526	17.15058	4.93
20.140	15.54	19.2	24.807	25.7741	12.00266	16.03547	8.9574	17.41761	4.98
18.979	15.68	21.9	26.204	26.4859	12.73062	16.89524	9.3365	17.5500	5.23
16.905	16.20	25.5	27.863	27.4274	13.58169	18.29033	9.9408	17.73092	5.24
13.519	16.75	27.9	29.019	28.7247	14.52498	18.32642	10.6268	18.33034	5.42
11.454	16.99	31.5	30.190	30.6642	15.65320	21.45367	11.2870	19.86265	5.17
7.593	17.11	33.3	30.539	32.5309	16.71547	20.32878	12.0995	20.87892	5.19
7.971	17.38	36.3	31.162	34.8886	17.65033	21.36015	12.7981	21.32028	5.14
8.211	17.42	38.9	31.550	37.0580	18.62966	21.55100	14.6167	22.07157	5.15
7.716	17.69	52.3	32.375	39.3513	19.67842	22.41800	15.8759	23.34093	4.88
5.534	18.41	54.9	32.664	40.7581	20.86389	23.5000	17.000	25.08595	4.7

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Contenido del fichero DIVOR.1 que produjo REGINT al ajustar la función que calcula la tasa de divorcios TDIV.

Caso: TDIV
Identificaci¢n: TASA_DE_DIVORCIOS
Variables independientes: 10
Funciones transformadas: 1
Puntos: 16
La variable x1 es PARA
La variable x2 es PEMA
La variable x3 es NORD
La variable x4 es COEN
La variable x5 es GPRS
La variable x6 es PIBH
La variable x7 es YMED
La variable x8 es GAST
La variable x9 es GMED
La variable x10 es CJUG
La variable Y es TDIV
Variables independientes que han sido seleccionadas:
X1 = PARA
X2 = PEMA
X3 = NORD
X4 = COEN
X5 = GPRS
X6 = PIBH
X7 = YMED
X8 = GAST
X9 = GMED
X10 = CJUG
Funciones elementales que han sido seleccionadas:
1. Identidad
2. Inversa
3. EXP(.1*X)
4. EXP(-.1*X)
5. LOG(X)
10. SQR(X)
Grado de confianza para estimar por intervalo los parámetros: 95.000000
El ajuste se ha realizado:

Con las funciones transformadas:
T1 = exp(-.1*X5)
Coeficientes de la ecuación:
Constante = 11.591269
Coeficiente de T1 = -12.030491
Coeficiente de determinación r2 = 0.985532
Coeficiente de correlaci¢n múltiple r = 0.992740
Error standard de estimación s = 0.080180
MATRIZ C
C(0,0) = 23.606967
MEDIAS DE LAS TRANSFORMADAS
T1

© Antonio Caselles Moncho ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

0.745145
REGRESION Y=a+b1*T1+b2*T2+...+bm*Tm
a = 11.591269   sa = 0.290978   10.967121 <= a <= 12.215417
b1 = -12.030491   sb1 = 0.389571   -12.866121 <= b1 <= -11.194862
s = 0.080180   s2 = 0.006429   r = 0.992740   r2 = 0.985532
t = 2.145000   gr. libertad = 14   gr. confianza = 95.000000
RESIDUOS

Y estimado	RESIDUO	Y
1.588506	0.126494	1.715000
1.827286	0.026714	1.854000
2.033347	-0.014347	2.019000
2.055762	0.045238	2.101000
2.178715	-0.046715	2.132000
2.294161	-0.040161	2.254000
2.360103	0.013897	2.374000
2.446606	-0.016606	2.430000
2.564474	0.002526	2.567000
2.737862	-0.173862	2.564000
2.901595	-0.070595	2.831000
3.104076	-0.032076	3.072000
3.286214	-0.069214	3.217000
3.474506	0.024494	3.499000
3.587894	0.112106	3.700000
3.587894	0.112106	3.700000

Contenido del fichero MATR.1 que produjo REGINT al ajustar la función que calcula la tasa de divorcios TMAT.

Caso: MATR
Identificaci¢n: TASA_DE_MATRIMONIOS
Variables independientes: 9
Funciones transformadas: 2
Puntos: 15
La variable x1 es PARA
La variable x2 es PEMA
La variable x3 es NORD
La variable x4 es COEN
La variable x5 es GPRS
La variable x6 es PIBH
La variable x7 es YMED
La variable x8 es GAST
La variable x9 es GMED
La variable Y es TMAT
Coeficiente de determinaci¢n m¡nimo aceptable: 1
Grado de confianza para estimar por intervalo los par metros: 95.000000
El mejor ajuste se obtiene:
Variables independientes que han sido seleccionadas:
X1 = PARA
X2 = PEMA
X3 = NORD
X4 = COEN
X5 = GPRS
X6 = PIBH
X7 = YMED

X8 = GAST
X9 = GMED
Funciones elementales que han sido seleccionadas:
1. Identidad
2. Inversa
3. EXP(0.100000*X)
4. EXP(-0.100000*X)
5. LOG(X)
6. COS(0.750000 ciclos X)
7. SEN(0.750000 ciclos X)
10. SQR(X)
(Se analizan todas las funciones posibles)
Con las funciones transformadas:
T1 = X4 * X6
T2 = sqr(X9)
Coeficientes de la ecuaci¢n:
Constante = 1.560639858531340e+001
Coeficiente de T1 = 5.676054183811840e-003
Coeficiente de T2 = -2.945479120065560e+000
Coeficiente de determinaci¢n r2 = 0.877583
Coeficiente de correlaci¢n m£ltiple r = 0.936794
Error standard de estimaci¢n s = 0.080517
MATRIZ C
C(0,0) = 0.000094296810
C(0,1) = -0.043011114874
C(1,0) = -0.043011114874
C(1,1) = 20.206693726304
MEDIAS DE LAS TRANSFORMADAS

T1	T2
403.368005	4.333173

REGRESION Y=a+b1*T1+b2*T2+...+bm*Tm
a = 15.606399   sa = 1.258902   12.863250 <= a <= 18.349547
b1 = 0.005676   sb1 = 0.000782   0.003972 <= b1 <= 0.007380
b2 = -2.945479   sb2 = 0.361937   -3.734140 <= b2 <= -2.156818
s = 0.080517   s2 = 0.006483   r = 0.936794   r2 = 0.877583
t = 2.179000   gr. libertad = 12   gr. confianza = 95.000000
RESIDUOS

Y estimado	RESIDUO	Y
5.564317	0.025683	5.590000
5.243980	-0.083980	5.160000
5.103574	-0.003574	5.100000
5.015024	0.094976	5.110000
5.006254	-0.076254	4.930000
5.003660	-0.023660	4.980000
5.160478	0.069522	5.230000
5.351516	-0.111516	5.240000
5.388099	0.031901	5.420000
5.161458	0.008542	5.170000
5.044963	0.145037	5.190000
5.127917	0.012083	5.140000
5.104613	0.045387	5.150000
4.992218	-0.112218	4.880000
4.721928	-0.021928	4.700000

Coeficiente de determinaci¢n R2 = 0.877583

Apéndice 8

Entradas por pantalla y fichero de salida de EXTRAPOL para la tasa de natalidad TNAT en el modelo PAREJAS descrito en 2.4.3

Contenido del fichero de datos para REGINT que se llamará DTNAT.txt

```
Año1=1991      TNAT
1              10.17
2              10.16
3               9.84
4               9.42
5               9.23
6               9.19
7               9.32
8               9.19
9               9.52
10              9.88
11              9.98
12             10.14
13             10.52
14             10.65
15             10.75
```

Contenido del fichero de salida de REGINT que se llamará TNAT1.txt

```
Caso: tnat
Identificación: tasa_de_natalidad_2t
Variables independientes: 1
Funciones transformadas: 2
Puntos: 15
La variable x1 es X1
La variable Y es Y
Coeficiente de determinación mínimo aceptable: 1
Grado de confianza para estimar por intervalo los parámetros:
95.000000
El mejor ajuste se obtiene:
Variables independientes que han sido seleccionadas:
 X1 = X1
Funciones elementales que han sido seleccionadas:
         1. Identidad
         2. Inversa
```

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

```
         3. EXP(0.100000*X)
         4. EXP(-0.100000*X)
         5. LOG(X)
        10. SQR(X)
   (Se analizan todas las funciones posibles)
Con las funciones transformadas:
 T1 = X1 * X1
 T2 = exp(0.100000*X1)
Coeficientes de la ecuación:
 Constante = 1.743494578214888e+001
 Coeficiente de T1 = 9.862427162255570e-002
 Coeficiente de T2 = -6.446550161785034e+000
Coeficiente de determinación r2 = 0.943902
Coeficiente de correlación múltiple r = 0.971546
Error standard de estimación s = 0.137202
MATRIZ C
 C(0,0) = 0.004546066864
 C(0,1) = -0.311982737158
 C(1,0) = -0.311982737158
 C(1,1) = 21.472735164587
MEDIAS DE LAS TRANSFORMADAS
 T1             T2
 82.666667      2.439116
REGRESION Y=a+b1*T1+b2*T2+...+bm*Tm
 a = 17.434946   sa = 0.788982   15.715755 <= a <= 19.154137
 b1 = 0.098624   sb1 = 0.009251   0.078467 <= b1 <= 0.118782
 b2 = -6.446550   sb2 = 0.635774   -7.831902 <= b2 <= -
5.061198
 s = 0.137202    s2 = 0.018824    r = 0.971546    r2 =
0.943902
 t = 2.179000    gr. libertad = 12    gr. confianza =
95.000000
RESIDUOS
 Y estimado      RESIDUO        Y
10.409030      -0.239030      10.170000
9.955609      0.204391      10.160000
9.620632      0.219368      9.840000
9.395811      0.024189      9.420000
9.271988      -0.041988      9.230000
9.239039      -0.049039      9.190000
9.285777      0.034223      9.320000
9.399838      -0.209838      9.190000
9.567557      -0.047557      9.520000
9.773833      0.106167      9.880000
10.001976      -0.021976      9.980000
10.233541      -0.093541      10.140000
10.448143      0.071857      10.520000
10.623253      0.026747      10.650000
10.733973      0.016027      10.750000

 Coeficiente de determinaci¢n R2 = 0.943902
```

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Entradas por pantalla de EXTRAPOL para extrapolar 5 puntos con su intervalo de confianza con la función previamente ajustada con REGINT

PREGUNTA	RESPUESTA
Nombre del problema (hasta 7 caracteres)	TNAT
EL FICHERODE DATOS DE DEBE LLAMAR DTNAT.TXT Primera fila: año o periodo NUMERO (siempre 14 caracteres por columna incluidos espacios en blanco) Resto filas: 1 el primero de la serie 2 el segundo de la serie Etc. etc.	
¿CUÁNTOS PUNTOS DESEA EXTRAPOLAR?	5
¿GRADO DE CONFIANZA? (%) (90 / 95 / 99 / Otro) Si contesta con un valor distinto de 90, 95 ó 99 o si su tabla de datos tiene más de 32 puntos tendrá que introducir posteriormente la t de Student correspondiente	95
Introduzca los coeficientes b0 a b20 de la función a extrapolar previamente obtenida por ajuste de la serie a la sucesión: 1, 2, 3, ... F(X) = b0 + b1*X^b2 + b3*X^b4 + b5*exp(b6*X) + b7*log(X+b8) + b9*cos(b10*X+b11) + b12*cos(b13*X+b14) + b15*X/(b16+X) + b17*exp(b18*X)/(1+b19*exp(b20*X))	
b0 ?	1,743494578214888e+001
b1 ?	1,743494578214888e+001
b2 ?	2
b3 ?	0
b4 ?	0
b5 ?	-6,446550161785034e+000
b6 ?	0,1
Resto hasta b20 ?	0
¿Año o period inicial? (el primero de datos)	1991
En el fichero ETNAT tiene los resultados	

 ISBN: 978-84-370-7198-5 Dep. legal: V-3913-2008

Contenido del Fichero de Salida de EXTRAPOL que se llamará ETNAT.txt

Se extrapolan 5 puntos

El grado de confianza de los intervalos es: 95 %

El valor considerado para la t de Student es: 2,16

El número de puntos considerado es: 15

F(X) = 17.4349457821489 + 9.86242716225557E-02*X^(2) + 0*X^(0) + 6.44655016178503*exp(.1*X) + 0*log(X+0)
+ 0*cos(0*X+0) + 0*cos(0*X+0) +
+ 0*X/(0+X) + 0*exp(0*X)/(1+0*exp(0*X))

Recta de regresion: Y = 1,00000000000036 F(X) + -3,5491609651217E-12

Año o periodo	TENDENCIA	MINIMO	MAXIMO
1991	10,17		
1992	10,16		
1993	9,84		
1994	9,42		
1995	9,23		
1996	9,19		
1997	9,32		
1998	9,19		
1999	9,52		
2000	9,88		
2001	9,98		
2002	10,14		
2003	10,52		
2004	10,65		
2005	10,75		
2006	10,7527873407126	10,4313527525216	11,0742219289036
2007	10,649283837326	10,3336502909297	10,9649173837223
2008	10,3898539474034	10,0859302092903	10,6937776855165
2009	9,93735393933843	9,64309799464345	10,2316098840334
2010	9,25073364117086	8,94333585425644	9,55813142808528