



Universitat d'Alacant
Universidad de Alicante

Esta tesis doctoral contiene un índice que enlaza a cada uno de los capítulos de la misma.

Existen asimismo botones de retorno al índice al principio y final de cada uno de los capítulos.

[Ir directamente al índice](#)

Para una correcta visualización del texto es necesaria la versión de [Adobe Acrobat Reader 7.0](#) o posteriores

Aquesta tesi doctoral conté un índex que enllaça a cadascun dels capítols. Existeixen així mateix botons de retorn a l'índex al principi i final de cadascun dels capítols .

[Anar directament a l'índex](#)

Per a una correcta visualització del text és necessària la versió d' [Adobe Acrobat Reader 7.0](#) o posteriors.



TESIS DOCTORAL

UN NUEVO ALGORITMO PARA LA MODELIZACIÓN DE SISTEMAS ALTAMENTE ESTRUCTURADOS

Dpto. de Análisis Matemático y Matemática Aplicada
Universidad de Alicante

Mónica Cortés Molina
Alicante, 2000





Universitat d'Alacant
Universidad de Alicante

YOLANDA VILLACAMPA ESTEVE, CATEDRÁTICA DEL DEPARTAMENTO DE ANÁLISIS MATEMÁTICO Y MATEMÁTICA APLICADA EN LA ESCUELA POLITÉCNICA SUPERIOR DE LA UNIVERSIDAD DE ALICANTE

CERTIFICA: Que Mónica Cortés Molina, Ingeniero Superior en Informática, ha realizado bajo mi dirección la Memoria que lleva por título "Un nuevo algoritmo para la modelización de sistemas altamente estructurados" con el fin de que sea presentada como Tesis para aspirar al grado de doctor.

Lo que certifico para que conste según la ley vigente.

Alicante, 30 de Mayo de 2000

Fdo. Yolanda Villamcampa Esteve



Universitat d'Alacant
Universidad de Alicante

AGRADECIMIENTOS:

Agradezco la confianza y ayuda prestada por mi directora de tesis, Yolanda Villacampa Esteve, que con su trabajo incansable en la lectura, revisión, crítica y consejos ha hecho posible este proyecto.

También quiero manifestar mi agradecimiento a todas aquellas personas que me han apoyado en todo el proceso de elaboración, en especial a Juan Enrique, y a la atención que me ha ofrecido mi familia.



Universitat d'Alacant
Universidad de Alicante

Contenido

1	Introducción	1
2	Modelización de los sistemas altamente estructurados	7
2.1	Construcción de un lenguaje formal. Vocabularios de orden n	12
2.2	Algunos modelos en Matemáticas	15
2.2.1	Modelos dinámicos	15
2.2.2	Modelos matriciales	16
2.2.3	Modelos estocásticos	17
2.2.4	Modelos multivariantes	18
2.2.5	Modelos de optimización	20
2.2.6	Modelos de la teoría de juegos	21
2.2.7	Modelos de la teoría de catástrofes	21



Contenido

3	Conceptos previos	23
3.1	Estadísticos	23
3.2	Algunas distribuciones	25
3.3	Inferencia estadística	26
3.3.1	Pruebas de hipótesis	27
3.4	Test de significancia para los coeficientes	28
3.5	Prueba de Kolgomorov-Smirnov	30
3.6	Software existente	31
4	Metodología estadística	33
4.1	Planos de regresión	33
4.2	Medidas de la dependencia estadística y de la correlación	45
4.2.1	La varianza residual	46
4.2.2	Coefficiente de determinación ajustado o corregido	48
4.2.3	Coefficiente de determinación normal	49
5	Software	51
5.1	Estructuras de datos	54
5.2	Bloques de funciones	58
5.3	Ficheros generados	93
5.4	Orden de ejecución de las funciones	94



Contenido

5.5	Secuencia de llamadas	104
6	Aplicaciones	109
6.1	Caso lineal: $Y = aX + b$	110
6.1.1	Ejecución con MODELHSS	112
6.1.2	Ejecución con SPLUS	115
6.1.3	Ejecución con SPSS	119
6.2	Caso lineal múltiple	124
6.2.1	Ejecución con MODELHSS	125
6.2.2	Ejecución con SPLUS	135
6.2.3	Ejecución con SPSS	138
6.3	Aplicación de MODELHSS a una modelización medioam- biental	141
6.3.1	Ejecución con MODELHSS	145
6.3.2	1ª Ejecución con SPLUS	154
6.3.3	1ª Ejecución con SPSS	159
6.3.4	2ª Ejecución con SPLUS	163
6.3.5	2ª Ejecución con SPSS	167
6.4	Conclusiones	176
7	Conclusiones finales	179



Universitat d'Alacant
Universidad de Alicante

Contenido

8 Anexo	185
8.1 Anexo 1	185



Universitat d'Alacant
Universidad de Alicante

Capítulo 1

Introducción

En los sistemas altamente estructurados, como pueden ser los ecológicos, socioeconómicos, etc., y en general, todo ecosistema, el número de relaciones existentes entre sus distintos elementos, así como los procesos estudiados en los mismos, puede ser muy elevado. El análisis de estos sistemas y sus modelizaciones puede enmarcarse dentro de los estudios realizados en los sistemas llamados sistemas dinámicos, siendo de gran importancia el estudio de su evolución en el tiempo. En las modelizaciones se consideran el estudio y el comportamiento de atributos medibles asociados a los elementos (entes) del sistema. Esto conduce, en la mayoría de los casos, a la existencia de relaciones no lineales que conviene expresar matemáticamente para determinar el comportamiento y evolución de los fenómenos considerados en dichos sistemas.

En este proyecto se presenta una metodología donde, inicialmente, se construye computacionalmente un lenguaje regular con el que imple-



mentar ecuaciones matemáticas, en su mayoría, no lineales. Generadas familias de ecuaciones a partir de datos experimentales se seleccionan una o un conjunto de ellas que se ajusten en mayor o menor medida a los datos experimentales.

La metodología presentada, además, coincide con las metodologías existentes en el mercado, como SPLUS y SPSS, en los casos en que se desee calcular relaciones lineales simples y múltiples. En el caso de ecuaciones no lineales, la construcción de dichas ecuaciones que se ajusten a datos experimentales difiere de dichos programas.

La metodología se ha implementado en un programa al que hemos llamado MODELHSS, utilizando como lenguaje de programación el lenguaje C. La ejecución del programa, dependiendo de las funciones matemáticas elegidas, genera familias de ecuaciones que puedan incluir a dichas funciones y posibles composiciones de ellas (ecuaciones no lineales). De este modo, se dispone de un conjunto de expresiones que se ajustarán en diversos grados a los datos iniciales. Con ello, la variedad de ecuaciones almacenadas posibilitan al modelizador a elegir aquella o aquellas que más se acerquen a su propósito. No por ello tiene que ser la que más se ajuste a los datos, siendo posible la elección de alguna de ellas que de cierta manera, por ejemplo, mejor le sea de representar o estudiar. Aunque en los siguientes estudios se verá, generalmente, las ecuaciones que tienen un mayor nivel de ajuste.

Sin embargo, al ejecutar los programas estadísticos elegidos para comparar con MODELHSS la única forma posible de obtener ecuaciones no lineales es proponer un tipo de ecuación con coeficientes a determinar los cuales serán calculados a partir del método de ajuste de los mínimos cuadrados.



La posibilidad de almacenamiento de un conjunto de ecuaciones que se ajusten de manera más o menos idónea a los datos experimentales supone disponer de una herramienta para obtener la máxima información posible respecto al sistema que se está tratando.

En este proyecto se ha considerado un lenguaje regular pudiéndose en futuras investigaciones determinar computacionalmente otros lenguajes. La elección de éste, en concreto, ha sido motivada porque en las primeras modelizaciones realizadas con el mismo en los ecosistemas los resultados han sido satisfactorios. Si bien, con la misma estructura teórica que nos permite construir este lenguaje, se pueden construir otros mediante la utilización de distintas operaciones matemáticas en el árbol generado para la construcción del lenguaje regular.

En el **Capítulo 2**, “Modelización de los sistemas altamente estructurados”, se consideran los conceptos teóricos básicos de los sistemas altamente estructurados, así como la construcción de lenguajes formales para la modelización de los mismos. Las relaciones que nos determinan el planteamiento teórico de los modelos se han definido a través de la relación de comportamiento y ecuaciones de comportamiento de una variable. Además, en este capítulo se incluye una reseña de algunos modelos matemáticos que con más frecuencia se suelen aplicar en el estudio, en general, de los modelos dinámicos.

En el **Capítulo 3**, “Conceptos previos”, se ha hecho referencia a aquellos conceptos estadísticos que son necesarios para el mejor desarrollo de la metodología presentada: estadísticos, algunas distribuciones como son la distribución normal y la t-Student, la inferencia estadística con las pruebas de hipótesis que tomamos, el test de significancia para los coeficientes y el test de Kolgomorov-Smirnov. Así mismo, realizamos un breve comentario del software existente en el



mercado, como el SPLUS y SPSS.

Dentro del **Capítulo 4**, “Metodología estadística”, comentamos la metodología estadística aplicada al programa para la determinación de las ecuaciones. Se hace referencia al método de los mínimos cuadrados con la particularidad de tipificar y trasladar los datos iniciales de los que se dispone. Además se incluyen las medidas de la dependencia estadística y de la correlación, de modo que nos permitan calibrar el grado de representatividad de la función analítica ajustada a los datos obtenidos empíricamente por observación. Estas medidas vendrían a ser la varianza residual y el coeficiente de determinación.

La presentación del software se incluye en el **Capítulo 5**. Este software ha sido generado para implementar la metodología de forma computacional. En este sentido, hemos dividido el análisis de este programa en: las diversas estructuras de datos que se utilizan, los bloques de funciones que se implementan con distintos objetivos, los posibles ficheros que pueden ser generados, el orden de ejecución de las funciones y la secuencia de llamadas que se produce entre ellas.

En el **Capítulo 6**, “Aplicaciones”, se estudian diversas modelizaciones aplicando el programa MODELHSS y comparando los resultados con SPLUS y SPSS. Así, se empieza considerando los casos más sencillos: un caso lineal, donde existe solamente una variables independiente y una dependiente, y un caso lineal múltiple, donde se dispone de cuatro variables independiente y una dependiente. A continuación se expone una modelización medioambiental a partir de ficheros de datos experimentales de tamaño considerado con varias variables independientes, obteniendo ecuaciones de expresión no lineal. Se presenta algún caso problemático que MODELHSS soluciona, pero tanto SPSS como SPLUS no llegan a resolver correctamente.



En el capítulo, **Capítulo 7**, comentamos las conclusiones del proyecto presentado y las futuras investigaciones a realizar.

Como último capítulo se presenta un anexo en el **Capítulo 8**, donde se adjunta uno de los ficheros de datos empleados en la modelización medioambiental del capítulo de “Aplicaciones”, concretamente corresponde al ajuste del ozono en el área A de noche del mes de Junio, utilizada en la sección 6.3.2.



Universitat d'Alacant
Universidad de Alicante

Capítulo 2

Modelización de los sistemas altamente estructurados

La metodología de modelización considerada en el proyecto presentado, será utilizada de forma general en el estudio y modelización de sistemas, denominados en la literatura usual ([1, 14]) como sistemas dinámicos.

Por **sistema dinámico** se considera a todo sistema en el que las propiedades que se desean estudiar, así como sus relaciones son temporales, considerándose su evolución en el tiempo. De forma concreta esta metodología se construye para el mejor conocimiento de los sistemas dinámicos en los que exista alguna complejidad en sus interrelaciones, ya que en la mayoría de los casos, con los métodos usuales de modelización se pierde mucha información. Así mismo, es aplicada con satisfacción en los sistemas con una menor complejidad.

De forma intuitiva, se considera un sistema como un conjunto de



partes (entes) interrelacionadas. Asociados a estos entes se consideran atributos medibles de los que se desea analizar su comportamiento a través de su modelización matemática. Son precisamente las relaciones formales las que constituyen el modelo matemático. Formalmente consideraremos las definiciones de sistema y de modelo dadas en "Towards a formalization of Ecosystems and ecological models" ([40]), para cualquier sistema altamente estructurado.

Inicialmente se considera un conjunto \mathcal{T} al cual llamamos universo. Este conjunto esta formado por cosas o entes definidas por medio de un observador. Para un subconjunto $S \subset \mathcal{T}$ se consideran algunos atributos medibles y relaciones existentes entre ellos. Estos son la base conceptual de la construcción de un modelo. De esta forma se obtendrá un concepto de sistema a partir de los atributos de un conjunto que está formado por entes o cosas.

Se denomina **sistema conexión simple**, $S \equiv (M, R)$ al par formado por un conjunto M y un conjunto de relaciones binarias R de manera que se cumple que:

$$R \subset P(M \times M) = P(M^2).$$

Es decir:

$$\forall r \in R, \quad r \subset M \times M, \\ r = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots / (x_i, y_i) \in M \times M\}$$

El **sistema conexión simple vacío** está definido por $\emptyset \equiv (\emptyset, \emptyset)$.

Un **sistema conexión** definido en un conjunto M , es un **sistema conexión simple** o la **unión de sistemas conexión**: S es un **sistema conexión** si y solo si S es un **sistema conexión** $S = \cup S_i$



con S_i un sistema simple conexión. Se denotará $S \equiv (M, R)$ con $R \subset P \left(\bigcup_{\text{finitas}} M^2 \right)$, ([40]).

Cuando el sistema objeto M está formado por entes, se dirá que el sistema es un *sistema ontológico* y si M está formado por atributos asociados a los entes el sistema se llamará *semiótico*, ([40]).

Se considera como **sistema altamente estructurado** ([5]), aquellos sistemas en los que el número de variables y **relaciones puede ser elevado**, así como aquellos sistemas en los que en alguna relación exista complejidad. Ejemplos de estos sistemas pueden encontrarse en los sistemas socioeconómicos y los ecosistemas. La complejidad de las relaciones puede ser originada principalmente por la combinación en la descripción de la relación de no linealidad, interacción, retroalimentación y discontinuidad.

A lo largo de este proyecto, al considerar el estudio y modelización de un sistema, haremos referencia a un sistema altamente estructurado, denotado en lo sucesivo por **HSS**. Si bien se debe de tener siempre presente que al considerar un sistema HSS el número de relaciones a tener en cuenta puede ser muy elevado, pero también puede ser pequeño. En ambos casos, la metodología que se expone podrá ser aplicable. Esta metodología nos determina ecuaciones matemáticas que estudiarán la evolución en el tiempo de ciertas propiedades de dichos sistemas.

Se introduce el uso de un vocabulario formal en la modelización de sistemas ecológicos en ([22, 38]), que es aplicable también a los sistemas HSS. En este contexto, a sí mismo, se han definido los **vocabularios de primer y n-ésimo orden** ([39]). La utilización de estos nuevos vocabularios normalmente permite encontrar ecuaciones que obtengan



mejor información a partir de los datos medioambientales en particular, y desde un HSS en general. Por medio de un vocabulario de primer orden, utilizado correctamente en modelizaciones ecológicas, se puede obtener cualquier vocabulario de orden n -ésimo iterativamente ([41]). Esto nos conducirá a la obtención de un conjunto de lenguajes que serán utilizados para construir ecuaciones matemáticas que formarán los modelos de los sistemas HSS.

Supongamos que tenemos un HSS y un proceso que empezaremos definiéndolo por medio de un conjunto de variables $V_x = \{x_i\}_{i=1}^p$. Nuestro propósito es estudiar el comportamiento y la variación de estas variables en el tiempo.

Para cualquier conjunto de variables dado $V_x = \{x_i\}_{i=1}^p$, el **sistema de comportamiento asociado**, ABS_x , se define como el sistema formado por cualquier variable que tiene alguna influencia en el comportamiento de las variables, V_x . Claramente, estas variables pueden ser de varios tipos: variables internas del HSS, variables externas, etc.

Para un sistema de comportamiento ABS_x se dan las siguientes definiciones:

- **Relación comportamiento:** Una relación de comportamiento se define como cualquier relación $F(z_1, z_2, \dots, z_p)$ para un conjunto de variables dado $\{x_i\}_{i=1}^p$ en el sistema de comportamiento asociado ABS_x .
- **Ecuación comportamiento de la variable:** Para cualquier $x_i \in V_x$, la ecuación comportamiento asociada a x_i es aquella relación de comportamiento con la que se puede expresar cada



variable:

$$x_i = F^i(x_1^i, x_2^i, \dots, x_n^i) \quad \{x_j^i\}_j \in ABS_x$$

Para analizar un HSS, se supondrá que se dispone de un conjunto de ecuaciones comportamiento 2.1 asociado a las variables $\{x_i\}_{i=1}^p$.

$$\begin{cases} x_1 = F^1(x_1^1, x_2^1, \dots, x_{i_1}^1) \\ x_2 = F^2(x_1^2, x_2^2, \dots, x_{i_2}^2) \\ \vdots \\ x_p = F^p(x_1^p, x_2^p, \dots, x_{i_p}^p) \end{cases} \quad (2.1)$$

Estas ecuaciones tienen una expresión analítica desconocida, aunque los correspondientes datos experimentales asociados sí que se conocen. La metodología presentada en este proyecto construye un lenguaje computacional donde las ecuaciones comportamiento asociadas a las variables pertenecientes a un HSS se expresarán en términos de funciones matemáticas.

$$\{z'_{(k)}(t)\}_{k=1}^m \quad (2.2)$$

Un caso particular es aquel en el que el sistema se modeliza a partir de un sistema de ecuaciones diferenciales 2.2. Cada ecuación de 2.2 se puede expresar por medio de una función, que resultará ser una combinación de funciones de comportamiento:

$$\{z'_{(k)}(t)\}_k = F_k(x'_{(k_1)}(t), x'_{(k_2)}(t), \dots, x'_{(k_n)}(t)), \quad (2.3)$$

$$F_k : \mathbb{R}^n \rightarrow \mathbb{R}$$



Las variables diferenciales dadas en 2.2 se denominan normalmente **variables de estado** y las ecuaciones dadas en 2.3 **ecuaciones de flujo**, donde $x'_{(k_j)}(t)$ son las **variables de flujo** ([22]). Dichas variables de flujo dependen de un conjunto de variables, cuyas dependencias han sido estudiadas teóricamente en ([42]).

Si se tiene la siguiente ecuación comportamiento de la variable $x'_{(k_j)}(t)$,

$$x'_{(k_j)}(t) = f(x_{(k_j)}^1(t), x_{(k_j)}^2(t), \dots, x_{(k_j)}^1(t)), \quad (2.4)$$

entonces la metodología presentada en este trabajo se puede utilizar para calcular algunas ecuaciones matemáticas que expresen 2.4. Estas ecuaciones definirán el comportamiento de las variables de flujo. El conocimiento de las mismas permite obtener las ecuaciones 2.2 y 2.3 que se resolverán utilizando métodos numéricos de resolución de ecuaciones diferenciales.

De este modo, nuestra metodología basada en la construcción de lenguajes formales para determinar ecuaciones de comportamiento, en general, pueden utilizarse para resolver diversos problemas definidos en los sistemas HSS.

2.1 Construcción de un lenguaje formal. Vocabularios de orden n

Inicialmente, se consideran un conjunto de funciones elegidas por el modelizador. Este conjunto de funciones constituye lo que es llamado



vocabulario de orden 1, es decir, V^1 :

$$V^1 = \{T_1, T_2, \dots, T_n\} \quad (2.5)$$

Este vocabulario 2.5 se ha denominado, en estudios teóricos realizados en modelos ecológicos, **funciones transformadas de orden 1** ([41]).

Los vocabularios de mayor orden se pueden generar iterativamente por medio de la operación matemática *. En este proyecto, será la composición la operación matemática utilizada, pudiéndose construir nuevos lenguajes regulares para distintas operaciones. El vocabulario de orden 2 se genera a partir del vocabulario de orden 1 de la expresión 2.5 de la siguiente manera: los elementos de V^2 son todos los posibles $T_i * T_j, \forall i, j$. El resto de los vocabularios se obtienen iterativamente del mismo modo, lo que implica, en términos de programación, una estructura en forma de árbol.

El lexicón $L = \{V^1, V^2, \dots, v^n\}$ se construye a partir de todos los vocabularios generados $V^i, i = 1, \dots, n$ ([42, 39]). En la modelización de un HSS, dichos vocabularios actúan sobre un conjunto de variables $\{x_i\}_n$. De este modo, se puede considerar el lexicón asociado a una variable x definido por el conjunto de vocabularios actuando sobre x

$$L(x) = \left\{ \bigcup_{i=1}^m V^i(x) \right\}$$

$L(x)$ es el vocabulario utilizado por cada variable x para construir ecuaciones matemáticas que fijen el conjunto de datos y permita al investigador hacer predicciones. $L(x)$ es un lenguaje regular, lo cual, permite su correcta construcción computacional.

La modelización matemática de los HSS como la de todo sistema, nos conduce a la expresión formal de las relaciones en un lenguaje



matemático. La relación más simple utilizada para describir las relaciones en un sistema es la determinada a través de modelos de regresión lineal, que vienen siempre definidos por ecuaciones del tipo:

$$y = A + B_1x_1 + B_2x_2 + \dots + B_mx_m,$$

para la variable aleatoria y distribuida alrededor de su media, dependiente de las variables $\{x_i\}_{i=1,2,\dots,m}$.

En los sistemas HSS la gran parte de sus relaciones no se ajustan de forma lineal. Ejemplos sencillos de modelos no lineales, han sido considerados en el estudio de la densidad de población y el tiempo en ([22]), para el que se ha utilizado la función de crecimiento exponencial $x(t) = x_0e^{rt}$. Debido a que dicho crecimiento se estabiliza, el cual terminará cuando se agoten los recursos, también ha sido utilizado como modelo la función de crecimiento logístico, $x(t) = x_0/(1 - ke^{-rt})$. De este tipo son los modelos realizados por Leith, H. ([43]), que modeliza las relaciones de productividad primaria, medida como producción anual de materia seca, y el clima a partir de dos relaciones no lineales:

$$Y = \frac{3000}{1 + \exp(1.315 - 0.119T)},$$

$$Y = 3000(1 - \exp(-0.00064R))$$

siendo \mathbf{Y} la producción anual de materia seca en g/m^2 , \mathbf{T} la temperatura anual media en $^{\circ}C$ y \mathbf{R} la precipitación anual media en mm . Con estas ecuaciones, considerando además el valor mas bajo en el caso de que no coincidan, Leith ha construido mapas mundiales de la producción potencial de materia seca.



2.2 Algunos modelos en Matemáticas

El proceso de modelización siempre debería empezar con una clara exposición de los objetivos de la investigación, de los cuales la modelización es una parte ([22]). Los modelos que pueden encontrarse con más frecuencia son los siguientes:

- (a) Modelo dinámicos.
- (b) Modelos matriciales.
- (c) Modelos estocásticos.
- (d) Modelos multivariantes.
- (e) Modelos de optimización.
- (f) Modelos de teoría de juegos.
- (g) Modelos de la teoría de catástrofes.

2.2.1 Modelos dinámicos

Están basados en la teoría del servomecanismo. El uso de los modelos en cualquier aplicación práctica depende de la capacidad de los modernos ordenadores digitales de alta velocidad para resolver miles de ecuaciones en cortos períodos de tiempo. Estas ecuaciones son descripciones matemáticas más o menos complejas de la operación del sistema que se va a simular, y tienen la forma de expresiones para niveles de diferentes tipos, cuyo cambio es controlado por funciones de decisión.



Las ecuaciones de nivel serían acumulaciones dentro del sistema de variables y las ecuaciones de control dirigen los cambios de los niveles a lo largo del tiempo. Las funciones de decisión son las reglas que controlan el funcionamiento del sistema.

Los modelos dinámicos poseen una gran flexibilidad para describir las operaciones del sistema y permite la introducción de la no linealidad y la retroalimentación tanto positiva como negativa.

Estas características implican algunas desventajas, como por ejemplo el no poder incluir ecuaciones para todos los componentes del sistema, ya que la simulación se hace demasiado compleja, y, principalmente, la incertidumbre sobre si serán capaces de estimar los valores de los parámetros básicos.

Los modelos dinámicos pueden considerarse útiles en las primeras etapas del análisis de sistemas de un problema ecológico complejo, ya que concentran la atención en las relaciones básicas del sistema y definen las variables y subsistemas que el investigador considera críticos. Sin embargo, en etapas posteriores del análisis a menudo es preferible centrarse en otras familias de modelos.

2.2.2 Modelos matriciales

Los modelos matriciales constituyen una familia de modelos en la que, hasta cierto punto, la realidad es sacrificada a fin de obtener ventajas con las propiedades matemáticas de la formulación.

La propia formulación también limita la forma en la que los modelos pueden ser utilizados, pero esto queda compensado por la comodidad



en los cálculos y por la relativa facilidad a la hora de establecer los valores de los parámetros básicos.

2.2.3 Modelos estocásticos

Son el desarrollo lógico de las matemáticas aplicadas, es decir, las matemáticas que se aplican a la física. Son análogos matemáticos de los procesos físicos, en los que hay una correspondencia de uno a uno entre causa y efecto. Permiten expresar las relaciones en términos de probabilidades, de manera que la respuesta del modelo no es siempre la misma. Son muy útiles para simular la variabilidad y complejidad de ciertos tipos de sistemas como son los ecológicos.

Dentro de estos modelos podemos encontrar además:

- (a) *Modelos de distribución.* Como ejemplo, podemos citar la representación de los patrones espaciales de los organismos vivos.
- (b) *Modelos de análisis de la varianza.* Uno de los modelos más utilizados en la investigación científica, basado en el análisis de la varianza. Proporcionan un método para construir modelos de poblaciones ecológicas y para estimar los parámetros de los modelos a partir de las observaciones de las muestras.
- (c) *Modelos de regresión.* Los modelos lineales de análisis de la varianza son un caso particular de los modelos de regresión, que se representan por la siguiente expresión:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$



En esta expresión, y es una variable aleatoria distribuida alrededor de su media que depende de los valores de las p variables $x_1 \dots x_p$. Se considera que estas variables sólo afectan a la media de y , y que la varianza es constante. Cuando se realizan pruebas de significación, se acepta que y tiene una distribución normal alrededor de esta media. La media puede ser contemplada como una función lineal de las variables x_i , aunque también pueda haber relaciones funcionales entre ellas, de manera que en los modelos más generales pueden incluirse funciones polinomiales y otras funciones no lineales. Los parámetros del modelo se estiman generalmente minimizando la suma residual de cuadrados para una muestra de la población.

- (d) *Modelos de Markov*. Son un híbrido entre los modelos matriciales discutidos en la sección 2.2.2 y los modelos estocásticos discutidos en esta sección. El formato básico de estos modelos es el de una matriz que expresa las probabilidades de transición de un estado a otro en una etapa de tiempo concreta. Estos modelos son similares a los modelos matriciales, excepto en que las probabilidades de cada columna suman uno.

2.2.4 Modelos multivariantes

Una variante es una cantidad que puede tomar cualquiera de los valores de un conjunto con una determinada frecuencia relativa o probabilidad. Estas variantes se denominan también variables aleatorias y se definen, no únicamente por un conjunto de posibles valores como cualquier variable matemática ordinaria, sino por una función asociada de frecuencia o probabilidad que indica con qué frecuencia aparecen estos valores en



la aplicación objeto de estudio. Por ejemplo, hay muchas situaciones en ecología en las que los modelos deben captar el comportamiento de más de una variante; estos modelos se conocen en conjunto con el nombre de multivariantes y están relacionados con técnicas que, en conjunto, se denominan análisis multivariantes.

Los modelos multivariantes se pueden dividir en dos grandes categorías:

1. Modelos predictivos: Los que utilizan algunas variables para predecir otras. Se dividen según:
 - (a) El número de variables que predicen y
 - (b) Si todos los predictores son cuantitativos o no.
2. Modelos descriptivos: Los que tienen todas las variables del mismo tipo, y no predicen los valores de unas a partir de los de las otras. A su vez se dividen en:
 - (a) Los datos de entrada son cuantitativos.
 - (b) Los datos de entrada son cualitativos: modelo de ponderación recíproca.

Generalmente se encuentran, dentro de los modelos multivariantes, los siguientes tipos:

1. Análisis de componentes principales.
2. Análisis de clúster.
3. Análisis de ponderación recíproca y asociación.



4. Función discriminante.
5. Análisis de correlaciones canónicas.

2.2.5 Modelos de optimización

Se utilizan para encontrar el máximo o el mínimo de una determinada expresión o función matemática dando valores a ciertas variables que pueden variar libremente dentro de los límites marcados.

Se dividen en:

1. *Programación lineal*. Parte de una función lineal objetiva:

$$Y = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_{i=1}^n a_ix_i$$

de donde se pretende obtener un máximo o un mínimo sujetos a una o más restricciones.

2. *Programación no lineal*. Introduce la no linealidad en la función de objetivos, en las restricciones o en ambos, alcanzando niveles bastantes desproporcionados de dificultad a la hora de encontrar soluciones apropiadas.
3. *Programación dinámica*. Algunos problemas de optimización pueden ser reformulados en forma de una serie de pequeños problemas agrupados en secuencias de espacio, tiempo o ambos, para reducir el esfuerzo de cálculo a la hora de encontrar una solución.



2.2.6 Modelos de la teoría de juegos

Están muy relacionados con los modelos de programación matemática. Representan una aproximación interesante y poco explorada a la resolución de problemas estratégicos.

2.2.7 Modelos de la teoría de catástrofes

La teoría de catástrofes es un elegante desarrollo de la topología matemática aplicada a sistemas que tienen cuatro propiedades básicas: bimodalidad, discontinuidad, histéresis y divergencia.



Universitat d'Alacant
Universidad de Alicante

Capítulo 3

Conceptos previos

La técnica utilizada en el proyecto para la obtención de familias de ecuaciones lineales y no lineales que se ajusten en cierta medida a los datos experimentales, implica la aplicación de ciertas medidas estadísticas. A continuación resumiremos aquellos conceptos y resultados que se utilizarán posteriormente en este proyecto.

Con el objetivo de obtener información respecto a los parámetros desconocidos de la población se definen funciones sobre las variables aleatorias o **estadísticos**.

3.1 Estadísticos

Para una muestra $x_{i=1}^n$ serán utilizados los siguientes:



1. La media muestral de una muestra de tamaño n , definido por el estadístico:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

2. La varianza muestral, que para una muestra de tamaño n se define por:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

3. La desviación estándar, representado por S , que en tal caso coincide con la raíz cuadrada positiva de la varianza muestral, es decir:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

4. La covarianza entre dos variables x e y de las que se dispone de dos muestras del mismo tamaño $x_{i=1}^n$ y $y_{i=1}^n$:

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Para el análisis y ajuste de datos numéricos y, en definitiva, el de muestras de una población, considerada como conjunto de la totalidad de observaciones, resulta conveniente el estudio de algunas distribuciones.



3.2 Algunas distribuciones

Podemos destacar la **distribución normal**. Las mediciones físicas en áreas tales como los experimentos meteorológicos, los estudios acerca de las lluvias y las mediciones sobre partes manufacturadas se explican con una **distribución normal** en forma más que adecuada. Además, los errores en las mediciones científicas se aproximan hasta límites extremadamente pequeños gracias a la **distribución normal**.

La ecuación matemática para la distribución de probabilidad de la variable normal depende de su media (μ) y su desviación estándar (σ). Su función de densidad tiene la siguiente expresión:

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}, \quad -\infty < x < \infty.$$

Esta curva normal cumple las siguientes propiedades:

1. La moda, que es el punto sobre el eje horizontal donde la curva tiene su máximo, ocurre en $x = \mu$.
2. La curva es simétrica alrededor de su eje vertical donde se tiene la media μ .
3. La curva tiene sus puntos de inflexión en $x = \mu \pm \sigma$, es cóncava hacia abajo si $\mu - \sigma < X < \mu + \sigma$, y es cóncava hacia arriba en cualquier otro punto.
4. La curva normal se acerca al eje horizontal en forma asintótica en cualquiera de las dos direcciones, alejándose de la media.



5. El área total bajo la curva y arriba del eje horizontal es igual a 1.

Otra distribución importante que cabe destacar y de la que también se hará uso es la **distribución t-Student** o **distribución t**. Si el tamaño de la muestra es pequeño los valores de las varianzas poblacionales, S^2 , varían considerablemente de muestra a muestra y la distribución de la variable aleatoria $(\bar{X} - \mu)(S/\sqrt{n})$ se desvía en forma apreciable de una distribución normal estándar. Ahora se trata de una distribución de un estadístico llamada T , donde,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Al derivar la distribución muestral de T , se asumirá que la muestra aleatoria se seleccionó de una población normal. Se puede expresar entonces:

$$T = \frac{(\bar{X} - \mu)/(\delta/\sqrt{n})}{\sqrt{S^2/\delta^2}} = \frac{Z}{\sqrt{V/(n-1)}}$$

De modo que la distribución de la variable aleatoria T está dada por:

$$h(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < t < \infty,$$

que se conoce como **distribución t** con ν grados de libertad.

3.3 Inferencia estadística

La inferencia estadística consiste en aquellos métodos con lo cuales se pueden realizar generalizaciones acerca de una población. Existen,



generalmente, dos métodos para realizar la inferencia. Nos centraremos en el método clásico, por medio del cual las inferencias se basan en la información obtenida de una muestra aleatoria seleccionada de la población.

La inferencia estadística puede dividirse, principalmente, en 2 áreas: estimación y prueba de hipótesis. Se tratará principalmente la prueba de hipótesis, pues no se intentará estimar un parámetro, sino tomar una decisión correcta respecto a la hipótesis preestablecida, en nuestro caso, suponer que los coeficientes obtenidos en la ecuación son significativos.

3.3.1 Pruebas de hipótesis

La estructura de la prueba de hipótesis se formulará utilizando el término **hipótesis nula**. Esto se refiere a cualquier hipótesis que se desee probar y se representa por H_0 . El rechazo de H_0 da como resultado la aceptación de una **hipótesis alternativa**, representada por H_1 . Una hipótesis nula referente a un parámetro poblacional siempre será establecida en forma tal que especifique un valor exacto del parámetro, mientras que la hipótesis alternativa admite la posibilidad de varios valores. De aquí que, si H_0 es la hipótesis nula de que un coeficiente α_i no es significativo, es decir, que $\alpha = 0$, la hipótesis alternativa H_1 sería que sí que es significativo y por lo tanto $\alpha > 0$.

Cuando se rechaza la hipótesis nula siendo verdadera se dice que se está cometiendo un **error tipo I**. La probabilidad de cometer un error de este tipo se conoce como **error de significancia** y se representa por α .

En lo referente a la significancia de los coeficientes, estaríamos ante



una prueba de hipótesis estadística con alternativa bilateral, tal como:

$$\begin{aligned}H_0 &: \theta = \theta_0, \\H_1 &: \theta \neq \theta_0.\end{aligned}$$

Esta hipótesis estadística recibe el nombre de **prueba de dos colas**, ya que la hipótesis alternativa $\theta \neq \theta_0$ establece que $\theta < \theta_0$ ó $\theta > \theta_0$.

3.4 Test de significancia para los coeficientes

Se puede utilizar el estadístico

$$T = \frac{B_j - \beta_j}{S\sqrt{c_{jj}}}$$

con $n - k - 1$ grados de libertad, donde:

- B_j : es el coeficiente de x_j .
- β_j : es el valor hipotético que queremos que tome.
- S : Raíz cuadrada de la varianza residual.



- c_{jj} : Elemento (j, j) de A^{-1} , donde A es la matriz de varianza-covarianza de los coeficientes de regresión estimados:

$$A = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \dots & \sum_{i=1}^n x_{ki} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \dots & \sum_{i=1}^n x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \dots & \sum_{i=1}^n x_{ki}^2 \end{bmatrix}$$

- n : Número de datos de los que se dispone.
- k : Número de variables independientes.

De modo, que en nuestro caso, si queremos probar:

$$H_0 : \beta_j = 0,$$

$$H_1 : \beta_j \neq 0,$$

siendo la hipótesis nula H_0 la de que un coeficientes no es significativo, se calcula la estadística

$$t = \frac{B_j}{s\sqrt{c_{jj}}}$$

y no se rechaza H_0 si

$$-t_{\alpha/2} < t < t_{\alpha/2},$$

donde $t_{\alpha/2}$ tiene $n - k - 1$ grados de libertad. Además se eligirá un nivel de significancia de 0.05, es decir, $\alpha = 0.05$.



3.5 Prueba de Kolgomorov-Smirnov

Consiste en contrastar si dos muestras proceden de la misma distribución. Se basa en la diferencia máxima absoluta entre las funciones de distribución acumulada observadas para ambas muestras. Cuando esta diferencia es significativamente grande, se consideran diferentes las dos distribuciones.

D se define como el valor máximo de la diferencia absoluta entre dos funciones de distribución acumulativas. Así, para comparar un conjunto de datos $S_N(x)$ frente a una función de distribución acumulativa conocida $P(x)$, el estadístico $K - S$ es:

$$D = \max_{-\infty < x < \infty} |S_N(x) - P(x)|$$

La función que introduce el cálculo de la significancia se puede escribir con el siguiente sumando:

$$Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 \lambda^2}$$

la cual se corresponde con una función monótona con límites

$$Q_{KS}(0) = 1 \quad Q_{KS}(\infty) = 0$$

El nivel de significancia de un valor observado de D se da aproximadamente por la fórmula:

$$\text{Probabilidad (D > observado)} = Q_{KS} \left(\left[\sqrt{N_e} + 0.12 + 0.11/\sqrt{N_e} \right] D \right)$$

donde N_e es el número de datos.



En nuestro caso se comprobará si los residuos se adaptan a una distribución normal, para lo que se comprobará si la probabilidad es mayor o menor que 0.05. En el caso de que sea mayor se adaptará a una distribución normal, en caso contrario no.

3.6 Software existente

Existen dos programas en el mercado con los que es posible encontrar expresiones matemáticas que se ajusten en mayor o menor medida a datos experimentales, el SPLUS y SPSS, que serán utilizados para hacer comparaciones con la metodología presentada.

Brevemente podemos describir la metodología utilizada por dichos programas en los siguientes puntos:

- Ambos programas, al igual que nuestra metodología, construyen correctamente regresiones lineales simples y múltiples. Además, se pasan los correspondientes estadísticos que nos dicen la significancia de los coeficientes.
- Respecto a la obtención de ecuaciones no lineales múltiples podemos hacer algunas distinciones: En SPLUS y SPSS la forma de obtener una ecuación no lineal que se ajuste a los datos experimentales existentes se realiza proponiendo una ecuación no lineal con coeficientes a determinar, los cuales, se determinan en base a la tabla de datos. Esto supone la ejecución de dichos programas cada vez que se quiera obtener una ecuación no lineal. En nuestra metodología, en una primera etapa, hemos construido un lenguaje regular a partir de la concatenación de funciones a través



de una misma operación matemática *. Con dicho lenguaje podemos obtener familias de funciones en cada ejecución del programa. Lo que supone disponer de una mayor información respecto a los datos existentes. En una segunda etapa serán construidos otros lenguajes regulares.

- También se observará en las aplicaciones del software presentado que podemos resolver algunas de las deficiencias que surgen al ejecutar los programas SPSS y SPLUS en tablas de datos que presenten ciertas restricciones.



Universitat d'Alacant
Universidad de Alicante

Capítulo 4

Metodología estadística

En este capítulo veremos el método estadístico de cálculo de regresiones por mínimos cuadrados, así como las medidas que nos determinan la dependencia estadística y la correlación.

4.1 Planos de regresión

La correlación múltiple se encarga de estudiar la dependencia entre n variables.

Se denotará por $x_1 = x_1(x_2, x_3, \dots, x_n)$ un conjunto de puntos situados en una superficie denominada *superficie de regresión* de x_1 sobre (x_2, x_3, \dots, x_n) .

Se tienen, además, los siguientes datos numéricos:



x_1	x_2	x_3	\cdots	x_n
x_{11}	x_{21}	x_{31}	\cdots	x_{n1}
x_{12}	x_{22}	x_{32}	\cdots	x_{n2}
x_{13}	x_{23}	x_{33}	\cdots	x_{n3}
\vdots	\vdots	\vdots	\ddots	\vdots
x_{1m}	x_{2m}	x_{3m}	\cdots	x_{nm}

$$\sum_{i=1}^m x_{1i} \quad \sum_{i=1}^m x_{2i} \quad \sum_{i=1}^m x_{3i} \quad \cdots \quad \sum_{i=1}^m x_{ni}$$

Cuyas medias son:

$$\bar{x}_1 = \frac{\sum_{i=1}^m x_{1i}}{m}; \quad \bar{x}_2 = \frac{\sum_{i=1}^m x_{2i}}{m}; \quad \bar{x}_3 = \frac{\sum_{i=1}^m x_{3i}}{m}; \quad \cdots; \quad \bar{x}_n = \frac{\sum_{i=1}^m x_{ni}}{m};$$

Varianzas:

$$s_{x_1}^2 = \frac{\sum_{i=1}^m (x_{1i} - \bar{x}_1)^2}{m-1}; \quad s_{x_2}^2 = \frac{\sum_{i=1}^m (x_{2i} - \bar{x}_2)^2}{m-1}; \quad s_{x_3}^2 = \frac{\sum_{i=1}^m (x_{3i} - \bar{x}_3)^2}{m-1};$$

$$s_{x_n}^2 = \frac{\sum_{i=1}^m (x_{ni} - \bar{x}_n)^2}{m-1}$$

Realizamos un cambio de variable (desviaciones con respecto a la



media):

$$\begin{aligned}
 z_1 &= \frac{x_1 - \bar{x}_1}{s_1} + 4; & z_2 &= \frac{x_2 - \bar{x}_2}{s_2} + 4; \\
 z_3 &= \frac{x_3 - \bar{x}_3}{s_3} + 4; \quad \dots; & z_n &= \frac{x_n - \bar{x}_n}{s_n} + 4
 \end{aligned} \tag{4.1}$$

Destacar que:

$$\begin{aligned}
 \sum_{i=1}^m (x_{1i} - \bar{x}_1) &= 0 \\
 \sum_{i=1}^m (x_{2i} - \bar{x}_2) &= 0 \\
 \sum_{i=1}^m (x_{3i} - \bar{x}_3) &= 0 \\
 &\vdots \\
 \sum_{i=1}^m (x_{ni} - \bar{x}_n) &= 0
 \end{aligned} \tag{4.2}$$

Con lo cual, aplicando la ecuación 4.2:



$$\begin{aligned}
 \sum_{i=1}^m z_{1i} &= \sum_{i=1}^m \left(\frac{x_{1i} - \bar{x}_1}{s_{x_1}} + 4 \right) = \\
 &= \left(\frac{x_{11} - \bar{x}_1}{s_{x_1}} + 4 \right) + \left(\frac{x_{12} - \bar{x}_1}{s_{x_1}} + 4 \right) + \dots + \left(\frac{x_{1m} - \bar{x}_1}{s_{x_1}} + 4 \right) = \\
 &= \frac{x_{11} - \bar{x}_1}{s_{x_1}} + 4 + \frac{x_{12} - \bar{x}_1}{s_{x_1}} + 4 + \dots + \frac{x_{1m} - \bar{x}_1}{s_{x_1}} + 4 = \\
 &= \frac{1}{s_{x_1}} ((x_{11} - \bar{x}_1) + (x_{12} - \bar{x}_1) + \dots + (x_{1m} - \bar{x}_1)) + 4 + 4 + \dots + 4 = \\
 &= \frac{1}{s_{x_1}} \overbrace{\sum (x_1 - \bar{x}_1)}^0 + 4 \overbrace{(1 + 1 \dots + 1)}^m = 4m
 \end{aligned}$$

$$\begin{aligned}
 \sum_{i=1}^m z_{2i} &= \sum_{i=1}^m \left(\frac{x_{2i} - \bar{x}_2}{s_{x_2}} + 4 \right) = \\
 &= \left(\frac{x_{21} - \bar{x}_2}{s_{x_2}} + 4 \right) + \left(\frac{x_{22} - \bar{x}_2}{s_{x_2}} + 4 \right) + \dots + \left(\frac{x_{2m} - \bar{x}_2}{s_{x_2}} + 4 \right) = \\
 &= \frac{x_{21} - \bar{x}_2}{s_{x_2}} + 4 + \frac{x_{22} - \bar{x}_2}{s_{x_2}} + 4 + \dots + \frac{x_{2m} - \bar{x}_2}{s_{x_2}} + 4 = \\
 &= \frac{2}{s_{x_2}} ((x_{21} - \bar{x}_2) + (x_{22} - \bar{x}_2) + \dots + (x_{2m} - \bar{x}_2)) + 4 + 4 + \dots + 4 = \\
 &= \frac{2}{s_{x_2}} \overbrace{\sum (x_2 - \bar{x}_2)}^0 + 4 \overbrace{(1 + 1 \dots + 1)}^m = 4m
 \end{aligned}$$

⋮


 \vdots

$$\begin{aligned}
 \sum_{i=1}^m z_{ni} &= \sum_{i=1}^m \left(\frac{x_n - \bar{x}_n}{s_{x_n}} + 4 \right) = \\
 &= \left(\frac{x_{n1} - \bar{x}_n}{s_{x_n}} + 4 \right) + \left(\frac{x_{n2} - \bar{x}_n}{s_{x_n}} + 4 \right) + \dots + \left(\frac{x_{nm} - \bar{x}_n}{s_{x_n}} + 4 \right) = \\
 &= \frac{x_{n1} - \bar{x}_n}{s_{x_n}} + 4 + \frac{x_{n2} - \bar{x}_n}{s_{x_n}} + 4 + \dots + \frac{x_{nm} - \bar{x}_n}{s_{x_n}} + 4 = \\
 &= \frac{n}{s_{x_n}} ((x_{n1} - \bar{x}_n) + (x_{n2} - \bar{x}_n) + \dots + (x_{nm} - \bar{x}_n)) + 4 + 4 + \dots + 4 = \\
 &= \frac{2}{s_{x_n}} \overbrace{\sum (x_n - \bar{x}_n)}^0 + 4 \overbrace{(1 + 1 \dots + 1)}^m = 4m
 \end{aligned}$$

de modo que:

$$\bar{z}_1 = 4$$

$$\bar{z}_2 = 4$$

$$\bar{z}_3 = 4$$

 \vdots

$$\bar{z}_4 = 4$$



Las nuevas varianzas resultarían:

$$\begin{aligned}
 s_{z_1}^2 &= \frac{\sum_{i=1}^m (z_{1i} - \bar{z}_1)^2}{m-1} = \frac{\sum_{i=1}^m \left(\frac{x_{1i} - \bar{x}_1}{s_{x_1}} + 4 - 4 \right)^2}{m-1} = \\
 &= \frac{\sum_{i=1}^m \left(\frac{x_{1i} - \bar{x}_1}{s_{x_1}} \right)^2}{m-1} = \frac{\sum_{i=1}^m \frac{(x_{1i} - \bar{x}_1)^2}{s_{x_1}^2}}{m-1} \\
 &= \frac{1}{(m-1)s_{x_1}^2} \left(x_{11} - \bar{x}_1 \right)^2 + (x_{12} - \bar{x}_1)^2 + \dots + (x_{1m} - \bar{x}_1)^2 = \\
 &= \frac{1}{(m-1)s_{x_1}^2} \left(x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2 - 2\bar{x}_1(x_{11} + x_{12} + \dots + x_{1m}) + m\bar{x}_1^2 \right) = \\
 &= \frac{m}{(m-1)s_{x_1}^2} \left(\frac{x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2}{m} - 2\bar{x}_1^2 + \bar{x}_1^2 \right) = \\
 &= \frac{m}{(m-1)s_{x_1}^2} \left(\frac{x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2}{m} - \bar{x}_1^2 \right) = \\
 &= \frac{m}{(m-1)} \frac{(m-1)}{\sum_{i=1}^m (x_{1i} - \bar{x}_1)^2} \left(\frac{x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2}{m} - \bar{x}_1^2 \right) = \\
 &= \frac{m}{(m-1)} \frac{(m-1)}{m \left(\frac{x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2}{m} - \bar{x}_1^2 \right)} \left(\frac{x_{11}^2 + x_{12}^2 + \dots + x_{1m}^2}{m} - \bar{x}_1^2 \right) = 1
 \end{aligned}$$



Generalizando:

$$s_{z_2}^2 = \frac{\sum_{i=1}^m (z_{2i} - \bar{z}_2)^2}{m-1} = \frac{\sum_{i=1}^m \left(\frac{x_{2i} - \bar{x}_2}{s_{x_2}} + 4 - 4 \right)^2}{m-1} = 1$$

$$s_{z_3}^2 = \frac{\sum_{i=1}^m (z_{3i} - \bar{z}_3)^2}{m-1} = \frac{\sum_{i=1}^m \left(\frac{x_{3i} - \bar{x}_3}{s_{x_3}} + 4 - 4 \right)^2}{m-1} = 1$$

⋮

$$s_{z_n}^2 = \frac{\sum_{i=1}^m (z_{ni} - \bar{z}_n)^2}{m-1} = \frac{\sum_{i=1}^m \left(\frac{x_{ni} - \bar{x}_n}{s_{x_n}} + 4 - 4 \right)^2}{m-1} = 1$$

y las covarianzas

$$\begin{aligned} s_{1,2} &= \frac{\sum z_1 z_2}{m-1} = \frac{\sum z_2 z_1}{m-1} = s_{2,1} \\ s_{1,3} &= \frac{\sum z_1 z_3}{m-1} = \frac{\sum z_3 z_1}{m-1} = s_{3,1} \\ s_{1,n} &= \frac{\sum z_1 z_n}{m-1} = \frac{\sum z_n z_1}{m-1} = s_{n,1} \end{aligned}$$

Considerando las nuevas variables z_1, z_2, \dots, z_n si suponemos que la superficie de regresión es un plano, habrá que ajustar la nube de puntos por un plano de ecuación:

$$z_1 = \alpha_1 + \alpha_2 z_2 + \alpha_3 z_3 + \dots + \alpha_n z_n$$

con la condición de que

$$D = \sum (z_1 - \alpha_1 - \alpha_2 z_2 - \alpha_3 z_3 \dots - \alpha_n z_n)^2$$



sea mínima.

Utilizando el procedimiento de mínimos cuadrados, derivando respecto a $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$ e igualando a cero:

$$\frac{\partial D}{\partial \alpha_1} = 0; \quad \frac{\partial D}{\partial \alpha_2} = 0; \quad \dots; \quad \frac{\partial D}{\partial \alpha_n} = 0;$$

se obtiene un sistema de n ecuaciones con n incógnitas $(\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$, cuya resolución nos dará el plano de regresión de x_1 sobre (x_2, x_3, \dots, x_n) pedido.

Otro procedimiento es mediante las ecuaciones normales:

$$\begin{aligned} \sum (z_1 - \alpha_1 - \alpha_2 z_2 - \alpha_3 z_3 - \dots - \alpha_n z_n) &= 0 \\ \sum z_2 z_1 &= \sum z_2 (\alpha_1 - \alpha_2 z_2 - \alpha_3 z_3 - \dots - \alpha_n z_n) \\ \sum z_3 z_1 &= \sum z_3 (\alpha_1 - \alpha_2 z_2 - \alpha_3 z_3 - \dots - \alpha_n z_n) \\ &\vdots \\ \sum z_n z_1 &= \sum z_n (\alpha_1 - \alpha_2 z_2 - \alpha_3 z_3 - \dots - \alpha_n z_n) \end{aligned}$$

operando

$$\begin{aligned} \sum z_2 z_1 &= \alpha_2 \sum z_2^2 + \alpha_3 \sum z_2 z_3 + \dots + \alpha_n \sum z_2 z_n \\ \sum z_3 z_1 &= \alpha_2 \sum z_3 z_2 + \alpha_3 \sum z_3^2 + \dots + \alpha_n \sum z_3 z_n \\ \sum z_4 z_1 &= \alpha_2 \sum z_4 z_2 + \alpha_3 \sum z_4 z_3 + \dots + \alpha_n \sum z_4 z_n \\ &\vdots \\ \sum z_n z_1 &= \alpha_2 \sum z_n z_2 + \alpha_3 \sum z_n z_3 + \dots + \alpha_n \sum z_n^2 \end{aligned}$$

de donde

$$\begin{aligned} s_{21} &= \alpha_2 s_2^2 + \alpha_3 s_{23} + \dots + \alpha_n s_{2n} \\ s_{31} &= \alpha_2 s_{23} + \alpha_3 s_3^2 + \dots + \alpha_n s_{3n} \\ s_{41} &= \alpha_2 s_{24} + \alpha_3 s_{34} + \dots + \alpha_n s_{4n} \\ &\vdots \\ s_{n1} &= \alpha_2 s_{2n} + \alpha_3 s_{3n} + \dots + \alpha_n s_n^2 \end{aligned}$$



Llamaremos r_{ij} al coeficiente de correlación entre z_i y z_j , que es:

$$r_{ij} = \frac{s_{ij}}{s_{z_i} s_{z_j}}$$

Por comodidad de escritura, a partir de ahora se denotará:

$$s_i \Leftrightarrow s_{z_i}$$

resultando

$$\begin{aligned} r_{21}s_2s_1 &= \alpha_2s_2^2 + \alpha_3r_{23}s_2s_3 + \dots + \alpha_nr_{2n}s_2s_n \\ r_{31}s_3s_1 &= \alpha_2r_{23} + \alpha_3s_3^2 + \dots + \alpha_nr_{3n}s_3s_n \\ r_{41}s_4s_1 &= \alpha_2r_{24} + \alpha_3r_{34}s_3s_4 + \dots + \alpha_nr_{4n}s_4s_n \\ &\vdots \\ r_{n1}s_ns_1 &= \alpha_2r_{2n} + \alpha_3r_{3n}s_3s_n + \dots + \alpha_ns_n^2 \end{aligned}$$

Resolviendo por Cramer resulta:

$$\alpha_2 = \frac{\begin{vmatrix} r_{21}s_2s_1 & r_{23}s_2s_3 & \dots & r_{2n}s_2s_n \\ r_{31}s_3s_1 & s_3^2 & \dots & r_{3n}s_3s_n \\ r_{41}s_4s_1 & r_{24}s_2s_4 & \dots & r_{4n}s_4s_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}s_ns_1 & r_{2n}s_2s_n & \dots & s_n^2 \end{vmatrix}}{\begin{vmatrix} s_2^2 & r_{23}s_2s_3 & \dots & r_{2n}s_2s_n \\ r_{23}s_2s_3 & s_3^2 & \dots & r_{3n}s_3s_n \\ r_{24}s_2s_4 & r_{34}s_3s_4 & \dots & r_{4n}s_4s_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n}s_2s_n & r_{3n}s_3s_n & \dots & s_n^2 \end{vmatrix}} = \frac{s_1}{s_2} \begin{vmatrix} r_{21} & r_{23} & \dots & r_{2n} \\ r_{31} & 1 & \dots & r_{3n} \\ r_{41} & r_{24} & \dots & r_{4n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{2n} & \dots & 1 \end{vmatrix}$$



$$\alpha_3 = \frac{\begin{vmatrix} s_2^3 & r_{21}s_2s_1 & \cdots & r_{2n}s_2s_n \\ r_{23}s_2s_3 & r_{31}s_3s_1 & \cdots & r_{3n}s_3s_n \\ r_{24}s_2s_4 & r_{41}s_4s_1 & \cdots & r_{4n}s_4s_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n}s_2s_n & r_{n1}s_ns_1 & \cdots & s_n^2 \end{vmatrix}}{\begin{vmatrix} s_2^2 & r_{23}s_2s_3 & \cdots & r_{2n}s_2s_n \\ r_{23}s_2s_3 & s_3^2 & \cdots & r_{3n}s_3s_n \\ r_{24}s_2s_4 & r_{34}s_3s_4 & \cdots & r_{4n}s_4s_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n}s_2s_n & r_{3n}s_3s_n & \cdots & s_n^2 \end{vmatrix}} = \frac{s_1}{s_3} \begin{vmatrix} 1 & r_{21} & \cdots & r_{2n} \\ r_{23} & r_{31} & \cdots & r_{3n} \\ r_{24} & r_{41} & \cdots & r_{4n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n} & r_{n1} & \cdots & 1 \end{vmatrix}$$

⋮

$$\alpha_n = \frac{\begin{vmatrix} s_2^3 & r_{23}s_2s_3 & \cdots & r_{21}s_2s_1 \\ r_{23}s_2s_3 & s_3^2 & \cdots & r_{31}s_3s_1 \\ r_{24}s_2s_4 & r_{24}s_2s_4 & \cdots & r_{41}s_4s_1 \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n}s_2s_n & r_{2n}s_2s_n & \cdots & r_{n1}s_ns_1 \end{vmatrix}}{\begin{vmatrix} s_2^2 & r_{23}s_2s_3 & \cdots & r_{2n}s_2s_n \\ r_{23}s_2s_3 & s_3^2 & \cdots & r_{3n}s_3s_n \\ r_{24}s_2s_4 & r_{34}s_3s_4 & \cdots & r_{4n}s_4s_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n}s_2s_n & r_{3n}s_3s_n & \cdots & s_n^2 \end{vmatrix}} = \frac{s_1}{s_n} \begin{vmatrix} 1 & r_{23} & \cdots & r_{21} \\ r_{23} & 1 & \cdots & r_{31} \\ r_{24} & r_{24} & \cdots & r_{41} \\ \vdots & \vdots & \ddots & \vdots \\ r_{2n} & r_{2n} & \cdots & r_{n1} \end{vmatrix}$$



Llamando:

$$R = \begin{vmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ r_{31} & r_{32} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{vmatrix}$$

como $r_{11} = r_{22} = \cdots = r_{nn} = 1$, resultaría:

$$R = \begin{vmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ r_{31} & r_{32} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{vmatrix}$$

Además, el determinante es simétrico por ser $r_{ij} = r_{ji}$.

Si designamos por R_{ij} el adjunto de $r_{i,j}$, resulta:

$$\begin{aligned} \alpha_2 &= -\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}} \\ \alpha_3 &= -\frac{s_1}{s_3} \cdot \frac{R_{13}}{R_{11}} \\ \alpha_4 &= -\frac{s_1}{s_4} \cdot \frac{R_{14}}{R_{11}} \\ &\vdots \\ \alpha_n &= -\frac{s_1}{s_n} \cdot \frac{R_{1n}}{R_{11}} \end{aligned}$$



con lo que

$$\begin{aligned} z_1 &= \alpha_1 + \alpha_2 z_2 + \alpha_3 z_3 + \dots + \alpha_n z_n = \\ &= -\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}} z_2 - \frac{s_1}{s_3} \cdot \frac{R_{13}}{R_{11}} z_3 - \dots - \frac{s_1}{s_n} \cdot \frac{R_{1n}}{R_{11}} z_n \end{aligned}$$

Deshaciendo el cambio 4.1 se obtiene:

$$\begin{aligned} \frac{x_1 - \bar{x}_1}{s_{x_1}} + 4 &= \\ &= -\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}} \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} + 4 \right) - \frac{s_1}{s_3} \cdot \frac{R_{13}}{R_{11}} \left(\frac{x_3 - \bar{x}_3}{s_{x_3}} + 4 \right) - \\ &\quad - \dots - \frac{s_1}{s_n} \cdot \frac{R_{1n}}{R_{11}} \left(\frac{x_n - \bar{x}_n}{s_{x_n}} + 4 \right) \end{aligned}$$

y despejando x_1 :

$$\begin{aligned} x_1 &= \\ &= \left[-\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}} \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} + 4 \right) - \frac{s_1}{s_3} \cdot \frac{R_{13}}{R_{11}} \left(\frac{x_3 - \bar{x}_3}{s_{x_3}} + 4 \right) - \right. \\ &\quad \left. - \dots - \frac{s_1}{s_n} \cdot \frac{R_{1n}}{R_{11}} \left(\frac{x_n - \bar{x}_n}{s_{x_n}} + 4 \right) - 4 \right] s_{x_1} + \bar{x}_1 = \\ &= -s_{x_1} \frac{s_1}{R_{11}} \left[\frac{R_{12}}{s_2} \left(\frac{x_2 - \bar{x}_2}{s_{x_2}} \right) + \frac{R_{13}}{s_3} \left(\frac{x_3 - \bar{x}_3}{s_{x_3}} \right) + \dots \right. \\ &\quad \left. \dots + \frac{R_{1n}}{s_n} \left(\frac{x_n - \bar{x}_n}{s_{x_n}} \right) + 4 \left(\frac{R_{12}}{s_2} + \frac{R_{13}}{s_3} + \dots + \frac{R_{1n}}{s_n} + \frac{R_{11}}{s_1} \right) \right] + \bar{x}_1 \end{aligned}$$



Si nombramos a los coeficientes de la siguiente manera:

$$\beta_1 = -4s_{x_1} \frac{s_1}{R_{11}} \left(\frac{R_{12}}{s_2} + \frac{R_{13}}{s_3} + \dots + \frac{R_{1n}}{s_n} + \frac{R_{11}}{s_1} \right) + \bar{x}_1$$

$$\beta_2 = -s_{x_1} \frac{s_1}{R_{11}} \cdot \frac{R_{12}}{s_2}$$

$$\beta_3 = -s_{x_1} \frac{s_1}{R_{11}} \cdot \frac{R_{13}}{s_3}$$

⋮

$$\beta_n = -s_{x_1} \frac{s_1}{R_{11}} \cdot \frac{R_{1n}}{s_n},$$

tendríamos una expresión de los coeficientes de regresión (coeficientes de las variables independientes y coeficiente independiente) donde estaríamos buscando una ecuación del tipo:

$$x_1 = \beta_1 + \beta_2 \frac{x_2 - \bar{x}_2}{s_{x_2}} + \beta_3 \frac{x_3 - \bar{x}_3}{s_{x_3}} + \dots + \beta_n \frac{x_n - \bar{x}_n}{s_{x_n}}$$

que es la ecuación del plano de regresión de x_1 sobre (x_2, x_3, \dots, x_n) .

4.2 Medidas de la dependencia estadística y de la correlación

Se hace ahora necesario complementar el análisis de la regresión con la obtención de unas medidas o coeficientes que permitan calibrar el



grado de dependencia estadística existente entre las variables (Nodos) o, dicho de otro modo, el *grado de representatividad* de la función analítica ajustada a los datos obtenidos empíricamente por observación.

Se definirá y comentará la *varianza residual* y el *coeficiente de determinación*, como medidas de la dependencia estadística existente entre dos variables.

4.2.1 La varianza residual

Una vez ajustada la ecuación de regresión y^* a una determinada nube de puntos, se producirán unas diferencias entre los distintos *valores ajustados o teóricos*, y_i^* , y los correspondientes *valores observados o empíricos*, y_i .

Así el *error* o *residuo* que se produce al sustituir la nube de puntos por la ecuación ajustada puede escribirse para cada punto en la forma:

$$y_i = y_i^* + e_i$$

La *varianza residual* se define como la varianza de la serie de errores o residuos, es decir, la varianza de la serie de e_i o de la serie de diferencias $y_i^* - y_i$. Llamando S_e^2 a dicha varianza residual, su expresión será:

$$S_e^2 = \frac{\sum_{i=1}^m (e_i - \bar{e})^2}{m - NI - 1} \quad (4.3)$$

donde NI se corresponde con el número de variables independientes de las que se dispone en el fichero de datos.



Por el método utilizado, el de mínimos cuadrados, se cumple que la media de los residuos es nula:

$$\bar{e} = \frac{\sum_{i=1}^m e_i}{m} = 0$$

de modo que 4.3 se reduce a:

$$S_e^2 = \frac{\sum_{i=1}^m (e_i - \bar{e})^2}{m - NI - 1} = \frac{\sum_{i=1}^m e_i^2}{m - NI - 1} = \frac{\sum_{i=1}^m (y_i - y_i^*)^2}{m - NI - 1}$$

Parece lógico utilizar una medida de dispersión, para juzgar el grado de representatividad del ajuste efectuado, que permita calibrar la capacidad de sustitución correcta que pueda tener la ecuación de regresión ajustada al sintetizar la nube de puntos observada.

Esta medida de dispersión es precisamente la varianza residual, que mide la dispersión existente entre las ordenadas de los valores observados, y_i , y las ordenadas de sus valores medidos, y_i^* , de forma que, si esta varianza residual, S_e^2 , es grande, la representatividad de la ecuación de regresión será pequeña y si S_e^2 es pequeña, la representatividad será grande. En el caso de que $S_e^2 = 0$, nos encontraremos ante una dependencia perfecta entre las variables, puesto que ser nulo S_e^2 implica a su vez que sea nulo cada sumando, esto es, $e_i^2 = 0$ para todo i , lo cual indica que no existirán residuos o diferencias entre los valores observados, y_i , y los valores medidos, y_i^* .



4.2.2 Coeficiente de determinación ajustado o corregido

Se puede expresar el coeficiente de determinación, R^2 , en función de la varianza de la variable dependiente, es decir, de la varianza total, y de la varianza residual:

$$R^2 = 1 - \frac{S_e^2}{S_y^2} \quad (4.4)$$

Se puede afirmar que el cociente $\frac{S_e^2}{S_y^2}$ que aparece en el segundo miembro del coeficiente de determinación 4.4, será en todo caso positivo, pues se trata de un cociente entre números positivos. Por otra parte si nos basamos en la relación:

$$S_y^2 = S_{y^*}^2 + S_e^2 \quad (4.5)$$

podrá tomar como máximo el valor 1, ya que, como se desprende de 4.5 será $S_y^2 \geq S_e^2$.

Así, el coeficiente de determinación, siempre es positivo y menor o igual que la unidad.

Este coeficiente presenta la ventaja de ser una medida de tipo abstracto y, en consecuencia, de su aplicación práctica pueden obtenerse conclusiones de orden cuantitativo, que permiten, además, comparaciones con otras distribuciones bidimensionales. Así, tal como se desprende del análisis de las fórmulas (4.4) y (4.5), si $R^2 = 1$, la dependencia o correlación entre las variables será perfecta o exactamente representada por la ecuación ajustada; si por el contrario, $R^2 = 0$, la ecuación ajustada no representa en absoluto la posible relación de dependencia entre las variables en estudio.



Por otra parte, no ha de perderse de vista que un valor bajo o incluso nulo del coeficiente de determinación, no indica más que lo inadecuado de la función analítica elegida como representativa de la relación entre las variables y cabe, por tanto, que la adecuada modificación de la función de ajuste conduzca a la elevación del correspondiente coeficiente de determinación, sin que, pese a ello, pueda asegurarse, por razones meramente estadísticas, la existencia de una dependencia causal.

4.2.3 Coeficiente de determinación normal

Se utiliza la expresión:

$$R^2 = 1 - \frac{\sum_{i=1}^m e_i^2}{(n-1) \sum_{i=1}^m (y_i - \bar{y})}$$



Universitat d'Alacant
Universidad de Alicante

Capítulo 5

Software

La metodología presentada en este proyecto tiene como primer objetivo la construcción de familias de ecuaciones matemáticas, generadas a partir de un lenguaje. De las familias construidas, lo principal es determinar un conjunto de ecuaciones concretas que se ajusten en mayor o menor medida a los datos experimentales, objeto de la modelización matemática. Dichos planteamientos nos conducen a la necesidad de memoria para el almacenamiento y el cálculo operacional.

Analizando el objetivo general se pensó en utilizar un lenguaje de programación que incluyera librerías matemáticas y que facilitara la implementación de ciertos cálculos y operaciones estadísticas. El problema surge debido a la necesidad de gran cantidad de memoria para almacenar y para operar. Por ello, se ha elegido el **lenguaje C** como lenguaje de programación, pues, aunque la mayoría de las funciones matemáticas no están implementadas (lo que supone una mayor di-



ficultad en la programación) si que es cierto que permite una mejor gestión de la capacidad de memoria, lo que resulta muy importante en la implementación de nuestra metodología. Otro motivo de elección de este lenguaje es que es de dominio público y permite su utilización tanto en un PC como en una estación de trabajo, indispensable en ciertos casos debido a la gran cantidad de cálculos realizados y de memoria necesitada.

Para la obtención de familias de ecuaciones que se ajusten a los datos experimentales se ha escogido el ajuste por mínimos cuadrados, de la misma forma que se aplica en los métodos estadísticos de regresión. Es además posible seleccionar una o un conjunto de ecuaciones de acuerdo con los criterios establecidos previamente.

La metodología presentada aporta respecto a las usuales ([13, 27]) la particularidad de generar además familias de ecuaciones no lineales. Es además posible considerar como variables independientes las variables construidas como fruto de la aplicación de ciertas funciones a las variables inicialmente independientes. Además, se permite elegir entre una serie de funciones disponibles en un menú de entrada.

Como se observa en la tabla 5.1, todas las funciones disponen de un coeficiente A , que también será cuestionado, de modo que se puede trabajar con x^2 , x^3 , $\exp(0.1x)$, $\sqrt[4]{x}$, etc. Esto permitirá, en algunas ocasiones, considerar funciones que mejor representen los fenómenos y, por consiguiente, poder aproximarnos más a los datos.

Uno de los aspectos fundamentales para la ejecución del software es la elección del vocabulario de orden uno, ya que a partir de él se genera el lenguaje con el que se modelizan las variables del sistema (HSS).

Resulta importante la elección de estas funciones por lo que siem-



Funciones a elegir	
1.	x^A
2.	$x^{1/A}$
3.	$\frac{1}{Ax}$
4.	$\text{sen}(Ax)$
5.	$\exp Ax$
6.	$\frac{1}{\exp Ax}$
7.	$\text{arctg}(x)$
8.	$\ln Ax $

Tabla 5.1: Funciones a elegir

pre será recomendable realizar cualquier estudio previo de las características del sistema y de los datos experimentales. Estos análisis deben de realizarse con anterioridad a la elección de los vocabularios de orden 1, y aunque son independientes del software aquí presentado, pueden aportar una información adicional que beneficiará al modelizador en la obtención de ecuaciones que mejor representen los fenómenos.

A toda variable independiente se le puede aplicar más de una función, por ejemplo: $\text{sen}(x^2)$. Como se observa, realmente correspondería a la composición de dos funciones sen y x^2 . Para ello, se pregunta por el orden de composición de las funciones, o lo que en el programa se ha llamado: orden de las transformadas (*ord*). De modo que si se elige $\text{ord} = 3$ entonces se trabajará con funciones de hasta orden 3 incluido.



5.1 Estructuras de datos

Para la construcción del programa ha sido indispensable la utilización de memoria dinámica debido a la gran complejidad de los datos y el gran número de ellos que se pueden almacenar y crear. El programa se estructura en 5 grandes grupos de funciones, cada uno de los cuales engloba un determinado fin. Del mismo modo se han definido **tres estructuras principales de datos**, la primera en forma de árbol, la segunda de tipo lista con doble puntero y la tercera en forma de matriz.

La primera estructura contiene nodos llamados *Npila* en los cuales se almacenan:

1. Las funciones elegidas y posteriormente su composición y combinaciones lineales (en tipo cadena),
2. su longitud (debido a la concatenación y composición de las mismas esta longitud irá aumentando),
3. el nivel (profundidad del árbol donde se encuentra el nodo) y
4. el tipo (si se ha producido la concatenación).

Como se ha dicho anteriormente esta estructura tendrá forma de árbol utilizando memoria dinámica para ello. Si generalizamos para n funciones elegidas con un orden de transformadas de k se puede observar una representación gráfica de cómo almacenaría internamente la composición de las funciones.



x^2	$(\text{sen}(x^2))^2$	$\text{sen}(\text{sen}(\exp(0.1x)))$
$\text{sen}x$	$((\exp(0.1x))^2)^2$	$\text{sen}(\exp(0.1x^2))$
$\exp(0.1x)$	$(\text{sen}x^2)^2$	$\text{sen}(\exp(\text{sen}x))$
$(x^2)^2$	$(\text{sen}(\text{sen}x))^2$	$\text{sen}(\exp(0.1(\exp(0.1x))))$
$(\text{sen}x)^2$	$(\text{sen}(\exp(0.1x)))^2$	$\exp(0.1(x^2)^2)$
$(\exp(0.1x))^2$	$(\exp(0.1x^2))^2$	$\exp(0.1(\text{sen}x)^2)$
$\text{sen}(x^2)$	$(\exp(0.1\text{sen}x))^2$	$\exp(0.1(\exp(0.1x))^2)$
$\text{sen}(\text{sen}x)$	$(\exp(0.1(\exp(0.1x))))^2$	$\exp(0.1\text{sen}x^2)$
$\text{sen}(\exp(0.1x))$	$\text{sen}((x^2)^2)$	$\exp(0.1\text{sen}(\text{sen}x))$
$\exp(0.1x^2)$	$\text{sen}(\text{sen}x^2)$	$\exp(0.1\text{sen}(\exp(0.1x)))$
$\exp(0.1\text{sen}x)$	$\text{sen}(\exp(0.1x))^2$	$\exp(0.1\exp(0.1x^2))$
$\exp(0.1\exp(0.1x))$	$\text{sen}(\text{sen}x^2)$	$\exp(0.1\exp(0.1\text{sen}x))$
$((x^2)^2)^2$	$\text{sen}(\text{sen}(\text{sen}x))$	$\exp(0.1\exp(0.1(\exp(0.1x))))$

Tabla 5.2: Composición de funciones.

Por ejemplo, en el caso de elegir las funciones 1, 4 y 5, siendo los coeficientes respectivamente 2, 1 y 0.1 y un orden de transformadas de 3, las funciones generadas se pueden observar en la tabla 5.2.

Esta tabla cabe interpretarla, es decir, para saber el orden de generación de las funciones hay que leerla de arriba a abajo y de derecha a izquierda.

En la **segunda estructura** los nodos se denominan *Nlista*. Esta lista se utiliza para almacenar los valores del fichero que contiene los datos de las variables. Así mismo se procede al tratamiento de los mismos: tipificarlos, trasladarlos y aplicarles las funciones correspondientes, de modo que se aplica el método de los mínimos cuadrados para obtener las ecuaciones que se aproximan en mayor o menor medida a los datos experimentales. Para todo ello, se necesitan varios campos:



1. El valor original del fichero de datos,
2. el valor correspondiente de la aplicación de las funciones y
3. un campo auxiliar para almacenar ciertos cálculos o para realizar operaciones intermedias (por ejemplo para hallar la media, para almacenar la varianza, etc.).

Una representación de esta segunda estructura se puede observar en la tabla 5.3.

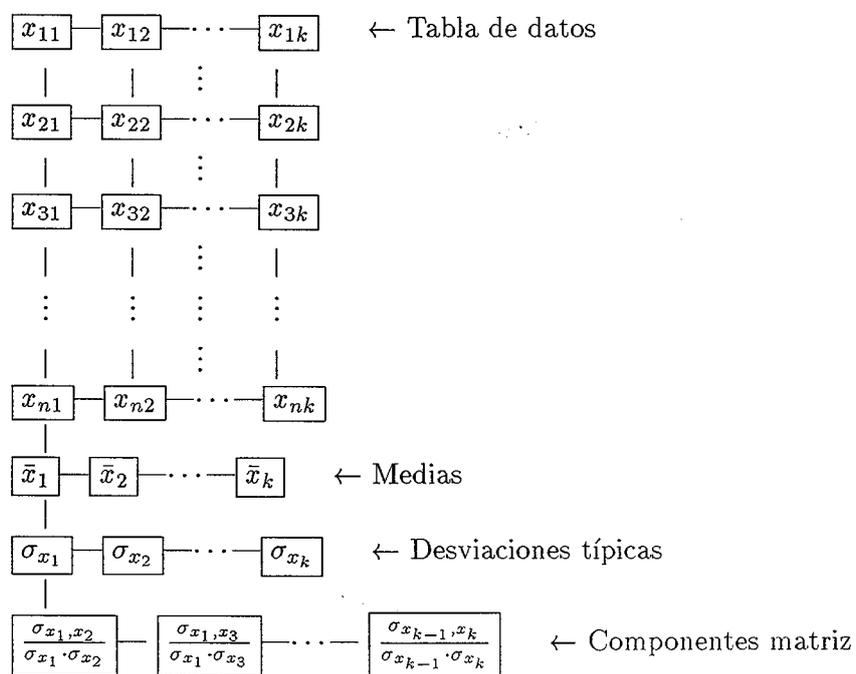


Tabla 5.3: Estructura de datos



En la **tercera estructura** se dispone de una matriz que es utilizada para construir, almacenar y manejar los datos necesarios en el test de significancia, pues se requiere construir una matriz de datos y trabajar con su inversa, lo que resulta imprescindible para calcular la función t-Student. Para la construcción de esta matriz son necesarios algunos cálculos que nos permiten calcular sus componentes. En la matriz 5.1 se puede ver qué expresión posee cada componente:

$$\begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{ki} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \cdots & \sum_{i=1}^n x_{ki}^2 \end{bmatrix} \quad (5.1)$$

5.2 Bloques de funciones

Se pueden considerar 5 grandes bloques de funciones:

1. Generación, cálculo y combinación de funciones.
2. Lectura y tratamiento estadístico del fichero de datos.
3. Funciones sobre la matriz de correlaciones.
4. Construcción de la matriz que se utiliza para poder pasar el test de significancia.



5. Cálculo de la función t-Student.

A continuación se comentarán las funciones incluidas en cada uno de los bloques:

1. **Generación, cálculo y combinación de funciones.** Las funciones incluidas en este grupo llevan el distintivo *F*. Estos procesos actúan como un traductor de las funciones debido a que, internamente, son tomadas como cadenas. Elegidas las funciones, se generan las distintas composiciones y combinaciones lineales, dependiendo del número de funciones elegidas, el número de variables independientes y el orden de las transformadas.

El número de ecuaciones que se pueden obtener aumenta respecto al orden de transformadas y el número de variables independientes, de la misma forma que la potencia (entero positiva) de una serie geométrica de razón ≥ 1 . Por este motivo se necesita gran cantidad de memoria.

Proposición

Sean $\mathcal{F} = \{f_1, f_2, \dots, f_{func}\}$ el conjunto de transformadas elementales. Sean $\mathcal{X} = \{x_1, x_2, \dots, x_{VI}\}$ el conjunto de variables independientes consideradas en la modelización. Se desea conocer el número de ecuaciones hasta orden *ord* que resultarían para $\{1, 2, \dots, VI\}$ variables independientes.

El número máximo de ecuaciones (NME) que se pueden generar con *func* funciones elementales de hasta orden *ord* y con *VI* variables independientes es la *VI* potencia de la suma de *ord* términos de la serie geométrica de razón *func*, lo que se puede



además expresar como:

$$\text{NME} = \left(\sum_{n=1}^{\text{ord}} \text{func}^n \right)^{\text{VI}} \quad (5.2)$$

Demostración

- (a) Para $\text{VI} = 1$ el número máximo de ecuaciones que se pueden generar vendría expresado por:

$$\text{NME} = \sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \quad (5.3)$$

Demostremos dicha ecuación 5.3 por inducción sobre el orden máximo de de las transformadas elementales:

- Para **ord=1**:
Aplicaríamos *func* funciones elementales sobre la variable independiente de la que disponemos, luego:

$$\text{NME} = \text{func} = \text{VR}_{\text{func}}^1$$

- Para **ord=2**:
Podríamos aplicar *func* funciones elementales de orden 1 más aquellas que se generan de orden 2 a la variable independiente:

$$\text{NME} = \text{VR}_{\text{func}}^1 + \text{VR}_{\text{func}}^1 = \sum_{n=1}^2 \text{VR}_{\text{func}}^n$$



- Supongamos que se cumple para $\text{ord}=k$:

$$\text{NME} = \sum_{n=1}^k \text{VR}_{\text{func}}^n \quad (5.4)$$

- Veamos entonces que se cumple para $\text{ord}=k+1$:

$$\begin{aligned} \text{NME} &= \sum_{n=1}^k \text{VR}_{\text{func}}^n + \text{VR}_{\text{func}}^{k+1} = \\ &= \text{VR}_{\text{func}}^1 + \text{VR}_{\text{func}}^2 + \text{VR}_{\text{func}}^3 + \dots + \text{VR}_{\text{func}}^k + \text{VR}_{\text{func}}^{k+1} = \\ &= \sum_{n=1}^{k+1} \text{VR}_{\text{func}}^n \quad \text{c.q.d.} \end{aligned}$$

- (b) Vamos a generaliza para VI variables independientes. En este caso, para $func$ funciones elementales de hasta ord orden, el número máximo de ecuaciones que se pueden generar será:

$$\text{NME} = \left(\sum_{n=1}^{\text{ord}} \text{func}^n \right)^{VI} \quad (5.5)$$

Pasemos a demostrar la ecuación 5.5 por inducción sobre el número de variables independientes:

- Para $VI=1$:
El número máximo de ecuaciones utilizando $func$ funciones transformadas elementales de hasta orden ord viene dado por la ecuación 5.6:

$$\text{NME} = \sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \quad (5.6)$$



que como se ve es equivalente a la ecuación 5.4, demostrada anteriormente.

- Para $VI=2$:

Cada ecuación obtenida con una sola variable independiente, se puede combinar con todas las obtenidas con la segunda variable independiente. En total:

$$\begin{aligned} NME &= \sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \cdot \sum_{n=1}^{\text{ord}} VR_{\text{func}}^n = \\ &= \left(\sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \right)^2 = \\ &= VR^2 \left(\sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \right) \end{aligned}$$

- Supongamos que se cumple para $VI=k$:

$$NME = VR^k \left(\sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \right)$$

- Veamos entonces que se cumple para $VI=k+1$:
Para k variables independientes se obtiene:

$$NME = VR^k \left(\sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \right) = \left(\sum_{n=1}^{\text{ord}} VR_{\text{func}}^n \right)^k$$



Al añadir una nueva variable independiente, cada una de las ecuaciones obtenidas para las k variables anteriores, se puede combinar con $\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n$ ecuaciones. Por tanto el número máximo de ecuaciones será:

$$\begin{aligned} \text{NME} &= \left(\text{VR}^k \left(\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \right) \right) \cdot \left(\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \right) = \\ &= \left(\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \right)^k \cdot \left(\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \right) = \\ &= \left(\sum_{n=1}^{\text{ord}} \text{VR}_{\text{func}}^n \right)^{k+1} \quad \text{c.q.d} \end{aligned}$$

En la tabla 5.4 se puede observar un ejemplo de las ecuaciones que se generarían como máximo en el supuesto de que el fichero de datos tuviera 3 variables independientes y se eligieran 1, 2, 3 ó 4 funciones elementales:

Las ecuaciones generadas, con coeficientes a determinar, tendrán tantos sumandos como de variables independientes se disponga en el fichero de datos. Cada uno de estos sumandos, formados por una función o las correspondientes composiciones de funciones, son aplicados sobre los datos numéricos obteniéndose unas nuevas tablas de datos.

Las siguientes funciones son las que se incluyen en este grupo:



Tabla 5.4: Número máximo de ecuaciones

func	ord	vi	NME
1	1	3	1
1	2	3	8
1	3	3	27
2	1	3	8
2	2	3	216
2	3	3	2744
3	1	3	27
3	2	3	1728
3	3	3	59319
4	1	3	64
4	2	3	8000
4	3	3	592704

- **Apilar_F:** Apila la función elegida en el árbol en el que se expresan los elementos del lenguaje con el que se construyen las ecuaciones matemáticas.

Parámetros de entrada: Código de la función elegida, longitud de la función y nivel del árbol donde se apilará, que en nuestro caso será 1.

Parámetros de salida: Ninguno.

- **BajarLinea_F:** Lee una línea más del fichero que contiene las funciones elegidas. A su vez, va escribiendo en el fichero "result.txt" las funciones en concreto.



Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **Calcular_F:** Calcula numéricamente $f(x)$, donde f es la función en la que se está trabajando y x es el dato numérico al cual se le aplica la función f . Debido a que f puede ser resultado de una composición de funciones se ha hecho indispensable hacer esta función recursiva.

Parámetros de entrada: Código de las composiciones de las funciones a calcular, longitud de dicho código y número sobre el cual queremos calcular esa función.

Parámetros de salida: Resultado de aplicar la función sobre el número.

- **CombinarFunciones_F:** Con esta función se procede a realizar la combinación lineal de las funciones hasta completar el número de variables independientes que hay en el fichero de datos. De este modo quedarán casi implementadas las ecuaciones que se aplicarán a los datos para poder determinar los coeficientes respectivos. Por ejemplo, si hubiera 3 variables independientes se irían sumando las funciones obtenidas en la composición hasta obtener 3 términos o sumandos necesitando posteriormente la determinación de los coeficientes respectivos. Estas combinaciones se van añadiendo al árbol de de las funciones elementales y sus composiciones.

Hay que destacar que como en la ecuación resultante se tienen que utilizar todas las variables independientes, está claro que cuando haya obtenido la combinación de 3 funciones a partir de las de 2, podré eliminar estas últimas liberando con ello memoria.



Parámetros de entrada: Número de variables independientes.

Parámetros de salida: Ninguno.

- **Desapilar_F:** Esta función permite eliminar todo el árbol de funciones elementales, sus composiciones y combinaciones, de manera que se libera toda la memoria utilizada para este propósito.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **DesapilarHorizontal_F:** Elimina una fila del árbol de funciones, debido a que ya no va a ser utilizada y con ello se ayudará a disponer de más memoria.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **EscribirFunciones_F:** Según el parámetro que se le pase, almacena las funciones elegidas elementales en el fichero de salida "gen.txt" o las combinaciones lineales en el fichero "comblin.txt" con coeficientes indeterminados.

Parámetros de entrada: Valor para indicar si se almacenan las funciones en "gen.txt" o "comblin.txt" y el número de variables independientes.

Parámetros de salida: Ninguno.

- **EstaEnLista_F:** Función que se utiliza para comprobar si una función elemental ha sido ya elegida o no, pues no se permite repetir funciones elementales.

Parámetros de entrada: La lista de funciones que ya han sido elegidas y la función elegida en ese momento.



Parámetros de salida: 1 ó 0 dependiendo de si ha sido elegida o no.

- **EsVacía_F:** Comprueba si el árbol de funciones elementales está o no vacío.

Parámetros de entrada: Ninguno.

Parámetros de salida: 1 ó 0 dependiendo de si el árbol está vacío o no.

- **Fin_F:** Comprueba si se ha llegado ya al final del fichero que almacena los códigos de las funciones elementales elegidas.

Parámetros de entrada: Ninguno.

Parámetros de salida: 1 ó 0 dependiendo de si se ha llegado al final del fichero o no.

- **GenerarCadena_F:** Las funciones internamente son almacenadas mediante códigos numéricos, pues ocupan menos espacio en memoria, de manera que en la salida se necesita traducir estos códigos a las cadenas de las funciones correspondientes. Se podría decir que tiene la función de un traductor, pues tiene que añadir los paréntesis necesarios para enlazar las funciones, así como los coeficientes elegidos en las funciones elementales. Las traducciones correspondientes se van almacenando en los ficheros "gen.txt" o "comblin.txt" dependiendo de desde donde es llamada la función.

Parámetros de entrada: Código de la función a traducir, longitud de ese código, indicativo para ver si se imprime en el fichero "comblin.txt" o "gen.txt", número de la variable independiente a la que se le aplica la función, indicativo de si se ha llegado al final de la combinación lineal de funciones,



número de variables independientes y un indicativo de si el código corresponde o no a una combinación lineal.

Parámetros de salida: Ninguno.

- **GenerarFunciones_F:** Genera las composiciones de las funciones elegidas y combinaciones generadas dependiendo de la profundidad del árbol que en nuestro caso es el orden de las transformadas.

Parámetros de entrada: La profundidad del árbol generada.

Parámetros de salida: Ninguno.

- **GenerarSubcadena_F:** Las ecuaciones internamente podemos decir que se componen de términos, los cuales se distinguen por estar separados por un signo + y estar formados a su vez de secuencias de números indicando las funciones con las que se realiza la composición. Estos términos son pasados uno a uno a esta función para hacer la traducción correspondiente de las funciones e ir construyendo las ecuaciones y visualizarlas en el fichero "result.txt" con sus coeficientes respectivos.

Parámetros de entrada: Códigos de la cadena de funciones a traducir, longitud de esa cadena, número de la variable independiente a la que se le está aplicando la función, indicativo de si es final de cadena, número total de variables independientes, indicativo de si la cadena es fruto de alguna combinación lineal y el número total de variables.

Parámetros de salida: Ninguno.

- **LeerFichero_F:** Lee el fichero donde están contenidas las funciones elementales elegidas, es decir, sus códigos respectivos.



Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

Como ejemplo pondremos la función a partir de la cual se generan las composiciones de las funciones elegidas y combinaciones dependiendo de la profundidad del árbol (**GenerarFunciones-F**) y la función que se utiliza para almacenar las funciones elegidas en el fichero de salida "gen.txt" o las posibles combinaciones lineales de ellas en el fichero de salida "comblin.txt", dependiendo del código que se le pase ,0 ó 1 respectivamente (**EscribirFunciones-F**):

GenerarFunciones-F

```
h1=h2= primer nodo
```

```
pr= 1
```

```
mientras pr<profundidad hacer
  mientras h1<>NULL hacer
    crear nodo
    copiar h1 a nodo
    concatenar nodo y h2
    guardar suma longitudes
    aumentar nivel
    h2= siguiente
  mientras h2<>NULL
    crear nodo
    copiar h1 a nodo
    concatenar nodo y h2
    guardar suma longitudes
```



```

        h2=siguiente
    fin mientras
    h1=siguiente
    h2=siguiente nivel
fin mientras
pr=pr+1
si pr<profundidad hacer
    h1=primer nodo
    h2=siguiente nivel
fin si
fin mientras

```

EscribirFunciones-F

```

cont=i=j=k=final=n=0
comp=cadena de caracteres vacia

h1=h2= primer nodo
si codigo=1 entonces
    mientras h2<>NULL hacer
        mientras h1<>NULL hacer
            GenerarCadena-F(funcion, longitud, codigo,
                cont, final, VI, tipo de la funcion, h1);
            h1=siguiente
        fin mientras
        h2=abajo;
        h1=h2;
    fin mientras
fin si

```



```

sino
  si VI>1 y hay alguna combinacion lineal hacer
    h1=h2=primera combinacion lineal
  fin si
sino
  si VI=1 hacer
    h1=h2=primer nodo
  fin si
fin sino
mientras h2<>NULL hacer
  mientras h1<>NULL hacer
    largo=longitud de la combinacion de las
    funciones
    repetir
      Copiar caracter i en comp
      i=i+1
      k=k+1
      si final de combinacion hacer
        final=1
        GenerarCadena-F(comp)
        i=cont=0
      fin si
      si caracter=+ hacer
        j=k
        k=k+1
        GenerarCadena-F(comp)
        i=0
        cont=cont+1

```



```
        fin si
    mientras k<>largo
        h1=siguiente
        final=cont=i=j=k=0
    fin mientras
    h2=abajo
    h1=h2
    cont=0
    fin mientras
fin sino
```

2. **Lectura y tratamiento estadístico del fichero de datos.** Las funciones se distinguen por finalizar en `_D`. Dentro de este grupo se encuentran las funciones que tratan y operan estadísticamente sobre los datos numéricos. Así, podemos citar: lectura y almacenamiento de los datos, cálculo de la media, la varianza, la covarianza, tipificación de los datos, coeficiente de determinación, varianza residual, test de Kolgomorov-Smirnov (KS), etc.

Funciones incluidas en este grupo:

- **AplicarFuncionesADatos_D:** Se aplican las funciones elegidas y sus posibles composiciones a las tablas de datos correspondientes generándose por tanto unas nuevas tablas de datos que serán con las que se trabajarán para aplicar el método de los mínimos cuadrados.

Parámetros de entrada: La cadena con los códigos de las combinaciones y composiciones de las funciones y el número total de tuplas de datos.

Parámetros de salida: Ninguno.



- **BorrarDatos_D**: De la lista de datos que está compuesta por los nodos *Nlista* borra los campos que se utilizan para hacer cálculos, de manera que sólo deja el valor original del fichero leído. Pero sí que elimina los nodos que almacenan las medias, varianzas y varianzas-covarianzas. Es necesario realizar esta función cuando el número de ecuaciones posibles que se puede generar es mayor que 1. Hay que realizar este paso previo cuando se pasa a probar con cada una de las posibles ecuaciones, salvo con la primera porque esta lista todavía no ha sido generada.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **CoefficienteDeterminacion_D**: Calcula el coeficiente de determinación, tanto el coeficiente de determinación ajustado o corregido, como el normal, devolviendo, en su caso, el mayor.

Parámetros de entrada: Número de tuplas de datos y la varianza residual.

Parámetros de salida: El coeficiente de determinación.

- **Covarianzas_D**: Calcula los términos de la matriz de varianzas y covarianzas que se utiliza en el método de los mínimos cuadrados. Por ejemplo, el elemento [1,3] de la matriz se corresponde con la covarianza entre la primera variable y la tercera dividida entre el producto de sus desviaciones típicas. Las variables son numeradas de manera que la dependiente es la número 1 y las independientes son numeradas correlativamente a ésta, es decir, la primera variable independiente será la número 2, la segunda será la número 3, y



así sucesivamente. De manera que en este caso se correspondería a la relación entre la variable dependiente y la segunda variable independiente.

Parámetros de entrada: La cantidad de tuplas de datos.

Parámetros de salida: Ninguno.

- **Desapilar_D:** Elimina todos los nodos de la lista que contiene los datos del fichero, sus medias, varianzas y varianzas-covarianzas. De modo que la lista queda totalmente vacía.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **Erfcc_D:** Es la complementaria de la función de error. Su definición es:

$$\operatorname{erfcc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$$

Esta función tiene los siguientes valores extremos y simetrías:

$$\operatorname{erfcc}(0) = 1 \quad \operatorname{erfcc}(\infty) = 0 \quad \operatorname{erfcc}(-x) = 2 - \operatorname{erfcc}(x)$$

Está relacionada con la función gamma incompleta por medio de:

$$\operatorname{erfcc}(x) = Q\left(\frac{1}{2}, x^2\right) \quad (x \geq 0)$$

Parámetros de entrada: Valor al que se le quiere aplicar esta complementaria de la función de error.

Parámetros de salida: El resultado de la aplicación de la complementaria de la función error al parámetro de entrada.



- **Fidis_D:** Función de distribución de la normal asociada al valor que se le pasa como parámetro, que en nuestro caso es el valor del residuo tipificado.

Parámetros de entrada: Valor al que se le quiere aplicar la función de distribución de la normal.

Parámetros de salida: Resultado de aplicar la función de distribución de la normal al parámetro de entrada.

- **Fmax_D:** Devuelve el máximo de entre 2 valores, en caso de ser iguales devuelve el primero.

Parámetros de entrada: Dos valores de los cuales se quiere calcular el máximo.

Parámetros de salida: El máximo de los valores introducidos.

- **Ksone_D:** Dado un array de datos, ordenado ascendentemente, y dada una función suministrada por el usuario de una sola variable, que es una función de distribución acumulativa cuyos rangos son 0 (para valores muy pequeños de sus argumentos) y 1 (para valores muy grandes de sus argumentos), esta rutina devuelve el estadístico KS y el nivel de significancia. Valores pequeños del nivel de significancia indica que la función de distribución acumulativa de los datos es significativamente diferente de la función suministrada por el usuario.

En nuestro caso se compara si los residuos siguen una distribución normal.

Parámetros de entrada: Array que contiene los residuos ordenados, el número total de tuplas de datos, el estadístico KS (indefinido) y el nivel de significancia (indefinido).



Parámetros de salida: El estadístico KS y el nivel de significancia.

- **LeerFichero_D:** Va leyendo los datos numéricos almacenados en el fichero de datos guardándolos en la lista formada por nodos *Nlista*.

Parámetros de entrada: Número total de variables de las que dispone el fichero (dependiente + independientes)

Parámetros de salida: Número de tuplas de datos que contiene el fichero.

- **Medias_D:** Calcula las medias de cada variable a partir de la lista construida con los datos. Estas medias son almacenadas en una fila de nodos a continuación de la última fila de datos.

Parámetros de entrada: Número de tuplas de datos.

Parámetros de salida: Ninguno.

- **PermutarColumnaFichero_D:** De forma general, en las modelizaciones llevadas a cabo ([5, 41]), se han utilizado ficheros de datos donde la variable dependiente figura en la última columna. Cuando se lee el fichero se almacena en la misma posición, pero por comodidad de acceso se pasa la última columna a la primera, de modo que en la lista de nodos de datos la primera columna corresponde a la variable dependiente y el resto de las columnas a las variables independientes.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **Probks_D:** Devuelve la función de probabilidad de Kolgomorov-Smirnov.



Parámetros de entrada: Valor al que se le quiere aplicar la función de probabilidad de Kolgomorov-Smirnov.

Parámetros de salida: Resultado de aplicar esta función de probabilidad al parámetro de entrada.

- **Residuos_D:** Calcula los residuos, es decir, la diferencia entre el valor real de la variable dependiente y el valor teórico que se obtiene con la ecuación resultante. Estos valores se almacenan en la primera columna de la lista de nodos de datos, en un campo auxiliar, y la media de estos.

Parámetros de entrada: Número de tuplas de datos, número total de variables y la media de los residuos (indefinido).

Parámetros de salida: La media de los residuos.

- **Sort_D:** Ordena los elementos de un array de datos en orden ascendente utilizando el algoritmo de ordenación Quicksort, el algoritmo de ordenación más rápido conocido para un gran número de datos.

Parámetros de entrada: Número de tuplas de datos y el array que contiene los datos a ordenar.

Parámetros de salida: Ninguno.

- **TipificarResiduos_D:** Tipifica los residuos.

Parámetros de entrada: Número de tuplas de datos y la varianza residual.

Parámetros de salida: Ninguno.

- **TipificarYTrasladar_D:** Tipifica y traslada los datos originales del fichero de datos. Los traslada sumando 4 para pasar del intervalo $[-2, 2]$ al $[2, 6]$ de modo que se evita el 0 y con ello posibles divisiones por 0.



Parámetros de entrada: Número de tuplas de datos y el número total de variables.

Parámetros de salida: Ninguno.

- **Varianzas_D:** Realmente esta función calcula las desviaciones típicas de las variables que son almacenadas en una nueva fila de nodos *Nlista* a continuación de la fila que contiene las medias. A su vez calcula la suma de los cuadrados de las variables y lo almacena en un campo auxiliar que contienen los nodos de las medias y que se utilizará en cálculos posteriores.

Parámetros de entrada: Número de tuplas de datos y el número total de variables.

Parámetros de salida: Ninguno.

- **VarianzaResidual_D:** Cálculo de la varianza residual.

Parámetros de entrada: Número de tuplas de datos.

Parámetros de salida: La varianza residual.

A continuación se dan algunas de las funciones del test KS: **Fidis-D**, **Erfcc-D** y **Probks-D**.

Ksone-D

```

en= num_datos
fo=0.0
*d=0.0
mientras j<num_datos hacer
    fn=j/en
    ff>(*Fidis_D)(array_residuos[j]);

```



```

dt=Máximo(|fo-ff|,|fn-ff|)
si (dt>*d) hacer
    *d=dt
fin si
fo=fn
j++
fin mientras
en=raíz(en)
*prob=Probks_D((en+0.12+0.11/en)*(*d))

```

Fidis-D

```
return(1.0-Erfcc_D(z/sqrt(2.0))/2.0);
```

Erfcc_D(x)

```

z= ValorAbsoluto(x);
t= 1.0/(1.0+0.5*z);
ans= 0.0;
ans= ans+t*0.17087277;
ans=t*(-0.82215223+ans);
ans=t*(1.48851587+ans);
ans=t*(-1.13520398+ans);

```



```
ans=t*(0.27886807+ans);
ans=t*(-0.18628806+ans);
ans=t*(0.09678418+ans);
ans=t*(0.37409196+ans);
ans=t*(1.00002368+ans);
ans=t*exp(-z*z-1.26551223+ans);
si x>=0 hacer
    t= ans;
fin si
sino
    t=2.0-ans;
fin sino
return t
```

Probks-D(y)

```
fac= 2.0;
sum= 0.0;
termbf= 0.0;
a2= -2.0*y*y;
j= 1;
mientras j<=100 hacer
    term= fac*exp(a2*j*j);
    sum += term;
    si |term|<=EPS1*termbf o |term|<=EPS2*sum hacer
        return sum;
```



```
    fin si
    fac= -fac;
    termbf= |term|
    j= j+1;
    fin mientras
    return 1.0;
```

3. **Funciones sobre la matriz de correlaciones.** Se distinguen por llevar M . Estas funciones tratan y manejan la matriz de varianzas y covarianzas. Es aquí donde, indirectamente, se aplica el método de los mínimos cuadrados, siendo para ello necesario la utilización de funciones que pertenecen a bloques distintos. De modo que se obtendrán los coeficientes de las ecuaciones correspondientes.

Funciones incluidas en este bloque:

- **ABS_M:** Devuelve el valor absoluto de un número.
Parámetros de entrada: Número del que se quiere calcular su valor absoluto.
Parámetros de salida: Valor absoluto del número pasado como parámetro.
- **BuscoUltimaFDist0_M:** Busca en la matriz de varianzas-covarianzas la siguiente fila que contenga un 0 en esa misma columna.
Parámetros de entrada: Número de la fila y número total de variables.
Parámetros de salida: La siguiente fila de la que se pasa como parámetro que contiene un 0 en la misma columna.



- **Coefficientes M:** Calcula los coeficientes de la ecuación planteada. Para ello sigue unos ciertos pasos:
 - (a) Calcula los coeficientes de regresión R_j (adjuntos correspondientes de los elementos r_{1j} de la matriz de varianzas-covarianzas), lo que ayudará a calcular los coeficientes de la ecuación.
 - (b) Calcula los residuos,
 - (c) la varianza residual,
 - (d) tipifica los residuos,
 - (e) se pasa el test de Kolgomorov-Smirnov,
 - (f) calcula el coeficiente de determinación,
 - (g) se pasa el test de significancia (en el caso de que se haya elegido pasarlo),
 - (h) destipifica la variable dependiente de la ecuación construida y
 - (i) escribe la ecuación resultante en el fichero.

En el caso de que la varianza de la variable dependiente sea 0 se aborta el programa para evitar una división por 0, dando la posibilidad de volver a ejecutar el programa eligiendo otro fichero de datos.

Parámetros de entrada: Número de tuplas de datos, número total de variables, modo de almacenamiento de las funciones, rango del mayor coeficiente de determinación permitido, rango del menor coeficiente de determinación permitido, códigos de las funciones compuestas y combinadas, número de las variables independientes y si es una función elemental o no.

Parámetros de salida: Ninguno.



- **CopiarMatriz_M:** Crea una imagen de la matriz de varianzas-covarianzas para poder hacer operaciones intermedias. Dependiendo del caso hace una copia de la original o de la auxiliar.

Parámetros de entrada: Tipo de copia que se debe de realizar, si es una copia de la matriz original o es una copia de la matriz auxiliar, y el número total de variables.

Parámetros de salida: Ninguno.

- **Diagonalizar_M:** Diagonaliza la matriz para poder hallar su determinante por medio del producto de los elementos de su diagonal.

Parámetros de entrada: Número de variables independientes.

Parámetros de salida: Determinante de la matriz.

- **EscribirEcuacionEnFichero_M:** Escribe en el fichero "result.txt" la/s ecuación/es obtenida/s. Además también escribe en el fichero la varianza residual, el coeficiente de determinación, el valor del test de Kolgomorov-Smirnov y el test de significación para cada uno de los coeficientes, en el caso de haber decidido pasarlo. Hay casos en los que, a pesar de haber decidido pasar el test de significancia, éste no se puede llevar a cabo. Esto sucede cuando el número de tuplas de datos menos el número de variables (dependientes e independientes) menos 1, es 0 (pues ocasiona división por 0), y además cuando se produce una división por 0 en cálculos intermedios (a causa, en ocasiones, de disponer de 2 columnas de datos iguales). En estos últimos casos puede ser que el valor que aparezca por pantalla sea NA, significando que no ha sido posible obtenerlo.



Parámetros de entrada: Número total de variables, varianza residual, coeficiente de determinación, cadena de códigos de las composiciones y combinaciones de funciones, número de variables independientes, si se trata de una función elemental o no, número de tuplas de datos y si se ha producido alguna división por 0 en el cálculo del test de significancia.

Parámetros de salida: Ninguno.

- **InicializarMatriz_M:** Inicializa la matriz de varianzas-covarianzas a 0.

Parámetros de entrada: Ninguno.

Parámetros de salida: Ninguno.

- **PasarAMatriz_M:** Pasa los valores calculados de varianzas dividido por las covarianzas a los elementos correspondientes de la matriz, que resultará ser simétrica. Estos valores se encuentran almacenados en la siguiente fila de nodos de datos a la fila de nodos de desviaciones típicas de las variables en la lista de datos.

Parámetros de entrada: Número total de variables.

Parámetros de salida: Ninguno.

- **PermutarFila_M:** Permutar las filas i y j de la matriz de varianzas-covarianzas.

Parámetros de entrada: Número de las filas que se quieren permutar entre sí.

Parámetros de salida: Ninguno.

- **QuitoFilaColumna_M:** Elimina la fila i y la columna j de la matriz de varianzas-covarianzas, trasladando los datos y rellenando a ceros la última fila y columna vacía resultantes, para obtener la matriz adjunta del elemento i, j .



Parámetros de entrada: El número de la fila y de la columna que se quieren eliminar, así como el número total de variables.

Parámetros de salida: Ninguno.

- **Regresiones_M:** Calcula los coeficientes de regresión, eliminando la fila y columna correspondientes y diagonalizando la matriz resultante.

Parámetros de entrada: Número total de variables.

Parámetros de salida: Ninguno.

Como ejemplo, veremos la función que calcula los coeficientes indeterminados de la ecuación (**Coeficientes-M**) y una función auxiliar que se emplea para calcular el determinante de una matriz mediante el cálculo de ceros por debajo de la diagonal (**Diagonalizar-M**):

Coeficientes-M

```

Cálculo de  $R_i$  /*adjuntos de la fila 1 y
                columna i de la matriz R */
Si la  $varianza_0$  de la VD es 0 hacer
    exit()
fin si
fac= $varianza_0/R_0$ 
i=1
mientras i<numero de variables hacer
    det=  $R_i$ 
     $V_i = -1 * varianza_0/R_0 * R_i/varianza_i$ 
    i= i+1

```



```

fin mientras
i=0
mientras i<numero de variables hacer
    ind= ind + (Ri/varianzai)*mediai
    i= i+1
fin mientras
V0= varianza0 * R0 * ind

```

Cálculo de los residuos.
 Cálculo de la varianza residual.
 Tipificar los residuos.
 Ordenar los residuos (exigencias del ks).
 Guardar los residuos en "residuos.txt".
 Aplicar el KS.
 Almacenar ecuación en el fichero.

Diagonalizar-M

```

f=0
mul=0
mientras f<Num.de var. independientes hacer
    i=0
    si MatrizA[f,f]=0 hacer
        i=Ultima fila distinta de 0
        si existe i hacer
            Permutar las filas i y f
        fin si
    fin si
fin si

```



```

si f<Num.de var. independientes y i existe hacer
  i=f+1
  mientras i<Num.de var. independientes hacer
    di=Matriz[i,f]/Matriz[f,f]
    j=f
    mientras j<Num.de var. independientes hacer
      Matriz[i,j]=Matriz[i,j]-(di*M[f,j])
      j=j+1
    fin mientras
    i=i+1
  fin mientras
fin si
f=f+1
fin mientras

i=0
mientras i<Num.de var. independientes hacer
  mul= mul*Matriz[i,i]
fin mientras
return mul

```

4. **Construcción de la matriz que se utiliza para poder pasar el test de significancia.** Se necesita construir una matriz similar a 5.1 a partir de los datos iniciales tipificados y trasladados para poder calcular la función t-Student, valor necesario en el test de significancia.

Las componentes de la matriz son a su vez estructuras con 2 campos, lo que permitirá almacenar el valor original de esa posición



(num) y el valor correspondiente a su inversa en esa misma posición (inv).

Las funciones se distinguen por finalizar en `_S`.

- **BuscoUltimaFDist0_S**: Busca la primera fila de la matriz 5.1 a partir de la que se pasa como parámetro que contenga en esa misma posición un 0.
Parámetros de entrada: Número de la fila y número total de variables.
Parámetros de salida: La siguiente fila de la que se pasa como parámetro que contiene un 0 en la misma columna.
- **Inversa_S**: Calcula la inversa de la matriz 5.1.
Parámetros de entrada: Número total de variables.
Parámetros de salida: Ninguno.
- **PasarAMatriz_S**: Construye la matriz 5.1 después de tipificar y trasladar los datos, resultando ser simétrica.
Parámetros de entrada: Número de tuplas de datos.
Parámetros de salida: Ninguno.
- **PermutarFila_S**: Permuta los valores de las filas que se pasan por parámetro.
Parámetros de entrada: El número de las filas que se quieren permutar entre sí.
Parámetros de salida: Ninguno.

Por ejemplo, veamos el código de las funciones **PasarAMatriz-S** y **PermutarFila-S**:



PasarAMatriz-S

```

Nlista v, v1, h, h1;
acum=0.0, prod=0.0, cuad=0.0;
i=0, j=1, n, tj

S[0][0].num=num_datos;
S[0][0].inv=1;
h=v= Segundo de la lista;
  si (Siguiete de h==NULL) entonces
    desde n=0 y mientras n<num_datos hacer
      acum= acum + Contenido de v;
      cuad= cuad + Contenido de v al cuadrado;
      v= Abajo de v;
      n++;
    fin desde
  S[0][1].num= S[1][0].num= acum;
  S[1][1].num= cuad;
  S[0][1].inv= S[1][0].inv= 0;
  S[1][1].inv= 1;
  fin si
sino hacer
  h=v= Primero de la lista;
  mientras Siguiete de h!=NULL hacer
    h=v= Siguiete de h;
    acum= cuad= 0.0;
    desde n=0 y mientras n<num_datos hacer
      acum= acum + v->num;
      cuad= cuad + v->num*v->num;

```



```

        v= Abajo de v;
        n++
    fin desde
    S[0][j].num= S[j][0].num= acum;
    S[0][j].inv= S[j][0].inv= 0;
    S[j][j].num= cuad;
    S[j][j].inv= 1;
    i=j;
    j++;
    tj=j;
    h1=v1= h;
    mientras Siguiete de h1!=NULL hacer
        h1=v1= Siguiete de h1;
        v= h;
        prod= 0.0;
        desde n=0 y mientras n<num_datos hacer
            prod= prod + (Contenido de v)*
                (Contenido de v1);
            v= Abajo de v;
            v1= Abajo de v1;
            n++;
        fin desde
        S[i][tj].num= S[tj][i].num= prod;
        S[i][tj].inv= S[tj][i].inv= 0;
    fin mientras
    fin mientras
    fin sino
    fin;

```



PermutarFila-S(i,j)

```

temp= vector
k=0

mientras k<Num. de filas hacer
  temp[k]=S[i,k].num
  S[i,k].num=S[j,k].num
  S[j,k].num=temp[k]

  temp[k]=S[i,k].inv
  S[i,k].inv=S[j,k].inv
  S[j,k].inv=temp[k]
fin mientras
k= k+1

```

5. **Cálculo de la función t-Student.** Este bloque está compuesto por 3 funciones estadísticas, las cuales han sido extraídas del Numerical Recipes ([32]).

Se caracterizan por finalizar en `_T`:

- **Betai_T:** Devuelve la función beta incompleta $I_x(a, b)$.
Parámetros de entrada: El intervalo donde se quiere calcular la función y el valor de x .
Parámetros de salida: El valor de la función beta incompleta aplicada a los parámetros de entrada.
- **Betacf_T:** Calcula la fracción continua para la función beta incompleta por el método de Lentz's modificado.



Parámetros de entrada: El intervalo donde se quiere calcular la función y el valor de x .

Parámetros de salida: El valor de la fracción continua aplicada a los parámetros de entrada.

- **Gammln_T:** Devuelve el neperiano de la función gamma, $\ln[\Gamma(x)]$, para $x > 0$.

Parámetros de entrada: Valor al que se le quiere aplicar dicha función.

Parámetros de salida: Resultado de aplicar la función al parámetro de entrada.

Como ejemplo, veremos la función **Betai-T**:

Betai-T

```
double bt;
double res;

si x<0.0 ó x>1.0 entonces
    Error: "Incorrecta x";
fin si

si x==0.0 ó x==1.0 entonces
    bt=0.0;
si no hacer
    bt=exp(Gammln_T(a+b)-Gammln_T(a)-Gammln_T(b)+
        a*log(x)+b*log(1.0-x));
fin si
```



```
si  $x < (a+1.0)/(a+b+2.0)$  entonces
    devolver  $bt * \text{Betacf\_T}(a, b, x) / a$ ;
si no hacer
    devolver  $1.0 - bt * \text{Betacf\_T}(b, a, 1.0 - x) / b$ ;
fin si

devolver res;
```

5.3 Ficheros generados

- gen.txt: Fichero con las composiciones de las funciones elementales elegidas dependiendo del orden de las transformadas que se ha elegido.
- comblin.txt: Fichero que contiene únicamente las combinaciones lineales de las composiciones de funciones obtenidas, dependiendo de las variables independientes de las que dispone el fichero de datos.
- result.txt: Fichero que contiene la información resultante de aplicar el método de mínimos cuadrados al fichero de datos elegido, como es el nombre del fichero de datos, los nombres de las variables dependiente e independientes, el número de tuplas de datos, las funciones elementales elegidas, el orden de las transformadas, el número máximo de ecuaciones que se podrán obtener, el modo que se ha elegido para almacenar la/s ecuación/es, la/s ecuación/es obtenidas con sus respectivos coeficientes de determinación, varianza residual, valor del test de Kolgomorov-Smirnov,



test de significancia para cada coeficiente y el valor de cada variable independiente destipificada.

- `residuos.txt`: Fichero en el que son almacenados los residuos correspondientes, es decir, la diferencia entre el valor real de la variable dependiente y el valor teórico que se obtiene con la ecuación obtenida, ordenados ascendentemente.
- Además, el fichero que se tiene que pasar como parámetro en la línea de comando de ejecución del programa. Este fichero almacenará las funciones elementales elegidas (sus correspondientes códigos) para trabajar posteriormente con ellas.

5.4 Orden de ejecución de las funciones

Cuando se inicia el programa se pregunta si se dispone de fichero de datos o hay que generarlo. El fichero de datos debe de cumplir las siguientes condiciones:

1. En la primera línea deben de aparecer los nombres de las variables, compuestos por caracteres alfanuméricos (números y caracteres). Ello determinará el número de variables y por lo tanto las columnas de datos.
2. El nombre de la última variable corresponde con la variable dependiente y el resto serán las independientes.
3. En las siguientes líneas serán guardados los datos correspondientes a dichas variables formando columnas. Estos datos pueden



ser tanto de tipo entero, decimal (siendo el punto (.) el separador de los decimales) o en notación científica.

4. Tanto los datos como el nombre de las variables pueden ir separados por uno o varios espacios o tabuladores.

En el caso de que no se tenga el fichero de datos almacenado se puede construir en este momento: se pregunta por el nombre del fichero que se quiere crear, el nombre de la variable dependiente, el número de las variables independientes así como sus nombres, y por el número de tuplas de datos de las que se dispone. Seguidamente se van preguntando los datos uno a uno. Al finalizar todo quedará guardado en el fichero del que se ha preguntado el nombre.

A continuación aparece un menú con las posibles funciones a elegir 5.1 así como sus respectivos coeficientes, sin posibilidad de repetir ni de sobrepasar el número máximo de funciones que hay. Luego se pregunta por el orden de transformadas.

Seguidamente se pregunta por el modo de almacenamiento de las funciones:

Modo de almacenamiento de las funciones	
1.	Almacenarlas todas
2.	Almacenar aquellas cuyo coeficiente de determinación esté dentro de un rango
3.	Almacenar la que tenga mejor coeficiente de determinación

En el caso de elegir la segunda opción se preguntará por el rango mayor y menor.



Cabe la posibilidad de pasar el test de significancia a cada uno de los coeficientes de la ecuación, de manera que se pregunta si se quiere pasar o no.

Con todos estos datos el programa se ejecuta haciéndose los cálculos necesarios para poder aplicar el método de los mínimos cuadrados y obtener los distintos coeficientes de la ecuación.

En el siguiente recuadro se puede observar un ejemplo de ejecución del programa. En la primera línea se introduce el comando que ejecuta el programa nuestro llamado MODELHSS junto con el nombre del fichero en el que se almacenarán las funciones elementales elegidas, es decir, en "salida". A continuación, se elige un fichero que está almacenado conteniendo los nombres de las variables y los datos correspondientes llamado "datos.txt". Cuando finaliza la lectura de este fichero aparece el menú con las posibles funciones elementales a elegir. Se ha decidido, en este caso, elegir x^2 y $\exp(0.01x)$. Seguidamente se pregunta por el orden de transformadas, donde se ha elegido 2, de modo que hará composiciones de estas dos funciones hasta orden 2. Elegido el orden se procede a la composición de estas funciones sacando por pantalla el número máximo de ecuaciones que se podrán generar. Esto dará una idea de si se quieren almacenar todas, las que estén dentro de un rango o la que cumpla con el mejor coeficiente de determinación. Se ha elegido la primera opción porque en este caso son pocas las ecuaciones resultantes de modo que el fichero de salida no será demasiado grande. Se pregunta si se quiere pasar el test de significancia y posteriormente empieza realmente el proceso de buscar esas ecuaciones. Al finalizar, se indica que en el fichero "result.txt" están almacenados los resultados obtenidos.



```
C:\go32 ModelHSS salida go32
version 1.12 Copyright (C) 1994 DJ Delorie
El fichero de datos....
  1.- Esta almacenado?
  2.- Quiere introducir los datos?
Opcion (1 o 2)?: 1

Fichero que contiene los datos de las variables?: datos.txt
Leyendo fichero de datos...

      ELECCION DE LAS FUNCIONES A COMPONER
  1.- Potencia [x^A] (A>0)
  2.- Raiz [[x^(1/A)]
  3.- Inversa [1/Ax] (A>0)
  4.- Seno [sen(Ax)]
  5.- Exponencial [exp(Ax)] (A>0)
  6.- Inversa de la exponencial [1/exp(Ax)] (A>0)
  7.- Arcotangente [arctg(Ax)]
  8.- Logaritmo neperiano del valor absoluto [ln|Ax|]
Funcion ('f'/'F' para fin): 1
Factor: 2
Funcion ('f'/'F' para fin): 5
Factor: 0.01
Funcion ('f'/'F' para fin): f
Orden de transformadas?: 2

Realizando la composicion de funciones...
      En total se generaran 8 ecuaciones.
```

Continúa en la página siguiente...



```

Las funciones...
  1.- Quiere almacenarlas todas?
  2.- Quiere almacenar aquellas cuyo coeficiente de
      determinacion esten dentro de un rango?
  3.- Quiere almacenar la que tenga mejor coeficiente
      de determinacion?
Opcion [1, 2 o 3]....1
Desea pasar Test de Significancia (S/N)?: s

Proceso de tipificar y trasladar los datos....
....proceso finalizado.

Construyendo las ecuaciones...

      Para ver los resultados edite el fichero 'result.txt'.

```

1. Lectura y almacenamiento del fichero de datos. Se procede a la lectura del fichero de datos, siendo almacenados éstos en la segunda estructura de datos llamada *Nlista*. Posteriormente serán tratados estadísticamente junto con las funciones elegidas. Este procedimiento lo componen las funciones:

```

AlmacenarDatosVariables()
LeerFicheroD()
PermutarColumnaFicheroD()

```

Parámetros de entrada: Ninguno.

Parámetros de salida: Número de tuplas de datos.

2. Elección de las funciones de primer orden, donde se comprueba que no se repitan ni se exceda del número máximo de funciones a elegir. Son almacenadas en el fichero que se introduce como



parámetro en la línea de comando para la ejecución del programa, y recuperadas posteriormente. Formado por:

```
EleccionFunciones()  
  EstaEnLista.F()  
  LeerFichero.F()  
    BajarLinea.F()  
      Apilar.F()  
    Fin.F()
```

Parámetros de entrada: Nombre del fichero que contendrá las funciones elementales elegidas.

Parámetros de salida: Número de funciones elementales elegidas.

3. Preguntar por el orden de las transformadas para generar las composiciones y pasar a almacenarlas en el fichero "gen.txt". Este módulo está compuesto por:

```
ComposicionYCombinacion()  
  GenerarFunciones.F()  
  EscribirFunciones.F()  
  GenerarCadena.F()
```

Parámetros de entrada: Fichero que contiene las funciones elementales elegidas y el número de funciones elementales elegidas.

Parámetros de salida: Orden de las transformadas.

4. Generar las combinaciones lineales de las composiciones, según el número de variables independientes, y almacenarlas en el fichero "comblin.txt". Formado por:



```

ComposicionYCombinacion()
  CombinarFunciones2_F()
    DesapilarHorizontal_F()
      EsVacía_F()
        EscribirFunciones_F()
          GenerarCadena_F()

```

Parámetros de entrada: Fichero que contiene las funciones elementales elegidas y el número de funciones elementales elegidas.

Parámetros de salida: Orden de las transformadas.

5. Preguntar sobre el tipo de ecuaciones que se quieren almacenar, porque cabe la posibilidad de almacenar todas las que se generen, las que estén dentro de un rango referido al coeficiente de determinación o aquella que consigue el mejor coeficiente de determinación. Se incluye en una única función:

```
ModoAlmacenamientoFunciones()
```

Parámetros de entrada: Ninguno.

Parámetros de salida: Modo de almacenamiento de las funciones.

6. Tipificar y trasladar los datos iniciales. Esto se realiza para acotar los problemas que pueden surgir a la hora de aplicar las funciones en los datos debido a sobrepasar el rango máximo y mínimo del tipo de datos (aunque se ha escogido el tipo que permite el mayor rango dentro de C). Para tipificar se aplica la fórmula: $\frac{x_i - \bar{x}}{s_x^2}$, donde \bar{x} es la media de la variable x y s_x^2 es su varianza. Y para trasladar se suma 4 para evitar en todo lo posible el 0, ya que



nos podríamos encontrar con divisiones por 0. Es aquí donde se pregunta si se quiere pasar el test de significancia, ya que los cálculos realizados en este apartado sirven para construir la matriz que se utilizará en el test de significancia. De este modo si se contesta afirmativamente se pasará a la construcción de dicha matriz y si no, lo pasará por alto. Sus funciones son:

```
TipificarYTrasladarDatosIniciales()  
  Medias_D()  
  Varianzas_D()  
  TipificarYTrasladar_D()  
  PasarAMatriz_S()
```

Parámetros de entrada: Número de tuplas de datos.

Parámetros de salida: Ninguno.

7. Calcular los elementos de la matriz. Es el primer paso para construir las ecuaciones: inicialmente se necesita aplicar las combinaciones lineales de las funciones a los correspondientes datos de manera que se pasa a formar una nueva tabla de datos. Sobre esta nueva tabla de datos se calculan las medias, la desviación típica y las covarianzas entre ellos. De este modo se pueden calcular los elementos de la matriz de varianzas y covarianzas ya que el elemento a_{ij} se corresponde con el coeficiente de correlación (r_{ij}) entre las variables independientes i y j . Este coeficiente de correlación es determinado como la covarianza entre ellas dividida por el producto de sus desviaciones típicas. Se incluyen las siguientes funciones:

```
ConstruirEcuaciones()
```



```

AplicarFuncionesADatos_D()
  Calcular_F()
Medias_D()
Varianzas_D()
Covarianzas_D()

```

Parámetros de entrada: Cadena con los códigos de las funciones compuestas y combinadas, número de tuplas de datos, modo de almacenamiento de las funciones y si se trata de funciones elementales o no.

Parámetros de salida: Ninguno.

8. Construir la matriz y aplicar el método de los mínimos cuadrados para calcular los coeficientes de la ecuación, así como el coeficiente de determinación y la varianza residual que se obtienen. Distinguir que la matriz de correlaciones es simétrica por ser r_{ij} igual a r_{ji} , y por ser $r_{ii} = 1$ la diagonal principal es igual a 1. También es aquí donde se aplica el test de Kolgomorov-Smirnov y el test de significación, en el caso de que se haya solicitado. Además se van guardando en el fichero "result.txt" las ecuaciones obtenidas. Hay que recordar que previamente los datos iniciales han sido tipificados, de modo que se trabaja con las variables tipificadas. Al escribirlas en el fichero, la variable dependiente se destipifica, pero el resto de las variables no, indicando seguidamente a qué equivale cada variable. Las funciones que cumplen este propósito son:

```

ConstruirEcuaciones()
  InicializarMatriz_M()

```



```
PasarAMatriz_M()  
Coeficientes_M()  
  Regresiones_M()  
    CopiarMatriz_M()  
    QuitoFilaColumna_()  
    Diagonalizar_M()  
      BuscoUltimaFDist0_M()  
      PermutarFila_M()  
Residuos_D()  
VarianzaResidual_D()  
TipificarResiduos_D()  
Sort_D()  
Ksone_D()  
  Fidis_D()  
    Erfcc_D()  
      ABS_M()  
  Fmax_D()  
  Probks_D()  
    ABS_M()  
CoeficienteDeterminacion_D()  
Inversa_S()  
  BuscoUltimaFDist0_S()  
  PermutarFila_S()  
DeshacerCambioTipificar()  
EscribirEcuacionEnFichero_M()  
  GenerarSubcadena_F()  
  BetaI_T()  
    Gamln_T()  
    Betacf_T()
```



EscribirColetilla()

Parámetros de entrada: Cadena con los códigos de las funciones compuestas y combinadas, número de tuplas de datos, modo de almacenamiento de las funciones y si se trata de funciones elementales o no.

Parámetros de salida: Ninguno.

9. Repetir el proceso mientras queden posibles ecuaciones a generar. Hasta que no se hayan probado todas las combinaciones lineales de las composiciones de las funciones elegidas el algoritmo no habrá finalizado, de modo que se pasará a eliminar los datos generados y en su caso las funciones elegidas. Estas funciones serían:

```
main()  
  Desapilar_F  
  Desapilar_D  
  BorrarDatos_D
```

5.5 Secuencia de llamadas

Se expone a continuación el orden de llamada de las funciones, así como la dependencia existente entre ellas:

```
main();  
  AlmacenarDatosVaribles();  
  LeerFichero_D();
```



```
    PermutarColumnaFichero_D();
EleccionFunciones();
    EstaEnLista_F();
ComposicionYCombinacion();
    EleccionFunciones();
        LeerFichero_F();
            BajarLinea_F();
                Apilar_F();
                    Fin_F();
GenerarFunciones_F();
EscribirFunciones_F();
    GenerarCadena_F();
CombinarFunciones_F();
    DesapilarHorizontal_F();
        EsVacía_F();
EscribirFunciones_F();
    GenerarCadena_F();
ModoAlmacenamientoFunciones();
TipificarYTrasladarDatosIniciales();
    Medias_D();
    Varianzas_D();
    TipificarYTrasladar_D();
    PasaAMatriz_S();
ConstruirEcuaciones();
    AplicarFuncionesADatos_D();
        Calcular_F();
    Medias_D();
    Varianzas_D();
    Covarianzas_D();
```



```
InicializarMatriz_M();
PasarAMatriz_M();
Coeficientes_M();
  Regresiones_M();
    CopiarMatriz_M();
    QuitoFilaColumna_M();
    Diagonalizar_M();
      BuscoUltimaFDist0_M();
      PermutarFila_M();
Residuos_D();
VarianzaResidual_D();
TipificarResiduos_D();
Sort_D();
Ksone_D();
  Fidis_D();
    Erfcc_D();
      ABS_M();
  Fmax_D();
  Probks_D();
    ABS_M();
CoeficienteDeterminacion_D();
Inversa_S();
  BuscoUltimaFDist0_S();
  PermutarFila_S();
DeshacerCambioTipificar();
EscribirEcuacionEnFichero_M();
  GenerarSubcadena_F();
  Betail_T();
    Gammln_T();
```



```
        Betacf_T();
    ABS_M();
    GenerarSubcadena_F();
    Betai_T();
    Gamln_T();
    Betacf_T();
    EscribirColetilla();
    Desapilar_F();
    EsVacia_F();
    Desapilar_D();
    BorrarDatos_D();
fin.
```



Universitat d'Alacant
Universidad de Alicante

Capítulo 6

Aplicaciones

En este capítulo se aplicará la metodología presentada en el proyecto mediante la ejecución del software, a diversas modelizaciones, que nos expresen relaciones entre variables, conocidos los datos experimentales de las mismas.

Serán utilizados ajustes lineales y no lineales comparando además los resultados con los programas existentes en el mercado, como pueden ser el SPLUS o el SPSS. Son de interés los estudios que se realizarán en datos experimentales que presentan cierta problemática, ya que nos permitirá conocer el tratamiento de los mismos con el software presentado en el proyecto y los programas existentes comerciales.



6.1 Caso lineal: $Y = aX + b$

A continuación examinaremos el caso lineal para el software (MODEL-HSS), así como su comparación con SPLUS y SPSS, observando que los resultados obtenidos son los mismos.

Se ha extraído la tabla de datos 6.1 del libro de Probabilidad y estadística ([45]) y se corresponde con unos datos experimentales que se obtuvieron de 33 muestras.

Tabla 6.1: Mediciones de sólidos y demanda de oxígeno químico

Reducción de sólidos, x (%)	Demanda de oxígeno químico, y (%)
3	5
7	11
11	21
15	16
18	16
27	28
29	27
30	25
30	35
31	30
31	40
32	32
33	34
33	32

Continúa en la página siguiente...



Tabla 6.1: Mediciones de sólidos y demanda de oxígeno químico

Reducción de sólidos, $x(\%)$	Demanda de oxígeno químico, $y(\%)$
34	34
36	37
36	38
36	34
37	36
38	38
39	37
39	36
39	45
40	39
41	41
42	40
42	44
43	37
44	44
45	46
46	46
47	49
50	51

Estas muestras corresponden a los desperdicios que se tratan químicamente en el estudio “Chemical Treatment on Spent Vegetable Tan Liquor” realizado en la Virginia Polytechnic Institute and State University en 1970. Se registraron las lecturas de x : reducción porcentual del



total de sólidos y de y : reducción porcentual de demanda de oxígeno químico.

6.1.1 Ejecución con MODELHSS

Se le ha pasado la tabla de datos 6.1 al programa MODELHSS y se han obtenido los siguientes resultados:

```

El fichero 'p1' tiene como variable dependiente 'y' y dispone
de 1 variable independiente:
Variable 1= x

El fichero p1 contiene 33 tuplas de datos.

Han sido elegidas 1 funciones: x

El orden de transformadas es de 1
    Como maximo se generaran 1 ecuaciones.

Se ha elegido almacenar todas las ecuaciones generadas.

ECUACION N.1: y= -7.11312 + 10.2934*x$

Varianza Residual= [10.429913]

Coeficiente de Determinacion= [0.912941]  KS= [0.785577]

TEST DE SIGNIFICANCIA
Termino independiente: 0.004969
Coeficiente de x$: 0.000000
-----
DONDE: x$=0.087788*x+1.06308
  
```



Como se observa, la ecuación obtenida es:

$$y = -7.11312 + 10.2934x\$ \quad (6.1)$$

donde la variable independiente x se ha transformado por $x\$$, de manera que la equivalencia es la siguiente:

$$x\$ = 0.087788x + 1.06308 \quad (6.2)$$

Sustituyendo 6.2 en 6.1 la ecuación resultante quedaría:

$$\begin{aligned} y &= -7.11312 + 10.2934(0.087788x + 1.06308) = \\ &= 3.829588 + 0.903637x \end{aligned} \quad (6.3)$$

con un coeficiente de determinación del 0.912941.

El valor del test de Kolgomorov-Smirnov es del 0.785577, al ser mayor que 0.05 indica que si que lo cumple y, por lo tanto, los residuos siguen una distribución normal.

En cuanto al test de significancia se observa que ambos coeficientes, el independiente y el de x son significativos (por ser sus valores menores que 0.05).

Se puede observar en la gráfica 6.1 como la ecuación obtenida se corresponde a una recta, la cual se acerca lo mejor posible al conjunto de puntos reales de los que se dispone, es por ello que el coeficiente de determinación es bastante bueno. En el eje x se representan los valores de la variable independiente x y en el eje y los correspondientes de la variable dependiente y . Los puntos representados por el símbolo



+ se corresponden con los datos reales de la tabla 6.1. El símbolo o representa a los valores de y que se obtienen al sustituir la variable independiente x de la tabla 6.1 en la ecuación obtenida 6.3, es decir, representan a los valores teóricos que se obtendrían.

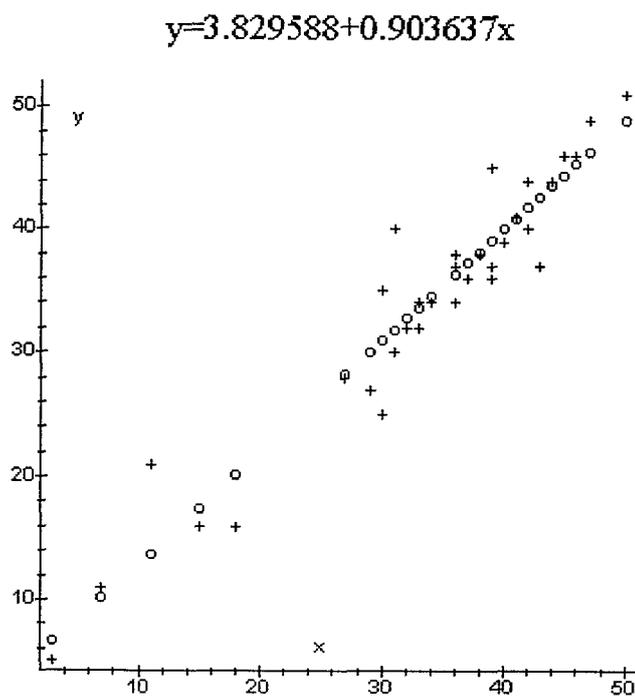


Figura 6.1: Representación gráfica



6.1.2 Ejecución con SPLUS

Se ha introducido la tabla 6.1 en el programa estadístico SPLUS y se ha ejecutado una regresión lineal, mediante el método de los mínimos cuadrados, para obtener los siguientes resultados:

```

*** Linear Model ***

Call: lm(formula = y ~ x, data = p1, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-5.939  -1.783  -0.228   1.506   8.157

Coefficients:
                Value  Std. Error  t value  Pr(>|t|)
(Intercept)  3.8296    1.7684     2.1655   0.0382
             x   0.9036    0.0501    18.0300   0.0000

Residual standard error: 3.23 on 31 degrees of freedom
Multiple R-Squared: 0.9129
F-statistic: 325.1 on 1 and 31 degrees of freedom, the
p-value is 0
  
```

En este caso, la ecuación resultante es:

$$y = 3.8296 + 0.9036x \quad (6.4)$$

Como se observa, esta ecuación 6.4 es la misma que la obtenida por el programa MODELHSS, 6.3, además de coincidir el coeficiente de determinación.

Ambos coeficientes son significativos por ser $Pr(> |t|) < 0.05$,



aunque el término independiente, en este caso, comparado con el obtenido con MODELHSS (con la ecuación que resulta cuando la variable independiente es tipificada y trasladada), tiene una significancia menor (está más cerca de 0.05).

Se realiza a continuación una segunda prueba: se creará una nueva tabla de datos donde la variable dependiente va a ser la misma (y), pero la variable independiente (x_1) estará tipificada y trasladada con respecto a la variable independiente original x . Así, podremos comparar los resultados que se obtengan directamente con los de MODELHSS, y veremos si los coeficientes o sus significancias difieren. La nueva tabla de datos se muestra a continuación. Para tipificar la variable x_1 se ha tomado su media $\bar{x} = 33.454545$ y su varianza $\sigma_x = 11.391035$:

Tabla 6.2: Tabla con la variable x de la tabla 6.1 tipificada y trasladada.

$x_1 = \frac{x - \bar{x}}{\sigma_x} + 4(\%)$	$y(\%)$
1.326602	5
1.677788	11
2.028973	21
2.380158	16
2.643547	16
3.433714	28
3.609306	27
3.697103	25
3.697103	35

Continúa en la página siguiente...



Tabla 6.2: Tabla con la variable x de la tabla 6.1 tipificada y trasladada.

$x_1 = \frac{x - \bar{x}}{\sigma_x} + 4(\%)$	$y(\%)$
3.784899	30
3.784899	40
3.872695	32
3.960492	34
3.960492	32
4.048288	34
4.223881	37
4.223881	38
4.223881	34
4.311677	36
4.399473	38
4.487270	37
4.487270	36
4.487270	45
4.575066	39
4.662862	41
4.750658	40
4.750658	44
4.838455	37
4.926251	44
5.014047	46
5.101844	46

Continúa en la página siguiente...

Caso lineal: $Y = aX + b$

C6

Tabla 6.2: Tabla con la variable x de la tabla 6.1 tipificada y trasladada.

$x_1 = \frac{x - \bar{x}}{\sigma_x} + 4(\%)$	$y(\%)$
5.189640	49
5.453029	51

Y los resultados obtenidos se muestran a continuación:

```

*** Linear Model ***
Call: lm(formula = y ~ x1, data = p1, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-5.939  -1.783  -0.228   1.506   8.157

Coefficients:
                Value  Std. Error t value Pr(>|t|)
(Intercept)  -7.1135    2.3518   -3.0247  0.0050
             x1  10.2925    0.5709   18.0300  0.0000

Redisual standard error: 3.23 on 31 degrees of freedom
Multiple R-Squared: 0.9129
F-statistic: 325.1 on 1 and 31 degrees of freedom, the
p-value is 0

```



Comprobamos que es la misma ecuación que la obtenida en 6.1, además, se tipifique o no, el valor del coeficiente de determinación coincide en ambos casos, así como el valor de los residuos. Sólo cambia el término independiente y por lo tanto su significancia (aumentando). Además, estos datos son los mismos que los obtenidos con MODELHSS.

6.1.3 Ejecución con SPSS

Se ha realizado la misma operación que con el SPLUS, es decir, se ha cogido la tabla de datos original 6.1 y se ha introducido en el SPSS para realizar un análisis de regresión lineal. Los resultados obtenidos se presentan a continuación:

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	X ^a	,	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Y

Caso lineal: $Y = aX + b$

C6

Resumen del modelo^b

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,955 ^a	,913	,910	3,23

a. Variables predictoras: (Constante), TIPX

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	3390,551	1	3390,551	325,080	,000 ^a
	Residual	323,327	31	10,430		
	Total	3713,879	32			

a. Variables predictoras: (Constante), X

b. Variable dependiente: Y

Coeficientes^a

Modelo		Coefic. no estand.		Coefic. estand.	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	3,830	1,768		2,166	,038
	X	,904	,050	,955	18,030	,000

a. Variables predictoras: (Constante), X

b. Variable dependiente: Y

En este caso, si se observa la tabla *Coeficientes^a*, se puede comprobar que los coeficientes de la ecuación son: como término independiente 3.830 y, como coeficiente de x , 0.904. En la columna denominada *Sig.* se indica la significancia de estos coeficientes, que como se comprueba son



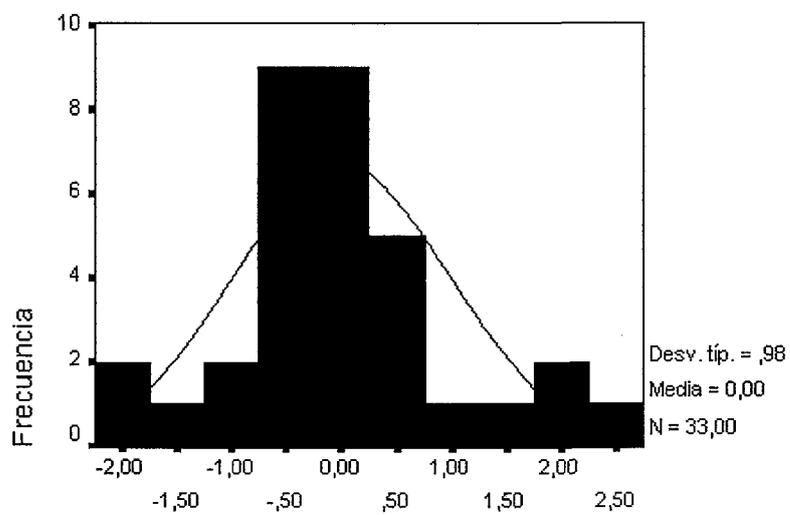
ambos significativos por ser este valor menor o igual que 0.05. Además en la tabla *Resumen del modelo*^b aparece el coeficiente de determinación, llamado *R cuadrado*, cuyo valor es 0.913. Estos resultados coinciden con los resultados obtenidos con el software implementado en MODELHSS y con SPLUS.

Como se obtienen los mismos resultados se puede extraer, para todos ellos, un histograma, según la figura 6.2, de los residuos tipificados para comprobar que, efectivamente, siguen una distribución normal.



Histograma

Variable dependiente: Y



Regresión Residuo tipificado

Figura 6.2: Histograma del residuos tipificado

Caso lineal: $Y = aX + b$

C6

Si ahora tipificásemos y trasladásemos la variable independiente obtendríamos la misma tabla 6.2 para obtener los resultados que a continuación se indican:

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	TIPX ^a		Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Y

Resumen del modelo^b

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,955 ^a	,913	,910	3,23

a. Variables predictoras: (Constante), TIPX

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	3390,551	1	3390,551	325,080	,000 ^a
	Residual	323,327	31	10,430		
	Total	3713,879	32			

a. Variables predictoras: (Constante), X

b. Variable dependiente: Y

Coeficientes^a

Modelo	Coefic. no estand.		Coefic. estand.	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	-7,113	2,352		-3,025	,005
TIPX	10,292	,571	,955	18,030	,000

a. Variables predictoras: (Constante), X

b. Variable dependiente: Y

En este caso solo se modifican los coeficientes obtenidos, resultando la ecuación:

$$y = -7.113 + 10.292 \text{ TIPX}$$

que coincide con la obtenida con MODELHSS. Como se observa, la significancia del término independiente ha subido con respecto al mismo ejemplo sin tipificar ni trasladar.

6.2 Caso lineal múltiple

Se examina en este apartado la obtención de linealidad múltiple desde el software presentado y su comparación con SPLUS y SPSS, al igual que en el anterior apartado.

Vamos a considerar en este apartado, la determinación de una relación lineal entre cinco variables, cuatro de las cuales son las variables



independientes. Supongamos que se refiere a la relación entre el comportamiento hacia el trabajo (y) de 12 trabajadores y las calificaciones de cuatro pruebas (x_1, x_2, x_3, x_4). El fichero de datos, en este caso, sería el mostrado en la tabla 6.3.

y	x_1	x_2	x_3	x_4
11.2	56.5	71.0	38.5	43.0
14.5	59.5	72.5	38.2	44.8
17.2	69.2	76.0	42.5	49.0
17.8	74.5	79.5	43.4	56.3
19.3	81.2	84.0	47.5	60.2
24.5	88.0	86.2	47.4	62.0
21.2	78.2	80.5	44.5	58.1
16.9	69.0	72.0	41.8	48.1
14.8	58.1	68.0	42.1	46.0
20.0	80.5	85.0	48.1	60.3
13.2	58.3	71.0	37.5	47.1
22.5	84.0	87.2	51.0	65.2

Tabla 6.3: Relación entre el trabajo y 4 pruebas

6.2.1 Ejecución con MODELHSS

Tomando el caso lineal $y = b_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$ con MODELHSS se obtienen los siguientes resultados:



```

El fichero 'p2' tiene como variable dependiente 'y' y
dispone de 4 variables independientes:
Variable 1= x1
Variable 2= x2
Variable 3= x3
Variable 4= x4

El fichero p2 contiene 12 tuplas de datos.

Han sido elegidas 1 funciones:
x

El orden de transformadas es de 1
    Como maximo se generaran 1 ecuaciones.

Se ha elegido almacenar todas las ecuaciones generadas.

ECUACION N.1: y= 2.74326 + 4.73603*x1$ - 2.01957*x2$ +
              0.070635*x3$ + 0.96667*x4$

Varianza Residual= [1.493313]

Coeficiente de Determinacion= [0.939009]  KS= [0.568057]

TEST DE SIGNIFICANCIA
Termino independiente: 0.121043
Coeficiente de x1$: 0.015617
Coeficiente de x2$: 0.242647
Coeficiente de x3$: 0.949206
Coeficiente de x4$: 0.622309

-----

DONDE:
x1$=0.088903*x1-2.34918
x2$=0.146457*x2-7.38581
x3$=0.231898*x3-6.09724
x4$=0.12895*x4-2.87842
  
```



De modo que:

- Ecuación obtenida:

$$\begin{aligned}
 y &= 2.74326 + 4.73603 x_1 - 2.01957 x_2 + 0.070635 x_3 + \\
 &\quad + 0.96667 x_4 = \\
 &= 2.74326 + 4.73603(0.088903 x_1 - 2.34918) - \\
 &\quad - 2.01957(0.146457 x_2 - 7.38581) + \\
 &\quad + 0.070635(0.231898 x_3 - 6.09724) + \\
 &\quad + 0.96667(0.12895 x_4 - 2.87842) = \\
 &= 3.320473 + 0.421047 x_1 - 0.29578 x_2 + 0.01638 x_3 + \\
 &\quad + 0.124652 x_4
 \end{aligned}$$

- Coeficiente de determinación: 0.939009
- Variables significativas: Solamente x_1 .

Si ahora probamos hacer una transformación de los datos aplicándoles alguna función veremos si es posible mejorar el coeficiente de determinación. Para ello consideramos las siguientes funciones: x , $\ln|x|$ y $\exp(0.1x)$, de manera que, para el orden de transformadas uno, se obtienen 81 ecuaciones. Eligiendo en el programa la opción de considerar sólo la ecuación con mayor coeficiente de determinación, se almacenará una ecuación:



El fichero 'p2' tiene como variable dependiente 'y' y dispone de 4 variables independientes:

Variable 1= x1

Variable 2= x2

Variable 3= x3

Variable 4= x4

El fichero p2 contiene 12 tuplas de datos.

Han sido elegidas 3 funciones:

x

$\ln|x|$

$\exp(0.10*x)$

El orden de transformadas es de 1

Como máximo se generaran 81 ecuaciones.

Solo se almacenar\ '{a}' la ecuaci\ '{o}'n con mejor CD.

ECUACION N.1: $y = -31.362492 + 36.2223*\exp(0.10*x1\$) - 8.60612*\ln(|x2\$|) - 1.14376*\ln(|x3\$|) + 5.38047*\exp(0.10*x4\$)$

Varianza Residual= [1.284625]

Coefficiente de Determinacion= [0.947532] KS= [0.674399]

TEST DE SIGNIFICANCIA

Termino independiente: 0.011873

Coefficiente de x1\$: 0.010281

Coefficiente de x2\$: 0.156606

Coefficiente de x3\$: 0.776649

Coefficiente de x4\$: -0.620869

 DONDE:

$x1\$ = 0.088903*x1 - 2.34918$

$x2\$ = 0.146457*x2 - 7.38581$

$x3\$ = 0.231898*x3 - 6.09724$

$x4\$ = 0.12895*x4 - 2.87842$



Sí que parece que se mejore un poco el coeficiente de determinación, además de recuperar el término independiente como significativo:

- Ecuación obtenida:

$$\begin{aligned}
 y &= -31.362492 + 36.2223 \exp(0.1x_1) - 8.60612 \ln |x_2| - \\
 &\quad -1.14376 \ln |x_3| + 5.38047 \exp(0.1x_4) = \\
 &= -31.362492 + 36.2223 \exp(0.0088903x_1 - 0.234918) - \\
 &\quad -8.60612 \ln |0.146457x_2 - 7.38581| - \\
 &\quad -1.14376 \ln |0.231898x_3 - 6.09724| + \\
 &\quad +5.38047 \exp(0.012895x_4 - 0.287842)
 \end{aligned}$$

- Coeficiente de determinación: 0.947532
- Variables significativas: Término independiente y x_1 .

En este caso podemos ver un pequeño ejemplo aplicado de igual manera al mismo fichero de datos de la tabla 6.3 cuando se elige como orden de transformadas un valor superior a 1, además de elegir que se almacenen las funciones que están dentro de un rango de coeficientes de determinación. Como se puede observar el número de ecuaciones aumenta considerablemente y, con ello, el tamaño del fichero de salida debido al conjunto de ecuaciones que cumplen el rango elegido. Se va a probar con las mismas funciones, con un orden de transformadas de hasta 2 y que se almacenen aquellas ecuaciones cuyo coeficiente de determinación esté entre 0.947 y 0.999 (es fácil comprobar que la ecuación resultante anterior se encuentra entre las que se almacenan ahora):



```
El fichero 'p2' tiene como variable dependiente 'y' y dispone
de 4 variables independientes:
Variable 1= x1
Variable 2= x2
Variable 3= x3
Variable 4= x4

El fichero p2 contiene 12 tuplas de datos.

Han sido elegidas 3 funciones:
x
ln|x|
exp(0.10*x)

El orden de transformadas es de 2
    Como maximo se generaran 20736.000000 ecuaciones.

Se almacenaran las ecuaciones cuyo CD este entre 0.947000 y
0.990000.
.
.
ECUACION N. 8: y= -26.4689 + 36.0416*exp(0.10*x1$) -
8.43447*ln(|x2$|) - 1.02807*ln(|x3$|) + 0.76276*x4$

Varianza Residual= [1.291405]

Coeficiente de Determinacion= [0.947255]    KS= [0.629295]

TEST DE SIGNIFICANCIA
Termino independiente: 0.016884
Coeficiente de x1$: 0.011311
Coeficiente de x2$: 0.160091
Coeficiente de x3$: 0.796835
Coeficiente de x4$: 0.646518
```

Continúa en la página siguiente...



ECUACION N. 9: $y = -31.3625 + 36.2223 \cdot \exp(0.10 \cdot x1) - 8.60612 \cdot \ln(|x2|) - 1.14376 \cdot \ln(|x3|) + 5.38047 \cdot \exp(0.10 \cdot x4)$

Varianza Residual= [1.284625]

Coefficiente de Determinacion= [0.947532] KS= [0.674399]

TEST DE SIGNIFICANCIA

Termino independiente: 0.011873

Coefficiente de x1\$: 0.010281

Coefficiente de x2\$: 0.156606

Coefficiente de x3\$: 0.776649

Coefficiente de x4\$: 0.620869

ECUACION N. 884: $y = 42.5669 + 35.5357 \cdot (\exp(0.10 \cdot x1)) - 74.9167 \cdot \exp(0.10 \cdot \ln(|x2|)) - 0.273302 \cdot (x3) + 5.90974 \cdot \exp(0.10 \cdot x4)$

Varianza Residual= [1.287405]

Coefficiente de Determinacion= [0.947419] KS= [0.566935]

TEST DE SIGNIFICANCIA

Termino independiente: 0.312032

Coefficiente de x1\$: 0.008161

Coefficiente de x2\$: 0.153660

Coefficiente de x3\$: 0.803481

Coefficiente de x4\$: 0.618082

Continúa en la página siguiente...



ECUACION N. 885: $y = 45.9433 + 35.3639 * (\exp(0.10 * x1\$)) - 72.8999 * \exp(0.10 * \ln(|x2\$|)) - 0.213771 * (x3\$) + 0.79735 * (x4\$)$

Varianza Residual= [1.296577]

Coefficiente de Determinacion= [0.947044] KS= [0.516780]

TEST DE SIGNIFICANCIA

Termino independiente: 0.330379

Coefficiente de x1\$: 0.009305

Coefficiente de x2\$: 0.158094

Coefficiente de x3\$: 0.841021

Coefficiente de x4\$: 0.652905

ECUACION N. 1471: $y = -324.161 + 301.543 * \exp(0.10 * \exp(0.10 * x1\$)) - 7.75088 * \ln(|x2\$|) - 0.709483 * (\exp(0.10 * x3\$)) + 2.33405 * \ln(|x4\$|)$

Varianza Residual= [1.297410]

Coefficiente de Determinacion= [0.947010] KS= [0.560675]

TEST DE SIGNIFICANCIA

Termino independiente: 0.010536

Coefficiente de x1\$: 0.011181

Coefficiente de x2\$: 0.154970

Coefficiente de x3\$: 0.911174

Coefficiente de x4\$: 0.711003

Continúa en la página siguiente...



ECUACION N. 1472: $y = -320.413 + 299.729 \cdot \exp(0.10 \cdot \exp(0.10 \cdot x1)) - 7.53674 \cdot \ln(|x2|) - 0.420413 \cdot (\exp(0.10 \cdot x3)) + 2.77805 \cdot \ln(|\ln(|x4|)|)$

Varianza Residual= [1.296334]

Coefficiente de Determinacion= [0.947054] KS= [0.503522]

TEST DE SIGNIFICANCIA

Termino independiente: 0.013737

Coefficiente de x1\$: 0.012605

Coefficiente de x2\$: 0.152076

Coefficiente de x3\$: 0.944525

Coefficiente de x4\$: 0.705617

ECUACION N. 1612: $y = -324.161 + 301.543 \cdot \exp(0.10 \cdot \exp(0.10 \cdot x1)) - 7.75088 \cdot (\ln(|x2|)) - 0.709483 \cdot (\exp(0.10 \cdot x3)) + 2.33405 \cdot \ln(|x4|)$

Varianza Residual= [1.297410]

Coefficiente de Determinacion= [0.947010] KS= [0.560675]

TEST DE SIGNIFICANCIA

Termino independiente: 0.010536

Coefficiente de x1\$: 0.011181

Coefficiente de x2\$: 0.154970

Coefficiente de x3\$: 0.911174

Coefficiente de x4\$: 0.711003

Continúa en la página siguiente...



ECUACION N. 1613: $y = -320.413 + 299.729 \cdot \exp(0.10 \cdot \exp(0.10 \cdot x1)) - 7.53674 \cdot (\ln(|x2|)) - 0.420413 \cdot (\exp(0.10 \cdot x3)) + 2.77805 \cdot \ln(|\ln(|x4|)|)$

Varianza Residual= [1.296334]

Coefficiente de Determinacion= [0.947054] KS= [0.503522]

TEST DE SIGNIFICANCIA

Termino independiente: 0.013737

Coefficiente de x1\$: 0.012605

Coefficiente de x2\$: 0.152076

Coefficiente de x3\$: 0.944525

Coefficiente de x4\$: 0.705617

.
. .
ECUACION N. 2065: $y = -253.515 + 304.043 \cdot \exp(0.10 \cdot \exp(0.10 \cdot x1)) - 67.7683 \cdot \exp(0.10 \cdot \ln(|x2|)) - 4.53449 \cdot \exp(0.10 \cdot \ln(|x3|)) + 3.01688 \cdot \ln(|\ln(|x4|)|)$

Varianza Residual= [1.295259]

Coefficiente de Determinacion= [0.947098] KS= [0.543047]

TEST DE SIGNIFICANCIA

Termino independiente: 0.017654

Coefficiente de x1\$: 0.015031

Coefficiente de x2\$: 0.154354

Coefficiente de x3\$: 0.888857

Coefficiente de x4\$: 0.683238

Continúa en la página siguiente...



```

ECUACION N. 2066: y= -250.168 + 308.816*exp(0.10*exp(0.10*x1$)) -
74.1137*exp(0.10*ln(|x2$|)) - 7.88024*exp(0.10*ln(|x3$|)) +
7.70331*ln(|exp(0.10*x4$)|)

```

```

Varianza Residual= [1.286745]

```

```

Coeficiente de Determinacion= [0.947446]   KS= [0.688422]

```

TEST DE SIGNIFICANCIA

```

Termino independiente: 0.018110

```

```

Coeficiente de x1$: 0.010895

```

```

Coeficiente de x2$: 0.160916

```

```

Coeficiente de x3$: 0.821396

```

```

Coeficiente de x4$: 0.647160

```

```

-----
.
.
.

```

Sólo se han elegido 10 ecuaciones al azar de las 2076 (coinciden con un 10% de las posibles ecuaciones - 20736- que se pueden generar con el número de funciones y el orden de transformadas elegidas) que están dentro del rango elegido.

6.2.2 Ejecución con SPLUS

A continuación se discutirán los resultados con el programa SPLUS y compararlos con los obtenidos con MODELHSS cuando han sido elegidas 3 funciones y un orden de transformadas de 1. Ya que se ha obtenido con el programa MODELHSS una familia de ecuaciones no lineales y elegido la de mejor ajuste, se ejecuta el SPLUS para que obtenga



una ecuación del mismo tipo. Para ello se propone como fórmula la ecuación:

$$Y = A + B \exp(0.1 X1) + C \ln |X2| + D \ln |X3| + E \exp(0.1 X4),$$

de manera que SPLUS calcula los coeficientes para determinar la ecuación que mejor se ajusta a los datos experimentales. Esto equivale a probar directamente el caso en el que se transforman los datos iniciales aplicando las funciones correspondientes a las variables independientes originales, es decir, $\exp(x)$ y $\ln |x|$. Así se obtiene:

```

*** Linear Model ***

Call: lm(formula = Y ~ exp(0.1 * X1) + log(abs(X2),
base=exp(1)) + log(abs(X3), base=exp(1)) + exp(0.1 * X4),
data = p2, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-2.532  -0.3037  0.07515  0.5774   2.49

Coefficients:
                Value Std. Error t value Pr(>|t|)
(Intercept)  -89.6253   77.9728  -1.1494  0.2881
exp(0.1 * X1)    0.0012    0.0006   2.0244  0.0826
log(abs(X2), base=exp(1))  8.7830   16.7173   0.5254  0.6155
log(abs(X3), base=exp(1)) 18.0674   12.8574   1.4052  0.2027
exp(0.1 * X4)  -0.0050    0.0089  -0.5692  0.5870

Residual standard error: 1.595 on 7 degrees of freedom
Multiple R-Squared: 0.8961
F-statistic: 15.1 on 4 and 7 degrees of freedom, the p-value
is 0.001494

```

Destacar en este caso, como el coeficiente de determinación baja de 0.939 a 0.8961, además de no obtener ninguna de las variables ni



el coeficiente independiente significativo. La causa puede ser debida a que aquí no se está ni tipificando ni trasladando las variables iniciales independientes, de manera que a continuación vamos a realizar este cambio antes de aplicar las funciones. Las nuevas variables tipificadas y trasladadas se nombrarán $X1T$, $X2T$, $X3T$ y $X4T$:

```

*** Linear Model ***

Call: lm(formula = Y ~ exp(0.1 * X1T) + log(abs(X2T),
base=exp(1)) + log(abs(X3T), base=exp(1)) + exp(0.1 * X4T),
data = p2, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-1.526  -0.695  0.07411  0.6388  1.326

Coefficients:
                Value Std. Error t value Pr(>|t|)
(Intercept)   -31.3625    9.2983  -3.3729  0.0119
exp(0.1 * X1T)  36.2223   10.4119   3.4789  0.0103
log(abs(X2T), base=exp(1)) -8.6061    5.4240  -1.5867  0.1566
log(abs(X3T), base=exp(1)) -1.1438    3.8789  -0.2949  0.7766
exp(0.1 * X4T)   5.3805   10.4005   0.5173  0.6209

Residual standard error: 1.133 on 7 degrees of freedom
Multiple R-Squared: 0.9475
F-statistic: 31.6 on 4 and 7 degrees of freedom, the p-value
is 0.001428

```

Ahora sí que coinciden los resultados porque según resumimos:

- Ecuación obtenida:

$$y = -31.3625 + 36.2223 \exp(0.1 X1T) - 8.6061 \ln |X2T| - 1.1438 \ln |X3T| + 5.3805 \exp(0.1 X4T)$$



- Coeficiente de determinación: 0.9475
- Variables significativas: Término independiente y $X1T$.

Además, mejoramos el coeficiente de determinación con respecto a la no tipificación y traslación y ganamos los dos coeficientes como significativos.

6.2.3 Ejecución con SPSS

De la misma forma que se han llevado a cabo las ejecuciones con SPLUS, vamos a hacer las mismas pruebas con el programa estadístico SPSS: primero creando unas nuevas variables como fruto de la aplicación de las funciones, llamadas $EXPX1$, $LNX2$, $LNX3$ y $EXPX2$; y segundo, también creando unas nuevas variables pero en este caso aplicando las funciones a las tipificadas y trasladadas, llamadas $EXPX1T$, $LNX2T$, $LNX3T$ y $EXPX2T$.

En el primer caso obtenemos lo siguiente:

VARIABLES INTRODUCIDAS/ELIMINADAS^b

Modelo	VARIABLES INTRODUCIDAS	VARIABLES ELIMINADAS	Método
1	EXPX1, LNX2, LNX3, EXPX4 ^a	,	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Y



Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,947 ^a	,896	,837	1,5948

a. Variables predictoras: (Constante), EXPX1, LNX2, LNX3, EXPX4

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	153,586	4	38,396	15,097	,001 ^a
	Residual	17,804	7	2,543		
	Total	171,389	11			

a. Variables predictoras: (Constante), EXPX1, LNX2, LNX3, EXPX4

b. Variable dependiente: Y

Coeficientes^a

Modelo		Coefic. no estand.		Coefic. estand.	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-89,625	77,973		-1,149	,288
	EXPX1	1,165E-03	,001	,590	2,024	,083
	LNX2	8,783	16,717	,196	,525	,616
	LNX3	18,067	12,857	,453	1,405	,203
	EXPX4	-5,043E-03	,009	-,252	-,569	,587

b. Variable dependiente: Y

En efecto, coinciden los resultados obtenidos con el SPLUS cuando también se aplicaban las funciones a las variables independientes origi-



nales.

Vamos ahora, entonces, a probar con la tipificación y traslación de estas variables y luego aplicaremos las funciones:

VARIABLES INTRODUCIDAS/ELIMINADAS^b

Modelo	VARIABLES INTRODUCIDAS	VARIABLES ELIMINADAS	Método
1	EXPX1T, LNX2T, LNX3T, EXPX4T ^a		, Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Y

RESUMEN DEL MODELO

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,973 ^a	,948	,918	1,1334

a. Variables predictoras: (Constante), EXPX1T, LNX2T, LNX3T, EXPX4T



Aplicación de MODELHSS a una modelización medioambiental C6

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	162,397	4	40,599	31,604	,000 ^a
	Residual	8,992	7	1,285		
	Total	171,389	11			

a. Variables predictoras: (Constante), EXPX1T, LNX2T, LNX3T, EXPX4T

b. Variable dependiente: Y

Coefficientes^a

Modelo		Coefic. no estand.		Coefic. estand.	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-31,362	9,298		-3,373	,012
	EXPX1T	36,222	10,412	1,368	3,479	,010
	LNX2T	-8,606	5,424	-,559	-1,587	,157
	LNX3T	-1,144	3,897	-,074	-,295	,777
	EXPX4T	5,380	10,400	,505	,517	,621

b. Variable dependiente: Y

Es evidente que los resultados coinciden en los tres casos: los obtenidos por MODELHSS, los de SPLUS y los de SPSS.

6.3 Aplicación de MODELHSS a una modelización medioambiental

En este apartado se presenta la modelización matemática de un fenómeno a partir de la ejecución del programa para obtener ecuaciones



Aplicación de MODELHSS a una modelización medioambiental C6

que ajusten a los datos disponibles. En este caso además se dispone de gran cantidad de datos experimentales. Destacar que cada tabla de datos contiene más de 1000 tuplas y como mínimo 4 variables independientes, de modo que el número de operaciones aumenta considerablemente respecto a los ejemplos anteriores hasta el extremo, incluso, con ciertos ficheros de datos, de tener que lanzar el programa en una estación de trabajo, pues en un PC se agotaba la memoria. Los resultados obtenidos además han sido expuestos en el artículo "A new methodology for modelling highly structured systems" ([5]).

Se presenta un problema medioambiental relacionado con la polución del aire en la provincia mediterránea de Castellón, donde se analizan las siguientes variables:

1. O_3 → Se corresponde con la cantidad de ozono medido en $\mu\text{g}/\text{m}^3$.
2. VEL → Velocidad del viento (m/s).
3. TEM → Temperatura ($^{\circ}\text{C}$).
4. PLU → Lluvia caída (l/m^2).
5. RAD → Radiación solar (w/m^2).
6. PSS → Partículas sólidas suspendidas en el aire ($\mu\text{ g}/\text{m}^3$).
7. NO → Monóxido de nitrógeno ($\mu\text{ g}/\text{m}^3$).
8. NO₂ → Dióxido de nitrógeno ($\mu\text{ g}/\text{m}^3$).

De cada una de las variables se han tomado medidas cada 15 minutos durante todo un año, 1997. Las variables han sido medidas utilizando



Aplicación de MODELHSS a una modelización medioambiental C6

estaciones *sampling* automáticas que han sido localizadas en 2 áreas importantes dentro de la provincia. Una de las áreas, denominada A, está situada en el mar y la otra, llamada B, en la tierra, cerca de un valle rodeado de montañas de alrededor de 500 metros de altura por encima del nivel del mar.

El análisis de los datos de la polución del aire ha sido principalmente tratado en ciertas bibliografías ([2, 8, 9]), utilizando técnicas de regresión lineal múltiple o simple ([17, 37]). En todos los casos aportados, las ecuaciones de regresión se basaban en las variables independientes originales sin aplicar transformadas.

Se han, además, considerado las siguientes cuestiones:

1. Los efectos de la radiación solar afecta a los modelos propuestos de forma distinta dependiendo de los bajos o altos valores que pueda tomar. De este modo, se ha decidido definir los modelos distinguiéndolos entre el día y la noche.
2. Después de un análisis preliminar se ha observado, claramente, un comportamiento cíclico entre la mayoría de las variables consideradas. Los ciclos se definen en períodos de seis meses. Consecuentemente, se han tomado únicamente los seis primeros meses, de enero a febrero de 1997.
3. El área en la que las estaciones están localizadas es realmente importante a la hora de encontrar diferencias en el comportamiento de la polución. Por ello, se han separado los modelos en cuanto al área donde se encontraban.
4. Se han analizado los siguientes modelos para predecir el ozono (A,B= área, N= noche, D= día):



Aplicación de MODELHSS a una modelización medioambiental C6

(a) $O_3^{A,N} = f(\text{VEL}, \text{TEM}, \text{PLU}, \text{RAD})$

(b) $O_3^{A,D} = f(\text{VEL}, \text{TEM}, \text{PLU}, \text{RAD})$

(c) $O_3^{B,N} = f(\text{VEL}, \text{TEM}, \text{PLU}, \text{RAD})$

(d) $O_3^{B,D} = f(\text{VEL}, \text{TEM}, \text{PLU}, \text{RAD})$

5. Los siguientes modelos han sido utilizados para predecir las partículas sólidas suspendidas en el aire (A,B= área, N= noche, D= día):

(a) $PSS^{A,N} = f(\text{NO}, \text{NO}_2, \text{VEL}, \text{TEM})$

(b) $PSS^{A,D} = f(\text{NO}, \text{NO}_2, \text{VEL}, \text{TEM})$

(c) $PSS^{B,N} = f(\text{NO}, \text{NO}_2, \text{VEL}, \text{TEM})$

(d) $PSS^{B,D} = f(\text{NO}, \text{NO}_2, \text{VEL}, \text{TEM})$

La mayoría de los modelos considerados en las distintas bibliografías y utilizados para predecir el ozono en términos de temperatura, radiación solar y velocidad del viento presentan coeficientes de determinación menores que el 40% ([2, 8, 9]). El objetivo es aplicar la metodología implementada para analizar estos modelos y proponer nuevos, de manera que se mejoren los coeficientes de determinación hasta el momento obtenidos. Destacar además, que se busca una ecuación para predecir de una manera más ajustada los valores del ozono en función de otras variables medidas bajo la hipótesis de que no exista una autocorrelación o correlación cruzada. Esto también se puede realizar utilizando análisis de series en el tiempo, que aquí no se van a tratar.



6.3.1 Ejecución con MODELHSS

Para este análisis en concreto, se han elegido como funciones transformadas de orden 1 las siguientes: x , x^2 , $\frac{1}{x}$, $\text{sen}(x)$, $\exp(x)$, $\exp(-x)$, $\arctg(x)$ y $\ln|x|$.

Para predecir el ozono en el área A, se analizaron las ecuaciones correspondientes para el día y la noche tomando un modelo para los seis meses en conjunto, desde enero a febrero, y, posteriormente, seis modelos más correspondientes a cada mes.

En la ejecución del programa se generaban alrededor de 4096 ecuaciones, de modo que se ha elegido la mejor ecuación, aquella con el coeficiente de determinación más alto. Si se hubiera optado por almacenar todas las ecuaciones, sólo habría que tener en cuenta que sería de un gran tamaño. No obstante, podríamos haber elegido en la ejecución la obtención de aquellas con coeficiente en un cierto intervalo, reduciendo posiblemente el tamaño del fichero.

En las ecuaciones de la 6.5 a la 6.11 figuran los resultados referentes a las medidas tomadas durante el día.

$$\begin{aligned}
 O_3^{A,D} = & 6.358873 - 1060.72 \exp(-0.551626 VEL - 2.72358) + \\
 & 44.4784 \ln|0.160217 TEM + 1.32271| + \\
 & 482.509 \exp(-0.309559 PLU - 4.0312) + \\
 & 4.44181(0.00405371 * RAD + 2.67821)
 \end{aligned} \tag{6.5}$$



Aplicación de MODELHSS a una modelización medioambiental C6

$$O_3^{A,D}(ENE) = 75.951256 - 190.476/(0.694795 VEL + 2.63743) + \\ 0.634025(0.335051 TEM + 0.217592)^2 - \\ 373.846 \exp(-23.9049 PLU - 3.81287) + \\ 0.26569(0.008033 RAD + 2.80456)^2 \quad (6.6)$$

$$O_3^{A,D}(FEB) = -24.063227 - 932.129 \exp(-0.851243 VEL - 2.58273) + \\ 1.34467 \arctg(0.285478 TEM + 0.189076) - \\ 0.003744(104.641 PLU + 3.93901)^2 - \\ 0.207726(0.005138 RAD + 2.58421)^2 \quad (6.7)$$

$$O_3^{A,D}(MAR) = 108.45 - 1957.06 \exp(-0.342174 VEL - 3.35433) + \\ 54.9861(0.107897 TEM + 2.42998) - \\ 53.2049(0.133364 PLU + 4.0754) - \\ 0.025965 \exp(0.004273 RAD + 2.50554) \quad (6.8)$$

$$O_3^{A,D}(ABR) = -23.536412 - 662.72 \exp(-0.720062 VEL - 2.17745) + \\ 2.09628 \arctg(0.226566 TEM + 0.188636) - \\ 603.052 \operatorname{sen}(0.384861 PLU + 4.02412) + \\ 140.276 \operatorname{sen}(0.003864 RAD + 2.63392) \quad (6.9)$$

$$O_3^{A,D}(MAY) = 239.578682 - 581.343 \exp(-0.617834 VEL + 2.12281) + \\ 32.9301 \ln |0.331226 TEM - 2.39959| - \\ 2.46229 \arctg(12.119 PLU + 3.93926) + \\ 0.393999(0.003809 RAD + 2.50267)^2 \quad (6.10)$$



Aplicación de MODELHSS a una modelización medioambiental C6

$$\begin{aligned}
 O_3^{A,D}(JUN) = & 68.9098 - 652.507 \exp(-0.74387 VEL - 2.15319) + \\
 & 1.35004(0.238933 TEM - 1.28992)^2 - \\
 & 343.563 \exp(-18.359 PLU - 3.93164) + \quad (6.11) \\
 & 86.637 \exp(-0.003805 RAD - 2.56267)
 \end{aligned}$$

En la tabla 6.4 se representan los distintos coeficientes de determinación (R^2) y de correlación (R) obtenidos para cada una de estas ecuaciones.

Ecuación	R^2	R
$O_3^{A,D}$	0.482870	0.694889
$O_3^{A,D}(ENE)$	0.449784	0.670659
$O_3^{A,D}(FEB)$	0.444729	0.666880
$O_3^{A,D}(MAR)$	0.550555	0.741994
$O_3^{A,D}(ABR)$	0.444967	0.667058
$O_3^{A,D}(MAY)$	0.518699	0.720208
$O_3^{A,D}(JUN)$	0.449527	0.670468

Tabla 6.4: Coeficientes de determinación R^2 y de correlación del ozono en el área A durante el día.

En las ecuaciones de la 6.12 a la 6.18 figuran los resultados referentes



Aplicación de MODELHSS a una modelización medioambiental C6

a la noche.

$$\begin{aligned}
 O_3^{A,N} = & -73.621447 + 976.313 \operatorname{sen}(0.88588 \text{VEL} + 2.92506) + \\
 & 12.9759(0.210532 \text{TEM} + 1.3633) - \\
 & 789.688 \exp(-15.2588 \text{PLU} - 3.89689) - \\
 & 0.191083 \exp(0.137189 \text{RAD} + 3.10695)
 \end{aligned} \quad (6.12)$$

$$\begin{aligned}
 O_3^{A,N}(\text{ENE}) = & -49.185252 + 58.8197 \ln|0.897031 \text{VEL} + 2.79127| + \\
 & 0.0624276 \exp(0.357057 \text{TEM} + 0.796449) - \\
 & 767.276 \exp(-10.3709 \text{PLU} - 3.81435) - \\
 & 1.53993/(0.936724 \text{RAD} - 9.95098)
 \end{aligned} \quad (6.13)$$

$$\begin{aligned}
 O_3^{A,N}(\text{FEB}) = & -7.769970 + 0.000573 \exp(1.14393 \text{VEL} + 3.01953) + \\
 & 1.35011(0.318672 \text{TEM} + 1.04122)^2 - \\
 & 0.591119(62.7618 \text{PLU} + 3.90702) - \\
 & 0.081803(0.555506 \text{RAD} - 4.07634)^2
 \end{aligned} \quad (6.14)$$

$$\begin{aligned}
 O_3^{A,N}(\text{MAR}) = & -59.202111 + 10.2706(1.15946 \text{VEL} + 2.89539) + \\
 & 2.14495(0.343782 \text{TEM} + 0.305561)^2 + \\
 & 1.58778(51.5524 \text{PLU} + 3.92399) - \\
 & 97.4949 \exp(-0.150011 \text{RAD} - 3.32062)
 \end{aligned} \quad (6.15)$$

$$\begin{aligned}
 O_3^{A,N}(\text{ABR}) = & -140.255141 + 117.435 \ln|0.659617 \text{VEL} + 2.99301| + \\
 & 758.834 \operatorname{sen}(0.33148 \text{TEM} - 0.360182) - \\
 & 801.513 \exp(-20.1699 \text{PLU} - 3.81908) - \\
 & 743.159 \exp(-0.441192 \text{RAD} - 3.78976)
 \end{aligned} \quad (6.16)$$



Aplicación de MODELHSS a una modelización medioambiental C6

$$\begin{aligned}
 O_3^{A,N}(MAY) = & 107.784441 - 424.175/(0.845167 VEL + 2.80017) + \\
 & 38.9571 \ln |0.336753 TEM - 1.48897| + \\
 & 0.0363(9.03937 PLU + 3.9363)^2 + \\
 & 0.093186(0.38929 RAD + 3.76071)^2
 \end{aligned} \tag{6.17}$$

$$\begin{aligned}
 O_3^{A,N}(JUN) = & -29.872448 + 707.071 \operatorname{sen}(1.07613 VEL + 2.71615) + \\
 & 2.40767(0.259854 TEM - 0.937019)^2 - \\
 & 601.968 \exp(-82.8847 PLU - 3.93272) \\
 & -579.027 \exp(-0.400378 RAD - 3.76601)
 \end{aligned} \tag{6.18}$$

En la tabla 6.5 se representan los distintos coeficientes de determinación (R^2) y de correlación (R) obtenidos para cada una de estas ecuaciones.

Como se ve, hay una clara diferencia entre los resultados referentes a las medidas de día y de la noche. En cualquier caso, los coeficientes de determinación obtenidos se pueden considerar bastante buenos y altos, si tenemos en cuenta el tamaño considerado de los ficheros, según refleja la tabla 6.6.

Sin embargo, las ecuaciones obtenidas no satisfacen la condición de que los residuos se ajusten a una distribución normal, aunque sus medias sí que valen 0. Si analizamos este modelo mes a mes, encontramos diferencias entre ellos. Estas diferencias pueden ser debidas a las distintas estaciones del año (Ecuaciones 6.5- 6.18). Es muy interesante destacar que en algunos de los meses analizados se obtienen unas ecuaciones con un alto coeficiente de determinación (mayor del 62%).



Aplicación de MODELHSS a una modelización medioambiental C6

Ecuación	R ²	R
O ₃ ^{A,N}	0.519852	0.721008
O ₃ ^{A,N} (ENE)	0.605012	0.777825
O ₃ ^{A,N} (FEB)	0.342933	0.585605
O ₃ ^{A,N} (MAR)	0.546391	0.739183
O ₃ ^{A,N} (ABR)	0.630576	0.794088
O ₃ ^{A,N} (MAY)	0.560705	0.748802
O ₃ ^{A,N} (JUN)	0.420136	0.648179

Tabla 6.5: Coeficientes de determinación R² y de correlación del ozono en el área A durante la noche.

Para predecir el ozono en el área B también se analizan las correspondientes ecuaciones para el día y la noche, donde primero se modelizan los seis meses juntos y luego por separado. Para resumir, se van a presentar únicamente los valores obtenidos para el conjunto de los 6 meses, como puede observarse en las ecuaciones 6.19 y 6.20. Y los coeficientes de determinación se presentan en la tabla 6.7.

$$\begin{aligned}
 O_3^{B,D} = & -59.381427 + 36.5922 \ln |0.856094 VEL + 2.79869| + \\
 & 42.7557 \ln |0.223283 TEM + 1.04125| + \\
 & 0.147759(0.288293 RAD + 1.82055)^2
 \end{aligned} \tag{6.19}$$

$$\begin{aligned}
 O_3^{B,N} = & -5.304502 - 378.891 \exp(-0.720632 VEL - 2.37296) + \\
 & 50.9717 \ln |0.180004 TEM + 0.759425| + \\
 & 310.289 \operatorname{sen}(0.00394257 RAD + 2.71549)
 \end{aligned} \tag{6.20}$$



Aplicación de MODELHSS a una modelización medioambiental C6

Fichero	Tuplas de datos	Fichero	Tuplas de datos
$O_3^{A,D}$	8820	$O_3^{A,N}$	8602
$O_3^{A,D}$ (ENE)	1188	$O_3^{A,N}$ (ENE)	1782
$O_3^{A,D}$ (FEB)	1201	$O_3^{A,N}$ (FEB)	1485
$O_3^{A,D}$ (MAR)	1413	$O_3^{A,N}$ (MAR)	1560
$O_3^{A,D}$ (ABR)	1479	$O_3^{A,N}$ (ABR)	1427
$O_3^{A,D}$ (MAY)	1656	$O_3^{A,N}$ (MAY)	1334
$O_3^{A,D}$ (JUN)	1665	$O_3^{A,N}$ (JUN)	1232

Tabla 6.6: Tuplas de datos de cada fichero del ozono en el área A durante el día y la noche.

Los correspondientes coeficientes de determinación (Tabla 6.7) son más pequeños que los obtenidos en el área A. Esto lleva a intuir que para predecir el ozono en el área B se necesitan especificar y medir más variables. En este caso, no se tiene la suficiente información para analizar el comportamiento del ozono, únicamente en términos de la velocidad del viento, las radiaciones y la temperatura. Denotar que la pluriometría no se ha considerado signficante en este modelo. De hecho, el área B está localizada en una zona industrial (industria cerámica) y la polución es causada principalmente por otras variables químicas y componentes del aire externos.

Vamos a tomar el segundo factor a estudiar que sería la cantidad de partículas sólidas suspendidas en el aire (PSS). Esta variable se mide en términos del monóxido de nitrógeno, el dióxido de nitrógeno, la velocidad del viento y la temperatura. En el ajuste PSS en ambas áreas, A y B, si analizamos estas ecuaciones (Ecuaciones 6.21- 6.24),



Aplicación de MODELHSS a una modelización medioambiental C6

Ecuación	R ²	R
O ₃ ^{B,D}	0.308890	0.555779
O ₃ ^{B,N}	0.486082	0.697196

Tabla 6.7: Coeficientes de determinación R² y de correlación del ozono en el área B durante el día y la noche.

la temperatura es un factor común, expresado en forma exponencial en todas las ecuaciones. Aparte de esto, las ecuaciones son distintas entre el día y la noche, incluso en la misma área. Destacar que PSS está mejor ajustada en el área A durante la noche, al igual que en el área B. Esto es principalmente consecuencia de la localización de estas áreas. Por ejemplo, en el área B, PSS está principalmente influida por industrias que se dedican a la fabricación de cerámica durante el día. Sin embargo, en cualquier caso, se deberían considerar más variables e información para poder ajustar mejor PSS en estos casos. Sus correspondientes coeficientes de determinación se pueden ver en la tabla 6.8

$$\begin{aligned}
 PSS^{A,D} = & -27.526930 + 1538.35 \operatorname{sen}(0.0520916 NO_2 + 2.96637) + \\
 & 158.442 \operatorname{sen}(0.0423282 NO + 3.55021) - \\
 & 0.059969(0.551626 VEL + 2.72358)^2 - \\
 & 160.547 \exp(-0.160217 TEM - 1.32271)
 \end{aligned} \tag{6.21}$$



Aplicación de MODELHSS a una modelización medioambiental C6

$$\begin{aligned}
 PSS^{A,N} = & -601.481519 + 2.24161(0.0436817 NO_2 + 2.55374)^2 + \\
 & 8.52463 \arctg(0.0338888 NO + 3.34648) + \\
 & 14.1687 \ln |0.88588 VEL + 2.92506| - \\
 & 149.432 \exp(-0.210532 * TEM - 1.3633)
 \end{aligned} \tag{6.22}$$

$$\begin{aligned}
 PSS^{B,D} = & 54.490469 + 187.801 \exp(-0.111146 NO_2 - 2.80984) + \\
 & 0.223778(0.0509734 NO + 3.47987)^2 + \\
 & 0.00373973 \exp(0.720632 VEL + 2.37296) - \\
 & 136.382 \exp(-0.180004 * TEM - 0.759425)
 \end{aligned} \tag{6.23}$$

$$\begin{aligned}
 PSS^{B,N} = & 0.025820 + 157.518 \text{sen}(0.128772 NO_2 + 2.85842) + \\
 & 20.1926 \ln |0.0609849 NO + 3.37027| + \\
 & 135.572 \exp(-0.856094 VEL - 2.79869) - \\
 & 127.224 * \exp(-0.223283 TEM - 1.04125)
 \end{aligned} \tag{6.24}$$

Ecuación	R ²	R
PSS ^{A,D}	0.199357	0.446494
PSS ^{A,N}	0.311425	0.558055
PSS ^{B,D}	0.030333	0.174164
PSS ^{B,N}	0.079248	0.281510

Tabla 6.8: Coeficientes de determinación R² y de correlación de PSS en las áreas A y B durante el día y la noche.



6.3.2 1ª Ejecución con SPLUS

Se ha elegido un fichero al azar para comparar los resultados que se obtienen si se utiliza un programa estadístico, como puede ser el SPLUS, para ajustar una ecuación a esos datos.

El fichero para la prueba ha sido el fichero de datos correspondiente al ajuste del ozono en el área A de noche del mes de Junio. La tabla de datos correspondiente se adjunta en el Capítulo 7 de Anexo. Ha sido llamado $O_3^{A,N}(JUN)$ con 1232 tuplas de datos cuya ecuación obtenida con MODELHSS se puede ver en 6.18 y su coeficiente de determinación viene expresado en la tabla 6.5 (0.420136).

En cuanto a la significancia de los coeficientes lo podemos ver en la siguiente tabla 6.9:

Coficiente	Significancia
Independiente	0.006283
VEL	0.000000
TEM	0.000000
PLU	0.233078
RAD	0.000011

Tabla 6.9: Significancia de los coeficientes de la ecuación obtenida con MODELHSS.

Únicamente la variable PLU no es significativa.

La utilización de SPLUS a partir de la tabla de datos conduce a



Aplicación de MODELHSS a una modelización medioambiental C6

ejecutar los ajustes que tiene programados. Para ello comenzar con linealidad y continuar con la obtención de ecuaciones no lineales previa propuesta de las mismas.

Cuando se ha introducido la tabla de datos en el SPLUS y se ha pedido un análisis por regresión lineal, ya nos hemos encontrado un problema: si no hubiéramos ejecutado el programa MODELHSS, no sabríamos que expresión tendría la ecuación que más se ajusta a los datos, de modo que tendríamos que haber hecho distintas pruebas para ver si se puede o no conseguir acercarse o superar al coeficiente de determinación obtenido con el MODELHSS. De este modo, se ha elegido una expresión de la ecuación similar a la obtenida con MODELHSS para comparar los resultados, pero sin tipificar ni trasladar las variables independientes.

Los resultados con SPLUS se muestran a continuación:



Aplicación de MODELHSS a una modelización medioambiental C6

Universitat d'Alacant
Universidad de Alicante

```

*** Linear Model ***

Call: lm(formula = O3 ~ sin((VEL*3.141593)/180) + TEM^2 +
exp(-PLU) + exp(-RAD), data = p1, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-80.51  -17.17   -3.348   14.78   89.46

Coefficients:
                Value  Std. Error t value Pr(>|t|)
(Intercept)    41.0166    63.0787   0.6502  0.5157
sin((VEL*3.141593)/180)  777.8804    44.0473  17.6601  0.0000
      I(TEM^2)     0.1196     0.0067  17.9219  0.0000
      exp(-PLU)  -53.1228    62.9937  -0.8433  0.3992
      exp(-RAD)  -13.9246     2.8904  -4.8176  0.0000

Residual standard error: 24.26 on 1227 degrees of freedom
Multiple R-Squared:  0.4119
F-statistic: 214.9 on 4 and 1227 degrees of freedom, the
p-value is 0

```

De modo que de esta tabla podemos extraer las siguientes conclusiones:

- Ecuación obtenida¹:

$$\begin{aligned}
 O3 &= 41.0166 + 777.8804 \operatorname{sen}\left(\frac{VEL * 3.141593}{180}\right) + \\
 &+ 0.1196 TEM^2 - 53.1228 \exp(-PLU) - \\
 &- 13.9246 \exp(-RAD)
 \end{aligned} \tag{6.25}$$

¹La función *seno* aparece multiplicada por π y dividida por 180 para pasarla a grados



Aplicación de MODELHSS a una modelización medioambiental C6

- Coeficiente de determinación:

0.4119

- Variables significativas: solamente VEL, TEM y RAD.

Como se observa, el coeficiente de determinación disminuye en una centésima y, además de la variable PLU, el coeficiente independiente tampoco es significativo, debido en tal caso a la no tipificación y traslación de las variables.

Construimos unas nuevas columnas de datos en el ficheros de manera que sean el resultado de aplicar las funciones *seno*, *cuadrado* y *exponenciales negativas* a las variables originales, tipificadas y trasladadas. Con estas nuevas variables independientes creadas llamadas *SEVEL*, *CUATEM*, *EXPPLU* y *EXPRAD*, obtenemos los siguientes resultados:



Aplicación de MODELHSS a una modelización medioambiental C6

Universitat d'Alacant
Universidad de Alicante

```

*** Linear Model ***

Call: lm(formula = O3 ~ SENVEL + CUATEM + EXPPLU + EXPRAD,
data = gnoo3j, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-85.35  -16.77   -2.908   14.33   88.45

Coefficients:
              Value  Std. Error t value Pr(>|t|)
(Intercept) -29.8752   10.9130  -2.7376  0.0063
    SENVEL    707.1073   40.9482  17.2683  0.0000
    CUATEM     2.4077    0.1296  18.5741  0.0000
    EXPPLU  -602.5358  505.0347  -1.1931  0.2331
    EXPRAD  -579.0272  131.1922  -4.4136  0.0000

Residual standard error: 24.1 on 1227 degrees of freedom
Multiple R-Squared: 0.4201
F-statistic: 222.3 on 4 and 1227 degrees of freedom, the
p-value is 0

```

Conviene observar que mediante la tipificación y traslación se consigue mejorar un poco el coeficiente de determinación (se obtiene el mismo que en el caso del *bf it* ModelHSS, además de la misma ecuación y significancia de las variables), aunque la significancia de las variables, en este caso (con respecto al caso que no se tipifica ni traslada), no varía:



Aplicación de MODELHSS a una modelización medioambiental C6

- Ecuación obtenida:

$$\begin{aligned}
 O3 &= -29.8752 + 707.1073 \textit{SENVEL} + 2.4077 \textit{CUATEM} - \\
 &\quad - 602.5358 \textit{EXPPLU} - 579.0272 \textit{EXPRAD} \quad (6.26)
 \end{aligned}$$

- Coeficiente de determinación:

0.4201

- Variables NO significativas: EXPPLU.

6.3.3 1ª Ejecución con SPSS

Se ha elegido el mismo fichero de datos para ajustarlo con el SPSS. Como se busca una ecuación del mismo tipo que la obtenida con el MODELHSS, cuando intentamos hacer un análisis de regresión lineal a partir de los datos iniciales nos encontramos con el problema de que tenemos que crear nuevas variables de datos. Esto es debido a que el análisis de regresión lineal en SPSS sólo permite encontrar ecuaciones del tipo $y = a_0 + a_1x_1 + a_2x_2 + \dots$, es decir, las variables x_1, x_2, \dots no pueden ser funciones de variables. Con lo cual, creamos 4 variables más en la tabla de datos, *SENVEL*, *CUATEM*, *EXPPLU* y *EXPRAD*, las cuales corresponden al seno de VEL, el cuadrado de TEM y las exponenciales negativas de PLU y RAD respectivamente. A estas variables se les aplica las funciones correspondientes quedando creadas 4 columnas más de datos. Al modelizar, las variables independientes serán las nuevas creadas y la dependiente seguirá siendo *O3*.

A continuación se muestran los resultados:



Aplicación de MODELHSS a una modelización medioambiental C6

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1 7	EXPRAD, EXPPLU, CUATEM, SENVEL ^a		Introducir

- a. Todas las variables solicitadas introducidas
b. Variable dependiente: O3

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,642 ^a	,412	,410	24,2649

- a. Variables predictoras: (Constante), EXPRAD, EXPLU, CUATEM, SENVEL

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	506071,621	4	126517,905	214,879	,000 ^a
	Residual	722440,220	1227	588,786		
	Total	1228511,841	1231			

- a. Variables predictoras: (Constante), EXPRAD, EXPLU, CUATEM, SENVEL
b. Variable dependiente: O3



Aplicación de MODELHSS a una modelización medioambiental C6

Modelo		Coeficientes ^a			t	Sig.
		Coefic. no estand.		Coefic. estand.		
		B	Error típ.	Beta		
1	(Constante)	41,017	63,079		,650	,516
	SENVEL	777,881	44,047	,399	17,660	,000
	CUATEM	,120	,007	,404	17,922	,000
	EXPPLU	-53,123	62,994	-,018	-,843	,399
	EXPRAD	-13,925	2,890	-,106	-4,818	,000

b. Variable dependiente: O₃

De aquí también podemos sacar las siguientes conclusiones:

- Ecuación obtenida¹:

$$\begin{aligned}
 O_3 = & 41.017 + 777.881 \text{ SENVEL} + 0.120 \text{ CUATEM} - \\
 & - 53.123 \text{ EXPPLU} - 13.925 \text{ EXPRAD} \quad (6.27)
 \end{aligned}$$

- Coeficiente de determinación:

0.412

- Variables significativas: solamente SENVEL, CUATEM y EXPRAD.

Se comprueba fácilmente que los resultados obtenidos por ambos programas estadísticos coinciden en gran medida según muestran las ecuaciones 6.25 y 6.27 y sus respectivos coeficientes de determinación.

Si ahora se **tipifica y traslada** cada variable independiente, es decir se crea una nueva columna de datos para poder tipificar y trasladar

¹La función *seno* recibe el parámetro en grados



Aplicación de MODELHSS a una modelización medioambiental C6

las iniciales y a partir de esta nueva columna se aplican las funciones correspondientes, los resultados serán los que a continuación se presentan:

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	EXPRAD, EXPPLU, CUATEM, SENVEL ^a		Introducir

- a. Todas las variables solicitadas introducidas
- b. Variable dependiente: O3

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,648 ^a	,420	,418	24,0952

- a. Variables predictoras: (Constante), EXPRAD, EXPLU, CUATEM, SENVEL

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	516141,662	4	129035,416	222,253	,000 ^a
	Residual	712370,179	1227	580,579		
	Total	1228511,841	1231			

- a. Variables predictoras: (Constante), EXPRAD, EXPPLU, CUATEM, SENVEL
- b. Variable dependiente: O3



Aplicación de MODELHSS a una modelización medioambiental C6

Modelo		Coeficientes ^a			t	Sig.
		Coefic. no estand.		Coefic. estand.		
		B	Error típ.	Beta		
1	(Constante)	-29,875	10,913		-2,738	,006
	SEVEL	707,107	40,948	,399	17,268	,000
	CUATEM	2,408	,130	,418	18,574	,000
	EXPPLU	-602,536	505,035	-,026	-1,193	,233
	EXPRAD	-579,027	131,192	-,096	-4,414	,000

b. Variable dependiente: O3

Ahora comprobamos que al tipificar y trasladar aumenta el coeficiente de determinación (pasa de 0.410 a ser 0.420). La causa puede ser la gran cantidad de datos que posee el fichero y por lo tanto hay despreciamientos de decimales. De este modo, vuelven a coincidir los datos con los obtenidos con el *bf it* ModelHSS, incluso la significancia, destacando que el término independiente, en este caso, pasa a ser significativo respecto a la no tipificación y traslación.

6.3.4 2^a Ejecución con SPLUS

Un caso que nos ha parecido curioso destacar es el que se refiere al fichero de datos correspondiente al ajuste del ozono en el área A de noche pero en el mes de Mayo. Se le ha nombrado $O_3^{A,N}(MAY)$ con 1334 tuplas de datos. La ecuación obtenida con MODELHSS se puede ver en 6.17 y su coeficiente de determinación viene expresado en la tabla 6.5 (0.560705). Todas las variables resultan ser significativas, a excepción de *RAD*\$.

Nos encontramos con el mismo problema anterior: en el análisis por regresión lineal, si no hubiéramos ejecutado el programa MODELHSS,



Aplicación de MODELHSS a una modelización medioambiental C6

no sabríamos qué expresión tendría la ecuación que más se ajusta a los datos. Por ello, se ha elegido una expresión de la ecuación similar a la obtenida con MODELHSS para comparar los resultados.

Ha surgido un problema cuando en el análisis de regresión lineal se ha intentado especificar la ecuación resultante, es decir que la variable dependiente $O3$ sea función de la inversa de VEL , el neperiano del valor absoluto de TEM , el cuadrado de PLU y el cuadrado de RAD , específicamente, en términos de SPLUS sería:

$$O3 \sim 1/VEL + \log(\text{abs}(TEM), \text{base}=\exp(1)) + PLU^2 + RAD^2$$

Al aceptar esta dependencia el SPLUS devuelve el siguiente mensaje de error:

```
Error in .Fortran ("dqrls",: subroutine dqrls: 9 inf
value(s) in argument 1
Dumped
```

El problema viene causado por intentar calcular el neperiano de 0 y algunas divisiones por 0, de modo que no puede llegar a ejecutar el método de los mínimos cuadrados y por tanto no muestra ningún resultado.

De modo que podemos hacer 3 cosas:

- O crear unas variables nuevas como resultado de aplicar las funciones a las variables originales, pero nos encontraremos con el



Aplicación de MODELHSS a una modelización medioambiental C6

mismo problema anterior: tendremos errores por intentar aplicar funciones a ciertos datos incompatibles,

- o crear también unas variables nuevas pero que sean el fruto de tipificación y traslación de las originales y posteriormente en el análisis de regresión especificar la expresión de la ecuación que busquemos,
- o bien, directamente crear las variables tipificadas y trasladadas a las que se les aplica las funciones correspondientes, de modo que en el análisis de regresión la expresión de la ecuación será lineal en cuanto a las variables nuevas.

Probamos directamente la segunda opción, donde las variables nuevas se nombran: *TIPVEL*, *TIPTEM*, *TIPPLU* Y *TIPRAD*. Los resultados obtenidos son los siguientes:



Aplicación de MODELHSS a una modelización medioambiental C6

```

*** Linear Model ***

Call: lm(formula = O3 ~ 1/TIPVEL + log(abs(TIPTEM), base=exp(1)) +
  TIPPLU^2 + TIPRAD^2, data = gnoo3y, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-86.75  -20.45   -1.176    19   64.09

Coefficients:
                Value Std. Error t value Pr(>|t|)
(Intercept)   -96.8861    4.3288  -22.3816  0.0000
TIPVEL %n% 1   21.5706    0.7506   28.7369  0.0000
log(abs(TIPTEM), base=exp(1))  43.6929    2.7630   15.8136  0.0000
I(TIPPLU^2)    0.0380    0.0189    2.0049  0.0452
I(TIPRAD^2)    0.1064    0.0570    1.8676  0.0620

Redisual standard error: 26.38 on 1329 degrees of freedom
Multiple R-Squared: 0.5169
F-statistic: 355.5 on 4 and 1329 degrees of freedom,
the p-value is 0

```

En este caso, el coeficiente de determinación, 0.5169, es más pequeño que el obtenido con MODELHSS, de manera que también difiere la ecuación resultante.

Vamos a probar la tercera opción: crear nuevas variables con las funciones correspondientes aplicadas al resultado de tipificar y trasladar las variables independientes originales. Sus nombres serán: *INVTVEL*, *LNTTEM*, *TPLU2* y *TRAD2*. Sin nos fijamos en la siguiente tabla se verá que se mejora el coeficiente de determinación, obteniendo el mismo valor que con *bf it ModelHSS* así como la misma ecuación:



Aplicación de MODELHSS a una modelización medioambiental C6

```

*** Linear Model ***

Call: lm(formula = O3 ~ INVTVEL + LNTTEM + TPLU2 + TRAD2,
  data = gnoo3y, na.action = na.omit)

Residuals:
  Min       1Q   Median       3Q      Max
-87.08  -17.74  -0.7942   18      70.8

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  107.7861    5.9014   18.2645  0.0000
INVTVEL      -424.1833   13.1498  -32.2578  0.0000
LNTTEM        38.9569    2.6630   14.6288  0.0000
TPLU2         0.0363    0.0181    2.0082  0.0448
TRAD2         0.0932    0.0543    1.7149  0.0866

Redisual standard error: 25.16 on 1329 degrees of freedom
Multiple R-Squared: 0.5607
F-statistic: 424.1 on 4 and 1329 degrees of freedom,
the p-value is 0
  
```

6.3.5 2ª Ejecución con SPSS

Ahora haremos la misma comparación pero tomando como programa estadístico el SPSS. Estamos en el mismo problema que en el caso anterior: cuando intentamos hacer un análisis de regresión lineal a partir de los datos iniciales nos encontramos con el problema de que tenemos que crear nuevas variables de datos. Debido a ello tenemos que crear 4 variables nuevas en la tabla de datos, *INVVEL*, *LNTTEM*, *PLU2* y *RAD2*, las cuales corresponden a $\frac{1}{VEL}$, el logaritmo neperiano en valor absoluto de TEM y los cuadrados de PLU y RAD respectivamente. A estas



Aplicación de MODELHSS a una modelización medioambiental C6

variables se les aplica las funciones correspondientes quedando creadas 4 columnas más de datos. Al modelizar, las variables independientes serán las nuevas creadas y la dependiente seguirá siendo O_3 .

A continuación se muestran los resultados:

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	RAD2, PLU2 INVVEL, LNTEM ^a		, Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: O_3

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,622 ^a	,386	,384	29,6516

a. Variables predictoras: (Constante), RAD2, PLU2, INVVEL, LNTEM



Aplicación de MODELHSS a una modelización medioambiental C6

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	729971,836	4	182492,959	207,563	,000 ^a
	Residual	1159685,252	1319	879,216		
	Total	1889657,088	1323			

a. Variables predictoras: (Constante), RAD2, PLU2, INVVEL, LNTEM

b. Variable dependiente: O3

Coefficientes^a

Modelo		Coefic. no estand.		Coefic. estand.	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-169,695	13,264		-12,794	,000
	INVVEL	-7,978	,425	-,412	-18,766	,000
	LNTEM	84,003	4,702	,393	17,864	,000
	PLU2	3,395	2,167	,034	1,567	,117
	RAD2	5,653E-02	,023	,053	2,456	,014

b. Variable dependiente: O3

Resumiendo:

- Ecuación obtenida:

$$O3 = -169.695 - 7.978 INVVEL + 84.003 LNTEM + 3.395 PLU2 + 5.653e - 02 RAD2 \quad (6.28)$$

- Coeficiente de determinación:

0.386



Aplicación de MODELHSS a una modelización medioambiental C6

- Variables significativas: solamente INVVEL, LNTEM y RAD2.

Como se ve, el coeficiente de determinación baja bastante con respecto al obtenido con MODELHSS y ello es debido que el SPSS se encuentra con operaciones que le es imposible calcular, como pueden ser divisiones por 0 ó $\ln(0)$. En estos casos, el programa lo que hace es anular esa tupla de datos, en cuanto a valores y resultados, pero a la hora de contabilizar la cantidad de tuplas si que se tiene en cuenta. Por ejemplo, para calcular la media, la varianza, la covarianza, etc. si que es considerada.

Esto lo podemos ver en los distintos “warnings” (mensajes) que da el propio SPSS en su ejecución:

```
>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 26 Current case: 53 Currrent splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 26 Current case: 54 Currrent splitfile group: 1
```



Aplicación de MODELHSS a una modelización medioambiental C6

```
>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 55 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 56 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 893 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 894 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 895 Current splitfile group: 1
```



Aplicación de MODELHSS a una modelización medioambiental C6

```
>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 1205 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.

>Command line: 186 Current case: 1316 Current splitfile group: 1

>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.
>Command line: 186 Current case: 1317 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 53 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 54 Current splitfile group: 1
```



Aplicación de MODELHSS a una modelización medioambiental C6

```
>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 55 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 56 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 893 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 894 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 895 Current splitfile group: 1
```



Aplicación de MODELHSS a una modelización medioambiental C6

```
>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 1316 Current splitfile group: 1

>Warning # 602
>The argument for the natural log function is less than or equal
>to zero.
>The result has been set to the system-missing value.

>Command line: 189 Current case: 1317 Current splitfile group: 1
```

Estos mensajes cabe interpretarlos de la siguiente manera: el número que va a continuación de “warning” es el código que utiliza el programa para identificar el correspondiente problema. A continuación especifica a qué ha sido debido el error y, en tal caso, la acción que desempeña para no tener que abortar la ejecución del programa (aunque por ello se pierda ajuste en los datos), además de indicar la tupla de datos donde se ocasiona el error.

En cambio tipificando y trasladando nos evitamos estos problemas y conseguimos un coeficiente de determinación más alto, además la variable PLU pasa a ser significativa pero en cambio, RAD será no significativa.



Aplicación de MODELHSS a una modelización medioambiental C6

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	TRAD2, PLU2, LNTTEM, INVTVEL ^a		Introducir

- a. Todas las variables solicitadas introducidas
 b. Variable dependiente: O3

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,749 ^a	,561	,559	25,1590

- a. Variables predictoras: (Constante), TRAD2, TPLU2, LNTTEM, INVTVEL

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1073715,191	4	268428,798	424,076	,000 ^a
	Residual	841221,527	1329	632,937		
	Total	1914936,088	1333			

- a. Variables predictoras: (Constante), TRAD2, TPLU2, LNTTEM, INVTVEL
 b. Variable dependiente: O3

Coeficientes^a

Modelo	Coefic. no estand.		Coefic. estand.	t	Sig.	
	B	Error típ.	Beta			
1	(Constante)	107,786	5,901		18,264	,000
	INVTVEL	-424,183	13,150	-,616	-32,258	,000
	LNTTEM	38,957	2,663	,279	14,629	,000
	TPLU2	3,628E-02	,018	,037	2,008	,045
	TRAD2	9,319E-02	,054	,031	1,715	,087

b. Variable dependiente: O3

6.4 Conclusiones

De todas las aplicaciones elegidas y de los resultados obtenidos con los tres programas estadísticos podemos sacar las siguientes conclusiones:

1. La tipificación y trasladación de los datos iniciales de las variables independientes es un factor clave para poder evitar errores en la aplicación de ciertas funciones, no modificando el coeficiente de determinación. Incluso puede hacer que ciertas variables pasen a ser significativas cuando anteriormente no lo eran.
2. La aplicación de funciones a los datos iniciales puede ocasionar, en ciertas situaciones, una mejora del coeficiente de determinación.
3. Si comparamos MODELHSS con los programas estadísticos SPLUS y SPSS podemos destacar la ventaja de que con MODELHSS se pueden obtener familias de ecuaciones con las posibles combinaciones de funciones a los datos originales, de manera que podemos elegir la ecuación que nos resulte más interesante en una



misma ejecución. En cambio, con SPLUS o SPSS tenemos que decidir qué funciones queremos que se apliquen y a qué variables, obteniendo así una única ecuación en una ejecución, sin tener en una misma visión un conjunto de ellas y ver si hay alguna mejor que otra. Además con MODELHSS se produce la tipificación y traslación de los datos directamente, en cambio, con SPLUS y SPSS hay que crear variables adicionales a partir de las originales antes de lanzar el proceso de regresión por mínimos cuadrados.



Universitat d'Alacant
Universidad de Alicante

Capítulo 7

Conclusiones finales

La metodología presentada en este proyecto genera un lenguaje regular, a partir del cual, se construyen ecuaciones matemáticas que modelizan procesos considerados en los sistemas dinámicos.

Con el lenguaje construido también se obtienen las ecuaciones conocidas lineales y no lineales simples de la misma manera que se obtienen en los programas SPLUS y SPSS. Es además posible obtener con dicha metodología familias de ecuaciones en su mayoría no lineales que se podrán almacenar y manipular posteriormente para obtener información sobre el sistema objeto de modelización. En este sentido, hay que recalcar la idónea aplicación a sistemas altamente estructurados en lo que las relaciones existentes son complejas y no lineales, por ejemplo, en los sistemas ecológicos, socioeconómicos, etc. y, en general, en el estudio y modelización de todo ecosistema.



Para obtener las ecuaciones matemáticas se ha utilizado la aproximación a los datos existentes por el método estadístico de regresión de los mínimos cuadrados. Si bien hay que destacar en el tratamiento inicial de los datos experimentales:

- la tipificación,
- la traslación,
- la aplicación de las funciones elegidas y, posteriormente, generadas a los datos obteniéndose, de este modo, nuevas tablas de datos, y
- cálculos intermedios de los datos iniciales, así como de los creados en segundo lugar para poder hallar ciertos estadísticos.

La metodología comentada en lo anteriormente citado se ha implementado en un programa escrito en lenguaje C y al que hemos llamado MODELHSS. La ejecución del mismo permite a partir de los datos experimentales:

- Selección, generación, cálculo y combinación de funciones: elegidas las funciones, se generan las distintas composiciones y combinaciones lineales, dependiendo del número de funciones elegidas, el número de variables independientes y el orden de las transformadas. En ocasiones, se puede necesitar gran cantidad de memoria y espacio debido al elevado número de ecuaciones que se pueden generar.



- Lectura y tratamiento estadístico del fichero de datos inicial: lectura y almacenamiento de los datos para su tratamiento, cálculo de diversos estadísticos, como pueden ser la media, la varianza, la covarianza, etc.; la tipificación de los datos, cálculo del coeficiente de determinación y varianza residual, test de Kolmogorov-Smirnov, etc.
- Funciones sobre la matriz de correlaciones: Son las funciones que tratan y manejan la matriz de varianzas y covarianzas. Es aquí donde indirectamente se aplica el método de los mínimos cuadrados, obteniéndose así los coeficientes de las ecuaciones correspondientes.
- Construcción de la matriz que se utiliza para poder calcular el test de significancia: esta matriz se genera a partir de los datos iniciales, tipificados y trasladados, necesaria para poder calcular la función t-Student, indispensable en el test de significancia.
- Cálculo de la función t-Student: a partir de tres funciones estadísticas extraídas del Numerical Recipes.

Como se puede observar en las aplicaciones del programa MODELHSS hay que destacar en comparación a las ejecuciones de los programas SPLUS y SPSS que la tipificación y traslación de los datos iniciales de las variables independientes es un factor clave para poder evitar errores en la aplicación de ciertas funciones, no modificando el coeficiente de determinación en los casos lineales simples.



En algunos casos, la aplicación de funciones matemáticas a los datos iniciales puede mejorar el coeficiente de determinación. No obstante, esta posible mejora no tiene por qué ser elevada pues, en cualquier caso, depende de los datos iniciales, de la distribución que tiene, etc.

Característica a destacar es la ventaja de que con MODELHSS se pueden obtener familias de ecuaciones con las posibles combinaciones de funciones elegidas aplicadas a los datos originales. Esto conlleva el poder elegir la ecuación o el conjunto de ecuaciones que nos resulten más interesantes dentro de una misma ejecución. En cambio, la ejecución con SPLUS y SPSS es totalmente distinta, pues se les indica el tipo de ecuación que se quiere obtener, de modo que solamente se calculan los coeficientes indeterminados de esa ecuación, con lo cual, para calcular otra ecuación habría que volver a lanzar el proceso repitiendo los pasos de definir la ecuación.

Por otra parte, el proceso que ayuda a evitar problemas de cálculo es la tipificación y traslación de los datos iniciales. En algunos casos nos encontramos con tuplas de datos que contienen valores nulos, de modo que el cambio de estos valores restringe los problemas de desbordamiento y errores de cálculo de forma significativa.

Las futuras investigaciones del proyecto presentado irán encaminadas en los siguientes aspectos:

- Estudiar algunos estadísticos más que se puedan aplicar a las ecuaciones para darnos una mejor información de los ajustes, representatividad y concordancia con respecto a los datos experi-



mentales.

- Aumentar el número de funciones elementales posibles a elegir, disponiendo por tanto de un conjunto mayor de funciones a aplicar a los datos iniciales.
- En la metodología presentada, a una misma variable, se le podía aplicar una función o una composición de funciones. Así pues, se probará introducir nuevas operaciones entre funciones como puede ser el producto.
- Además, serán estudiados aquellos aspectos que afecten a la rapidez y mejora del programa, sin que por ello se tenga que modificar su fin: que es el de obtener una ecuación o conjunto de ecuaciones que se ajusten, en mayor o menor medida, a los datos experimentales y representen un determinado fenómeno.



Universitat d'Alacant
Universidad de Alicante

Capítulo 8

Anexo

8.1 Anexo 1

Tabla de datos correspondiente al ajuste del ozono en el área A de noche del mes de Junio, utilizada en la sección 6.3.2 del Capítulo 6.

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2	18.2	0	0	82
1.9	18.1	0	0	83
0.6	17.9	0	0	75
0.3	17.7	0	0	75
0.4	17.6	0	0	59
0.9	17.4	0	0	40
0.9	17.1	0	0	38

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.2	17	0	0	23
1.4	17	0	0	30
1.4	16.9	0	0	42
1.2	16.7	0	0	39
1	16.6	0	0	32
1	16.5	0	0	19
0.4	16.6	0	0	12
0.1	16.5	0	0	8
0.6	16.5	0	0	4
0.6	16.3	0	0	2
0.7	16.2	0	0	3
0.5	16.1	0	0	2
1	16.2	0	0	2
1.4	16.2	0	15	2
1	21.4	0	5	84
1.2	20.9	0	0	65
1.6	20.6	0	0	73
1.9	20.5	0	0	80
2.2	20.5	0	0	85
1.3	20.3	0	0	83
0.8	20.2	0	0	81
0.7	19.6	0	0	74
1.1	19.1	0	0	64
0.8	18.6	0	0	55
0.3	18.3	0	0	36
1	17.8	0	0	20
0.9	17.6	0	0	10
1.3	17.4	0	0	13
0.9	17.3	0	0	13
1.5	17.4	0	0	9
1.4	17.9	0	0	10

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.4	18.1	0	0	5
2.1	18.6	0	0	25
1.5	19.3	0	0	38
1.2	18.9	0	0	36
1.6	18.7	0	0	47
2.6	19.7	0	0	64
1.8	19.8	0	0	57
1.6	19.3	0	0	51
0.6	18.9	0	0	46
1.9	18.7	0	0	28
1.8	17.9	0	0	2
0.7	17.4	0	0	6
1	17.4	0	0	29
1.1	17.4	0	0	38
0.7	17.1	0	0	36
0.1	16.8	0	0	32
0.3	16.9	0	0	19
1.3	17.4	0	0	11
0.9	17.4	0	0	14
0.8	17.4	0	0	16
1.1	17.4	0	0	9
2	17	0	0	6
1.7	16.7	0	15	4
2.4	21.2	0	15	104
2.4	21	0	5	106
2.3	21	0	0	100
2.1	20.8	0	0	93
1.7	20.6	0	0	90
1.1	20.5	0	0	90
0.8	20.4	0	0	84
0.9	20.2	0	0	85

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.3	19.9	0	0	58
0.9	19.2	0	0	27
1.3	18.8	0	0	24
1.4	18.4	0	0	32
1	18.2	0	0	25
0.7	18.2	0	0	22
1	18	0	0	8
0.6	17.8	0	0	2
0.6	17.5	0	0	2
0.8	17.6	0	0	2
1	17.5	0	0	2
1.2	17.3	0	0	2
0.5	16.9	0	0	2
0.3	16.7	0	0	2
1.2	17	0	0	2
1.2	17.5	0	0	2
1.9	18.1	0	0	2
2.2	18.5	0	0	2
1.6	18.1	0	0	2
1.1	17.3	0	0	2
1.1	16.5	0	0	2
0.4	15.9	0	0	2
0.5	15.7	0	0	2
0.2	15.6	0	0	2
0.8	15.2	0	0	2
0.4	15.2	0	0	2
0.8	14.8	0	0	2
0.2	14.5	0	0	2
0.7	14.5	0	0	2
0.3	14.3	0	0	2
0.6	14.2	0.2	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	14.2	0	0	3
0.3	14.3	0	15	3
0.9	21.1	0	5	93
0.9	20.9	0	0	89
0.6	20.6	0	0	84
0.3	20.4	0	0	80
0.2	20.2	0	0	68
0.1	19.8	0	0	41
0.3	19.7	0	0	46
0.5	19.6	0	0	30
0.9	19.1	0	0	21
0.6	18.8	0	0	9
0.8	18.6	0	0	2
0.9	18.3	0	0	2
0.9	18.1	0	0	2
1.1	17.9	0	0	2
0.9	17.7	0	0	2
0.7	17.6	0	0	2
0.7	17.4	0	0	2
0.6	17.4	0	0	4
0.7	17.4	0	0	3
0.4	17.1	0	0	2
0.5	17	0	0	2
0.9	17	0	0	2
1.5	17	0	0	21
2.2	18.2	0	0	73
2.3	19.1	0	0	79
2.5	19.1	0	0	80
2.6	19.2	0	0	81
2.4	19.3	0	0	87
1.6	19.5	0	0	86

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2.8	19.4	0	0	82
1.3	19.2	0	0	78
1.1	18.7	0	0	62
0.4	18.6	0	0	63
0.6	18.4	0	0	55
1.9	18.8	0	0	79
3.7	19.5	0	0	84
5	19.6	0	0	89
2.9	19.7	0	0	91
2.3	19.6	0	0	90
2.7	19	0.2	0	84
1.5	18.6	0	10	71
3.7	17.9	0	10	50
3.1	17.8	0	0	42
2.4	17.7	0	0	29
2.5	17.6	0	0	25
1.6	17.8	0	0	7
1	17.8	0	0	2
2.1	17.8	0	0	2
2.1	17.8	0	0	2
2	17.7	0	0	4
1.4	17.5	0	0	8
1.3	17.4	0	0	10
1.9	17.4	0	0	15
2.1	17.7	0	0	22
2	17.7	0	0	24
1.9	17.4	0	0	22
1.8	17	0	0	19
1.9	16.9	0	0	20
2.2	16.8	0	0	15
2.8	16.8	0	0	12

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
3	16.9	0	0	2
3.3	16.7	0	0	2
1.6	16.9	0	0	2
0.6	16.9	0	0	2
1.6	16.8	0	0	2
1	16.9	0	0	2
1.1	17.3	0	0	11
0.7	17.2	0	0	8
0.8	17.4	0	0	21
0.5	17.2	0	0	16
1.2	16.9	0	0	14
1.5	17.1	0	0	16
0.7	17.6	0	0	26
0.9	17.8	0	0	36
0.4	17.7	0	0	31
0.7	16.9	0	0	19
1	16.4	0	0	13
1	15.5	0	0	6
1	15	0	0	4
0.8	14.7	0	0	2
0.9	14.7	0	0	2
1.1	14.7	0	5	3
0.5	14.8	0	10	3
0.7	14.8	0	10	4
3.1	23.7	0	15	73
3.8	21.9	0	0	77
3.2	21.7	0	0	74
2.4	22	0	0	63
2	21.9	0	0	55
0.5	21.6	0	0	39
1.3	21	0	0	39

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1	20.8	0	0	47
0.2	20.5	0	0	40
0.1	20.5	0	0	36
0.5	20.1	0	0	23
0.5	19.7	0	0	20
0.4	19.6	0	0	16
1.2	19.5	0	0	8
1.8	19.6	0	0	10
1.9	20.2	0	0	3
1.9	20	0	0	11
2.2	19.6	0	0	17
2	19.7	0	0	21
1.7	19.7	0	0	24
2	19.5	0	0	16
1.9	19.4	0	0	4
2	19.6	0	0	2
2.1	20	0	0	2
1.2	20.1	0	0	2
2.1	19.8	0	0	2
2.5	20.1	0	0	2
2.3	19.8	0	0	3
2.1	19.5	0	0	25
1.8	19.4	0	0	29
1.2	19	0	0	29
1.2	18.7	0	0	30
1.7	18.7	0	0	30
1.9	18.7	0	0	24
2.4	18.5	0	0	27
2	18.3	0	0	27
1.8	18.2	0	0	12
1.5	18.2	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.3	17.8	0	0	2
2	17.8	0	5	7
1.6	22.5	0	5	78
1.6	21.8	0	0	86
1	21.6	0	0	81
1.1	21.4	0	0	70
1.3	21.2	0	0	80
2	21	0	0	85
1.7	21	0	0	85
1.4	21	0	0	85
1	20.9	0	0	87
1.2	20.9	0	0	81
0.9	20.7	0	0	81
0.6	20.6	0	0	81
0.4	20.6	0	0	76
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0.7	20.1	0	0	69
1.2	20.2	0	0	67
1.1	20.1	0	0	70
1.8	20.1	0	0	68
2	20.1	0	0	67
2.4	20.1	0	0	70
2	20.1	0	0	71
1.6	20.1	0	0	63
1.3	20	0	0	54
1.6	19.9	0	0	62
2.5	19.9	0	0	63
2.4	19.9	0	0	68

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2	19.9	0	0	67
2.3	19.8	0	0	58
2.3	19.8	0	0	56
2.2	19.7	0	0	53
1.8	19.7	0	0	54
1.9	19.6	0	0	59
0.7	19.6	0	0	50
0.5	19.2	0	0	34
0.7	18.9	0	0	9
0.8	18.8	0	0	18
0.1	18.7	0	0	23
0.2	18.7	0	5	28
0.2	22.3	0	5	75
0.1	22.1	0	0	56
0.3	21.9	0	0	55
1.1	21.5	0	0	28
1	21.3	0	0	25
0.2	21.2	0	0	23
0.2	20.8	0	0	20
1.1	20.5	0	0	13
0.5	20.5	0	0	39
0.2	20.5	0	0	27
0.3	20.1	0	0	13
0.4	19.6	0	0	6
0.5	19.4	0	0	7
0.4	19.4	0	0	7
0.6	19.1	0	0	3
0.4	19.1	0	0	2
0	18.9	0	0	2
0.6	18.7	0	0	2
0.9	18.6	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.8	18.5	0	0	2
0.8	18.5	0	0	2
0.6	18.4	0	0	2
0.4	18.3	0	0	2
0.5	17.9	0	0	2
0.4	17.7	0	0	3
0.4	17.5	0	0	3
0.7	17.8	0	0	2
0.7	17.7	0	0	2
0.8	17.5	0	0	2
0.6	17.3	0	0	2
0.4	17.2	0	0	2
0.1	17.3	0	0	3
0.5	17.2	0	0	3
0.3	17.1	0	0	3
0.4	17.2	0	0	2
0.3	17.1	0	0	3
0.3	16.9	0	0	2
0.4	17	0	0	3
0.4	16.9	0	0	2
0.4	16.7	0	10	2
0.7	23.5	0	5	75
0.4	23.2	0	0	73
0.5	23	0	0	71
0.5	22.6	0	0	56
0.5	22.3	0	0	45
0.2	21.9	0	0	34
0.1	21.6	0	0	20
0.3	21.4	0	0	11
0.8	21	0	0	6
0.7	20.9	0	0	7

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.4	20.6	0	0	7
0.2	20.5	0	0	3
0.2	20.3	0	0	5
0.7	20.1	0	0	5
0.7	19.6	0	0	3
0.7	19.4	0	0	3
1	19.4	0	0	4
0.4	19.1	0	0	3
0.6	19.3	0	0	3
0.3	19.4	0	0	3
1.2	19.1	0	0	4
0.8	18.1	0	0	3
0.1	17.8	0	0	3
0.5	18	0	0	3
1.1	18.3	0	0	4
1.3	18.1	0	0	2
0.9	17.7	0	0	2
0.4	17.6	0	0	2
0.9	17.4	0	0	2
0.9	17.4	0	0	4
1.1	17.4	0	0	4
0.8	17.4	0	0	5
0.4	17.4	0	0	6
0.7	17.5	0	0	5
0.7	17.4	0	0	6
0.5	17.1	0	0	6
1.1	17.4	0	0	8
1.6	18.2	0	0	12
1	19.1	0	5	18
3.8	21.9	0	15	77
3.4	21.9	0	5	77

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
3.8	21.8	0	0	75
3.3	21.8	0	0	74
3.1	21.9	0	0	70
2.6	21.9	0	0	67
2.2	21.8	0	0	67
2.5	21.8	0	0	68
3.2	21.8	0	0	76
3	21.6	0	0	76
2.8	21.4	0	0	74
3.2	21.4	0	0	77
2.3	21.4	0	0	76
2.5	21.4	0	0	74
2.2	21.3	0	0	71
2.6	21.2	0	0	69
2.6	21.2	0	0	64
1.1	21.1	0	0	62
1.4	21.2	0	0	73
1	21.1	0	0	74
0.6	20.8	0	0	61
0.4	20.2	0	0	37
1.5	19.9	0	0	25
1	19.8	0	0	18
1.2	19.6	0	0	25
0.5	19.4	0	0	26
0.3	19.5	0	0	12
0.5	19.6	0	0	5
0.7	19.7	0	0	4
1.2	19.7	0	0	2
1.1	19.8	0	0	8
0.2	20	0	0	22
0.9	19.3	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.7	18.7	0	0	2
1.1	18.5	0	0	2
0.4	18.3	0	0	2
0.9	18.1	0	0	2
0.9	18.2	0	0	2
1.1	18.2	0	0	2
0.9	18.1	0	0	2
0.5	18	0	10	2
0.4	21.3	0	10	59
0.5	21.3	0	0	52
0.9	21.4	0	0	52
1.8	21.3	0	0	56
3.5	21.2	0	0	60
2	21.1	0	0	65
0.9	21	0	0	63
2.2	21.1	0	0	54
1.3	21.1	0	0	41
0.8	21	0	0	44
1.2	21.1	0	0	51
0.9	21.1	0	0	37
1.2	21	0	0	46
1	20.8	0	0	48
0.9	20.8	0	0	45
0.5	20.8	0	0	38
0.8	20.8	0	0	35
0.8	20.6	0	0	29
1.1	20.6	0	0	24
1	20.5	0	0	19
1.2	20.4	0	0	12
1.2	20.4	0	0	7
0.9	20.6	0	0	6

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.1	20.5	0	0	11
1.4	20.3	0	0	6
0.8	20.3	0	0	5
0.7	20.2	0	0	4
0.7	20	0	0	2
0.8	19.8	0	0	2
0.2	19.5	0	0	2
0.2	19.3	0	0	2
0.2	19.3	0	0	2
0.2	19.2	0	0	2
0.5	19.2	0	0	2
0.2	19.2	0	0	2
0.1	19.1	0	0	2
0.2	19	0	0	2
0.3	19	0	0	2
1.3	18.8	0	0	2
1.5	18.7	0	0	2
1	18.8	0	15	2
2.2	23.2	0	10	95
1.4	22.8	0	0	90
1.4	22.5	0	0	91
1.5	22.4	0	0	87
1.1	22.3	0	0	87
1	22.2	0	0	81
0.4	21.9	0	0	75
0.3	21.9	0	0	80
0.6	21.9	0	0	81
0.3	21.8	0	0	79
0.2	21.6	0	0	69
0.1	21.4	0	0	44
0.4	21.1	0	0	43

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	21	0	0	24
0.4	20.8	0	0	28
0.1	20.7	0	0	22
0.2	20.4	0	0	17
0.2	20.4	0	0	22
0.4	20.4	0	0	21
0.6	20.3	0	0	15
0.3	19.7	0	0	16
0.1	19.3	0	0	11
0.5	19.4	0	0	16
0.7	19.6	0	0	11
1.1	19.4	0	0	11
0.7	19.3	0	0	4
0.1	19.2	0	0	4
0.2	19.4	0	0	5
0.3	19.3	0	0	2
0.9	19	0	0	3
1.3	19.2	0	0	9
0.6	19.7	0	0	8
0.6	18.6	0	0	2
0.5	18.6	0	0	2
0.4	18.9	0	0	2
0.2	18.7	0	0	2
0.5	18.2	0	0	2
0.7	18.2	0	0	2
0.9	18.3	0	0	2
0.4	18.3	0	10	2
1	23.1	0	15	70
0.8	22.7	0	5	69
0.6	22.6	0	0	60
0.6	22.4	0	0	53

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.5	22.2	0	0	50
0.1	22	0	0	49
0.1	21.8	0	0	37
0.5	21.5	0	0	30
0.5	21.2	0	0	24
0.7	21.2	0	0	23
0.3	20.9	0	0	12
0.5	20.8	0	0	12
0.3	20.8	0	0	13
0.1	20.7	0	0	10
0.8	20.6	0	0	10
1.2	20.7	0	0	8
1	21	0	0	21
0	21.2	0	0	23
0.5	21	0	0	15
0.4	20.7	0	0	7
0.8	20.3	0	0	3
1.2	20.3	0	0	5
1.4	20.1	0	0	4
0.4	20	0	0	2
0.3	19.8	0	0	2
0.2	19.7	0	0	2
0.2	19.6	0	0	2
0.3	19.6	0	0	2
0.6	19.8	0	0	2
0.3	20	0	0	2
0.6	19.8	0	0	2
0.7	19.6	0	0	2
0.5	19.4	0	0	2
0.1	19.3	0	0	2
0.4	19.3	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.6	19.6	0	0	2
0.2	19.6	0	0	2
0.6	19.7	0	0	2
1.3	19.7	0	0	2
0.8	19.7	0	0	2
0.7	19.7	0	10	2
0.9	23	0	5	75
0.6	22.8	0	0	71
0.3	22.6	0	0	70
0.3	22.4	0	0	73
0.5	22.3	0	0	71
0.9	22.3	0	0	63
0.8	22.2	0	0	59
1	22.2	0	0	62
0.7	22.1	0	0	60
0.7	22.1	0	0	58
0.8	22	0	0	54
0.1	21.8	0	0	42
0.4	21.8	0	0	49
0.6	21.8	0	0	42
0.5	21.7	0	0	29
1	21.3	0	0	29
0.4	21.1	0	0	28
0.1	20.9	0	0	20
0.1	20.7	0	0	13
0.1	20.8	0	0	11
0.4	20.8	0	0	7
0.8	20.8	0	0	2
0.8	20.6	0	0	2
0.6	20.3	0	0	2
0.1	20	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	19.9	0.2	0	2
0.2	20	0	0	2
0.2	20	0	0	2
0.4	19.8	0	0	2
0.3	19.7	0	0	2
0.4	19.7	0	0	2
0.6	19.7	0	0	3
0.5	19.6	0	0	3
0.5	19.5	0	0	3
0.1	19.2	0	0	3
0.1	19.3	0	0	3
0.6	19.3	0	0	2
0.3	19.1	0	0	3
0.2	19	0	0	3
0.4	19.1	0	10	3
3	23.2	0	10	81
3	22.9	0	0	78
2.6	22.8	0	0	75
2.8	22.7	0	0	75
2.6	22.5	0	0	76
2.8	22.5	0	0	77
2.7	22.6	0	0	75
2.4	22.5	0	0	74
2.5	22.4	0	0	74
2.8	22.3	0	0	75
2.7	22.3	0	0	75
2.1	22.3	0	0	73
2.4	22.3	0	0	74
2	22.3	0	0	71
2.6	22.2	0	0	73
2.2	22.1	0	0	69

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
3.1	22.1	0	0	70
3.1	22	0	0	69
2.3	21.9	0	0	64
2.3	21.8	0	0	63
2.3	21.6	0	0	63
2.1	21.6	0	0	64
1.7	21.5	0	0	61
1.7	21.2	0	0	55
1.4	21	0	0	55
1.2	20.9	0	0	54
1.5	20.8	0	0	55
1.1	20.8	0	0	54
0.8	20.5	0	0	42
0.9	20.2	0	0	45
0.8	20.1	0	0	41
0.8	20	0	0	36
0.9	19.8	0	0	30
0.8	19.6	0	0	29
0.5	19.6	0	0	29
0.4	19.5	0	0	24
0.6	19.3	0	0	19
0.7	19.3	0	0	18
0.5	19.2	0	0	15
0.2	19.1	0	15	15
1.1	22.3	0	10	70
0.6	22.1	0	0	71
0.8	21.9	0	0	75
0.8	21.8	0	0	72
0.4	21.8	0	0	74
0.2	21.7	0	0	70
0.4	21.6	0	0	62

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	21.5	0	0	50
0.3	21.5	0	0	52
0.4	21.5	0	0	37
0.2	21.4	0	0	35
0.3	21.3	0	0	29
0.3	21.2	0	0	23
0.1	21	0	0	11
0.1	20.8	0	0	16
0.3	20.8	0	0	21
0	20.8	0	0	33
0.9	20.7	0	0	34
0.8	20.6	0	0	32
0.5	20.5	0	0	31
0.7	20.5	0	0	27
1	20.5	0	0	41
0.6	20.7	0	0	45
0.3	20.5	0	0	30
0.1	20.4	0	0	33
0.1	20.6	0	0	37
0.2	20.7	0	0	26
0.5	20.6	0	0	27
0.3	20.8	0	0	32
0.2	20.8	0	0	35
0.5	20.7	0	0	19
0.1	20.5	0	0	23
0.2	20.8	0	0	35
0.3	20.9	0	0	38
0.2	20.9	0	0	44
0.4	21	0	0	45
0.5	21	0	0	46
0.1	21.1	0	0	43

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.1	21.1	0	0	40
0.2	21.1	0	0	39
0.1	21.2	0	0	40
0.2	21.3	0	15	35
1.2	23.1	0	15	69
0.7	22.9	0	10	67
0.7	22.7	0	0	61
1	22.6	0	0	58
0.8	22.4	0	0	56
0.7	22.3	0	0	55
0.5	22.2	0	0	47
0.8	22.3	0	0	47
0.7	22.2	0	0	41
0.4	22	0	0	36
0.4	21.9	0	0	34
0.6	21.7	0	0	23
0.8	21.5	0	0	22
0.5	21.5	0	0	29
0.1	21.7	0	0	31
0.1	21.7	0	0	34
0.4	21.7	0	0	31
0.5	21.5	0	0	34
1.2	21.5	0	0	45
1.9	21.9	0	0	62
2.1	22	0	0	61
1.9	22.1	0	0	64
2.3	22.3	0	0	66
2.1	22.5	0	0	66
2.7	22.5	0	0	64
2.4	22.4	0	0	61
2.2	22.3	0	0	64

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2.4	22.3	0	0	60
2.7	22.3	0	0	52
2	22.4	0	0	52
2	22.4	0	0	62
2.6	22.3	0	0	66
2.2	22.3	0	0	66
1.3	22.2	0	0	52
1.7	21.8	0	0	44
1.7	21.5	0	0	44
2.2	21.5	0	0	58
2	22.1	0	0	59
1.9	22.2	0	0	60
2	22.3	0	0	60
1.6	22.3	0	0	56
1.5	22.3	0	15	55
3.8	21.7	0	5	81
3.8	21.5	0	0	81
3.9	21.5	0	0	83
3.8	21.5	0	0	84
3.6	21.5	0	0	86
3.4	21.6	0	0	85
3.3	21.6	0	0	86
3.5	21.6	0	0	86
3.4	21.6	0	0	85
3.5	21.6	0	0	85
3.9	21.6	0	0	89
3.4	21.5	0	0	91
3.6	21.6	0	0	91
3	21.7	0	0	88
2.8	21.6	0	0	82
2.8	21.6	0	0	80

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2.6	21.6	0	0	78
2.9	21.5	0	0	78
2.8	21.4	0	0	79
2.8	21.4	0	0	81
3	21.4	0	0	83
3.1	21.4	0	0	84
3.1	21.5	0	0	86
2.2	21.4	0	0	85
2.6	21.4	0	0	85
2.9	20.4	0	0	80
2.8	19.3	0.2	0	72
2.6	19.2	0	0	78
2.5	19.1	0	0	79
2.7	19.1	0	0	83
2.7	19.2	0.1	0	82
3	18.7	0	0	82
2.7	18.9	0	0	80
2.7	19.2	0	0	80
2.3	19.6	0	0	80
2.6	19.8	0	0	80
3.1	20.1	0	0	81
3.4	20.1	0	0	81
3.3	20.3	0	0	80
2.8	20.1	0	0	77
2.6	20.1	0	0	76
2.3	20.1	0	5	74
2.5	20.2	0.1	0	76
2.7	19.8	0	5	72
1.4	0	0	15	38
2.3	0	0	15	25
2.9	0	0	15	42

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
2.8	0	0	15	56
2	0	0	15	62
1.2	0	0	5	50
0.8	0	0	5	45
0.4	0	0	5	23
0.4	0	0	0	11
0.5	0	0	0	2
1.2	0	0	0	3
0.9	0	0	0	2
0.7	0	0	0	2
0.8	0	0	0	2
0.7	0	0	0	2
0.5	0	0	0	2
0.5	0	0	0	2
0.9	0	0	0	4
0.9	0	0	0	8
0.6	0	0	0	7
1.3	0	0	0	2
1.6	0	0	0	2
1.4	0	0	0	2
2	0	0	0	2
1.8	0	0	0	2
1.8	0	0	0	5
2	0	0	0	7
2.4	17.9	0	0	15
1.7	18.2	0	0	9
1.5	18.1	0	0	3
1.8	18	0	0	2
1.7	18.2	0	0	19
1.1	18.3	0	0	23
1	18	0	0	11

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.2	18	0	0	21
0.9	18.1	0	0	22
1.5	17.8	0	0	19
1.5	17.9	0	0	20
1.6	17.9	0	0	20
1.4	17.7	0	0	19
1.4	17.5	0	0	6
1.3	17	0	0	5
0.5	16.5	0	0	8
0.5	16.3	0	0	5
0.8	16	0	0	3
0.2	15.8	0	0	3
0.4	15.8	0	15	4
1.4	22.8	0	5	119
1.6	22.5	0	0	113
2.3	22.1	0	0	120
2.1	21.8	0	0	122
2.3	21.7	0	0	122
2.2	21.7	0	0	122
2.4	21.6	0	0	126
2.6	21.6	0	0	126
2.3	21.6	0	0	126
2.2	21.5	0	0	127
2.1	21.5	0	0	124
2.3	21.5	0	0	123
2.2	21.5	0	0	121
2.1	21.4	0	0	117
1.7	21.3	0	0	113
0.8	21.1	0	0	103
0.3	20.6	0	0	77
0.4	20.1	0	0	54

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.7	19.6	0	0	33
1	19.2	0	0	21
1.2	18.7	0	0	13
1.2	18.3	0	0	2
0.9	17.9	0	0	2
0.7	17.8	0	0	2
0.1	18	0	0	2
0.6	18.2	0	0	12
0.5	18.7	0	0	17
0.5	18.7	0	0	25
0.8	18.8	0	0	24
1.2	18.7	0	0	21
1.3	18.1	0	0	11
0.8	17.7	0	0	13
0.6	17.5	0	0	13
1.2	17.6	0	0	8
0.6	17.4	0	0	4
0.7	17.1	0	0	2
0.9	16.9	0	0	2
1.3	16.7	0	0	2
1.4	16.5	0	0	2
0.6	16	0	10	2
0.8	21.8	0	10	121
0.6	21.5	0	0	120
0.5	21.4	0	0	125
0.2	21.2	0	0	107
0.2	21	0	0	94
0.1	20.9	0	0	106
0.3	20.8	0	0	93
0.4	20.7	0	0	64
0.6	20.5	0	0	46

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.6	20.1	0	0	35
0.5	19.9	0	0	19
0.2	19.7	0	0	19
0.3	19.6	0	0	12
0.2	19.6	0	0	5
0.5	19.4	0	0	3
0.4	19.4	0	0	5
0.5	19.4	0	0	8
0.2	19	0	0	5
0.3	18.8	0	0	3
0.1	18.7	0	0	2
0.7	18.5	0	0	2
0.5	18	0	0	2
0.5	17.8	0	0	2
0.3	17.6	0	0	2
0.3	17.4	0	0	2
0.6	17.5	0	0	2
0.4	17.6	0	0	2
0.4	17.4	0	0	2
1	17.2	0	0	2
1.4	16.8	0	0	2
1.5	16.6	0	0	2
1.4	16.6	0	0	2
1.5	16.5	0	0	2
1.8	16.6	0	0	2
1.9	16.7	0	0	2
1.9	16.9	0	0	2
2	17.3	0	0	2
1.6	17.5	0	0	2
1.6	17.6	0	0	2
0.7	17.7	0	10	3

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.8	25.9	0	5	87
0.6	25.2	0	0	48
0.4	24.7	0	0	36
1	24.2	0	0	30
0.5	23.7	0	0	23
0.7	23.2	0	0	17
1.1	22.9	0	0	17
1.9	22.8	0	0	9
2.3	23	0	0	2
2.5	24.2	0	0	10
3	24.8	0	0	14
3.2	26	0	0	24
2.4	26.5	0	0	24
1.9	26.3	0	0	2
1.8	25.7	0	0	2
2.3	25.1	0	0	3
0	25.4	0	0	23
2.2	25.5	0	0	30
3.2	26	0	0	41
4.3	27.1	0	0	45
4.3	27.4	0	0	46
3.9	27.1	0	0	51
2.9	26.9	0	0	53
2.7	26.4	0	0	53
2.2	26.4	0	0	54
1.6	25.8	0	0	47
0.6	24.6	0	0	34
0.3	24	0	0	33
1.7	23.5	0	0	28
0.6	23.3	0	0	33
0.6	23	0	0	16

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.5	22.8	0	0	23
1.3	22.7	0	0	62
2.2	22.4	0	0	86
2.2	22.3	0	0	89
1.6	22.2	0	0	79
0.7	22	0	0	50
0.7	21.9	0	0	50
1.3	21.6	0	0	42
1.9	21.8	0	5	38
0	0	0	0	0
0	0	0	0	0
0.8	20.8	0	0	58
1	20.7	0	0	63
0.7	20.7	0	0	66
0.6	20.6	0	0	63
0.7	20.5	0	0	59
0.2	20.4	0	0	59
0.2	20.3	0	0	49
0.6	20.1	0	0	40
0.7	19.7	0	0	32
0.9	19.4	0	0	18
0.7	19.1	0	0	10
0.6	19.1	0	0	11
0.6	18.8	0	0	6
0.7	18.6	0	0	2
0.5	18.3	0	0	6
0.8	18.3	0	0	14
0.3	18.1	0	0	23
1	18.2	0	0	23
0.9	18.2	0	0	19
0.4	18	0	0	19

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.1	17.8	0	0	21
0.4	18	0	0	17
0.5	18	0	0	13
1.1	17.8	0	0	11
1.2	17.7	0	0	16
0.8	17.5	0	0	16
0.2	17	0	0	11
0.8	16.7	0	0	8
0.8	16.8	0	0	9
0.7	16.7	0	0	10
1.4	16.7	0	0	11
0.8	16.4	0	0	6
0.4	16	0	0	2
0.8	15.8	0	0	4
0.4	15.9	0	0	6
0.6	15.7	0	0	6
1	15.6	0	0	6
1	15.6	0	15	4
0.9	21.4	0	15	83
0.5	21.4	0	5	79
0.3	21.3	0	0	76
0.5	21.1	0	0	74
0.1	21	0	0	74
0.1	20.8	0	0	61
0.1	20.8	0	0	59
0.1	20.8	0	0	53
0.2	20.9	0	0	52
0.4	20.9	0	0	47
0.5	20.7	0	0	39
0.7	20.6	0	0	45
0.4	20.6	0	0	53

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	20.8	0	0	54
0.3	20.8	0	0	43
0.1	20.7	0	0	46
0.5	20.5	0	0	45
0.5	20.2	0	0	30
0.5	19.8	0	0	24
0.9	19.7	0	0	6
1.1	19.6	0	0	2
0.4	19.3	0	0	9
0.3	19	0	0	10
0.6	18.9	0	0	5
0.4	18.7	0	0	6
0.5	18.7	0	0	10
0.4	18.6	0	0	7
0.7	18.3	0	0	7
0.4	18	0	0	5
0.6	18	0	0	6
0.4	18.1	0	0	6
0.4	18.1	0	0	4
0.8	18	0	0	7
0.8	17.9	0	0	9
0.6	17.8	0	0	13
0.7	18.1	0	0	16
0.8	18	0	0	18
0.6	17.7	0	0	19
0.6	17.5	0	0	19
0.5	17.4	0	0	24
0.4	17.4	0	10	25
0.9	21.4	0	10	98
0.9	21.3	0	0	89
0.3	21.2	0	0	89

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.2	21.1	0	0	91
0.5	20.9	0	0	89
0.2	20.9	0	0	85
0.5	20.6	0	0	62
0.9	20.4	0	0	63
0.8	20.2	0	0	65
0.1	20.1	0	0	73
0.1	20	0	0	49
0.1	19.8	0	0	33
0.2	19.6	0	0	29
0.7	19.5	0	0	16
1	19.3	0	0	4
0.8	19.3	0	0	3
0.1	19.3	0	0	12
0.3	19.2	0	0	18
0.2	18.9	0	0	16
0.1	19	0	0	24
0.1	19.1	0	0	28
0.3	19.4	0	0	24
0.6	19.4	0	0	19
0.1	19.3	0	0	14
0.1	19.2	0	0	12
0.3	19.1	0	0	10
0.1	18.9	0	0	7
0.1	18.7	0	0	8
0.6	18.7	0	0	7
0.6	18.7	0	0	6
0.1	18.4	0	0	5
0.5	18.3	0	0	5
0.7	18.3	0	0	4
0.6	18	0	0	2

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}$ (JUN).

VEL	TEM	PLU	RAD	O3
0.5	17.8	0	0	2
0.8	17.7	0	0	2
0.7	17.8	0	0	2
0.6	17.7	0	0	2
0.6	17.7	0	0	2
0.5	17.6	0	10	2
2	22.6	0	10	76
1.7	22.3	0	0	73
1.6	22.1	0	0	70
1.8	21.9	0	0	69
1.5	21.9	0	0	65
1.5	21.8	0	0	64
1.3	21.7	0	0	64
1.2	21.6	0	0	66
1	21.5	0	0	63
1.2	21.4	0	0	63
1.2	21.4	0	0	61
1	21.3	0	0	63
0.8	21.3	0	0	57
1.2	20.9	0	0	34
1.3	20.5	0	0	30
0.8	20.3	0	0	30
0.3	20.2	0	0	24
0.6	19.9	0	0	17
0.9	19.6	0	0	10
1	19.5	0	0	6
0.9	19.3	0	0	5
0.6	19.2	0	0	5
0.2	19.2	0	0	2
0.6	19	0	0	5
1.2	18.7	0	0	12

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.1	18.8	0	0	15
0.8	18.7	0	0	13
0.5	18.6	0	0	9
0.2	18.5	0	0	10
0.4	18.3	0	0	11
0.3	18.1	0	0	6
0.2	17.9	0	0	6
0.2	17.9	0	0	5
0.7	17.8	0	0	2
0.8	17.8	0	0	4
0.6	17.7	0	0	4
0.7	17.5	0	0	3
0.7	17.5	0	0	2
0.5	17.2	0	0	2
1	17.4	0	10	4
2.4	23.3	0	5	94
2	23.2	0	0	92
2.1	23.2	0	0	93
2.1	23.1	0	0	89
1.7	23	0	0	85
1.2	22.9	0	0	77
0.9	22.8	0	0	72
0.9	22.6	0	0	60
2.8	22.6	0	0	83
3.6	23.1	0	0	93
3.9	22.9	0	0	92
4.1	22.7	0	0	87
3.5	22.4	0	0	90
3.5	22.4	0	0	83
3.3	22.5	0	0	79
4	22.4	0	0	79

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
4.6	22.3	0	0	80
4.8	22.2	0	0	82
4.6	22.2	0	0	81
4.4	22	0	0	79
4	21.9	0	0	76
3.9	21.9	0	0	72
3.2	21.9	0	0	60
2.8	21.9	0	0	58
2.4	21.6	0	0	56
2.4	21.5	0	0	53
2.1	21.5	0	0	50
1.7	21.5	0	0	46
1.4	21.4	0	0	43
1.7	21.1	0	0	44
1.9	21.1	0	0	43
1.3	21	0	0	42
1.8	20.9	0	0	46
1.5	20.8	0	0	48
1.8	20.6	0	0	49
1.5	20.4	0	0	52
1.7	20.2	0	0	54
2.3	20.2	0	0	60
2.6	20.1	0	0	63
2.9	19.9	0	0	65
1.8	19.8	0	5	60
1.3	19.7	0	5	63
2	19.6	0	10	61
0.7	17.6	0	15	48
0.7	17.5	0	5	44
0.9	17.5	0	0	38
0.7	17.4	0	0	40

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1	17.4	0	0	38
0.9	17.4	0	0	37
1.3	17.4	0	0	36
1.2	17.3	0	0	35
1	17.4	0	0	35
0.9	17.4	0	0	38
0.9	17.4	0	0	40
0.8	17.4	0	0	39
0.7	17.4	0	0	28
0.7	17.3	0	0	27
1.1	17.3	0	0	27
1.1	17.2	0	0	24
1	17.3	0	0	25
0.9	17.3	0	0	22
0.3	17.4	0	0	18
1.1	17.4	0	0	20
0.8	17.4	0	0	16
1.2	17.4	0	0	15
1	17.4	0	0	21
0.7	17.5	0	0	38
1.3	17.7	0	0	75
1.2	17.8	0	0	77
1.7	17.4	0	0	77
1.6	17.1	0	0	72
1.2	17	0	0	69
1.3	16.8	0	0	65
0.9	16.7	0	0	61
1.1	16.8	0	0	56
0.3	16.9	0	0	61
0.7	17	0	0	64
0.9	16.9	0	0	63

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.3	16.5	0	0	52
1.4	16.2	0	0	48
1.3	16.2	0	0	48
0.8	16.4	0	0	47
1.5	16.5	0	0	50
1.4	16.5	0	0	49
1.3	16.5	0	0	48
0.7	16.6	0	0	46
0.4	16.8	0	0	46
0.7	16.8	0	0	47
0.9	16.8	0	5	51
5.4	20.7	0	10	91
5.1	20.3	0	0	89
3.7	20.1	0	0	83
2.8	19.5	0	0	85
2.8	19.1	0	0	85
3.6	19	0	0	84
3.5	18.9	0	0	82
3	18.6	0	0	78
2.2	18.3	0	0	73
1.9	17.5	0	0	66
2.1	17.1	0	0	61
2.1	17.2	0	0	58
1.5	16.9	0	0	39
1.9	16.4	0	0	9
2.2	15.6	0	0	11
2.1	16.2	0	0	49
2.1	17	0	0	64
2.4	17.5	0	0	74
2.4	17.7	0	0	72
3	17.9	0	0	79

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1.9	17.8	0	0	79
2.8	17.7	0	0	82
3.1	17.7	0	0	82
2.8	17.4	0	0	76
2.3	16.9	0	0	68
2.1	16.4	0	0	69
1.2	16.3	0	0	72
0.3	16	0	0	64
2	16	0	0	66
1.3	16	0	0	62
1	14.8	0	0	55
0.5	14.5	0	0	41
1.3	14.6	0	0	31
1.2	14.1	0	0	13
1.3	14.1	0	0	13
1.7	14	0	0	8
1.7	13.8	0	0	15
0.8	13.1	0	0	9
0.4	12.8	0	0	17
1.5	20.8	0	5	73
1.7	20.5	0	0	67
1.7	20.2	0	0	67
2	20.1	0	0	75
2.1	20	0	0	75
1.9	20	0	0	83
1.8	19.9	0	0	84
1.5	19.8	0	0	84
1.4	18.9	0	0	37
0.5	18.2	0	0	25
0.8	17.9	0	0	6
1.1	17.7	0	0	3

Continúa en la página siguiente...

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
1	17.5	0	0	4
0.9	17	0	0	5
1	16.8	0	0	2
0.4	16.5	0	0	2
0.9	16.3	0	0	3
0.8	16.2	0	0	12
0.3	16	0	0	5
0.9	15.6	0	0	6
1	15.7	0	0	8
1.2	15.8	0	0	18
1	16.2	0	0	27
1.3	15.8	0	0	18
1.5	15.2	0	0	14
1.4	15.3	0	0	16
1	15.6	0	0	18
0.9	15.4	0	0	20
1	15	0	0	16
0.8	14.2	0	0	8
0.6	13.7	0	0	6
1	13.3	0	0	5
0.8	13.3	0	0	9
1.5	13.7	0	0	15
1.3	13.9	0	0	20
1.5	14	0	0	18
1.9	14.9	0	0	21
1.8	15.4	0	0	22
1.7	15.5	0	0	13
1.7	15.5	0	15	5
1.6	21.4	0	5	96
1.1	21.2	0	0	91
0.4	21	0	0	83

Continúa en la página siguiente...



Universitat d'Alacant

Anexo 1

C8

Universidad de Alicante

Tabla de datos de $O_3^{A,N}(JUN)$.

VEL	TEM	PLU	RAD	O3
0.3	20.5	0	0	63
0.8	20.3	0	0	76
1.1	20.4	0	0	73
1	20.1	0	0	73
0.7	20.1	0	0	71
0.9	19.8	0	0	56
1.4	19.2	0	0	45
1.3	18.7	0	0	29
1.5	18.3	0	0	27
0.5	18.1	0	0	35
0.2	17.9	0	0	24
0.5	17.5	0	0	11
1	17.1	0	0	7
1.1	16.9	0	0	18
1.1	16.9	0	0	17
0.8	16.7	0	0	13



Universitat d'Alacant
Universidad de Alicante

Bibliografía

- [1] Aracil, Javier: *Introducción a la dinámica de Sistemas*. Alianza Universal Textos, 2ª Edición, 1986
- [2] Chambers, J. M. & Hastie, T.J.: *Statistical Models in S*. Wadsworth & Brooks, Pacific Grove, CA, 1997
- [3] Checkland, P.B.: *Systems Theory, Systems Practice*. N.Y. John Wiley, 1981
- [4] Christian, Kaare: *Diccionario de C y UNIX*. Anaya Multimedia, 1990
- [5] Cortés, M., Villacampa, Y., Mateu, J. & Usó, J.L.: *A new methodology for modelling highly structured systems*. Environmental Modelling in Software. In press, 2000
- [6] Cuadras, C.M.: *Métodos de Análisis Multivariante*. Ed. PPU (Promociones y Publicaciones Universitarias), 1991.
- [7] Daget, J.: *Les modeles mathematiques en ecologia*. Masson, París, 1976



Bibliografía

- [8] Davis, J.M., Eder, B.K. & Bloomfield, P.: *Modelling Ozone in the Chicago Urban Area. In Case Studies in Environmental Statistics.* D. Nychka, W. Piegorsch & L. Cox (Eds.), 1998, 5-26
- [9] Davis, J.M., Eder, B.K. & Bloomfield, P.: *Regional and Temporal Models for Ozone Along the Gulf Coast. In Case Studies in Environmental Statistics.* D. Nychka, W. Piegorsch & L. Cox (Eds.), 1998, 27-50
- [10] Davis, M.: *GComputability, complexity and languages: fundamentals of computers science, 2ª edición.* Academic Press, 1994
- [11] Feller, William: *Introducción a la Teoría de Probabilidades y sus aplicaciones.* Ed. Limusa, 1970
- [12] Fernández, G. & Sáez, F.: *Fundamentos de informática.* Anaya, 1995
- [13] Ferrán Aranaz, Magdalena: *SPSS para Windows. Programación y análisis estadístico.* McGraw-Hill, 1996
- [14] Forrester, Jay W.: *Industrial Dynamics.* Productivity Press, 1961
- [15] Forrester, Jay W.: *Principles of systems.* Productivity Press, 1966
- [16] Forrester, Jay W.: *Urban Dynamics.* Productivity Press, 1969



Bibliografía

- [17] Gmurman, V.E.: *Teoría y problemas de las probabilidades y estadística matemática*.
Moscú Editorial "Mir", 1974
- [18] Goossens, M., Mittelbach, F. & Samarin, A.: *The Latex Companion*.
Addison-Wesley Publishing Company, 1994
- [19] Gross, M. & Lentin, A.: *Notions sur les grammaires formelles*.
Gautier-Villars, París, 1977
- [20] Hanson, Augie: *Aprenda C ya!*
Anaya Multimedia, 1990
- [21] Harrison, M.A.: *Introduction to formal language theory*.
Addison-Wesley, 1978
- [22] Jeffers, John N. R.: *Modelos en ecología*.
Oikos-tau, s.a., 1991
- [23] Kelley, D.: *Teoría de autómatas y lenguajes formales*.
Prentice-Hall, 1998
- [24] Kernighan, Brian W. & Dennis M. Ritchie: *The C Programming Language, 2ª ed.*
Englewood Cliffs, NJ: Prentice Hall, 1998
- [25] Lamport, Leslie: *Latex: A Document Preparation System. User's Guide and Reference Manual*.
Addison-Wesley Publishing Company, 1994
- [26] Loève, Michael: *Teoría de la probabilidad*.
Ed. Tecnos, 1976.



Bibliografía

- [27] Mathsoft: *S-Plus, Version 4.0, User's Guide, Data Analysis*. Products Division, Mathsoft, Seattle, Washington, 1997.
- [28] Mesarovic, M. & Takahara, Y.: *General System Theory. Mathematical Foundations*. Academic Press, New York, 1978
- [29] Microsoft-Anaya Multimedia: *Microsoft C/C++. Manual de referencia de la biblioteca de ejecución*. Anaya Multimedia, 1993
- [30] Norman, E. : *Teoría de la Información y Codificación*. Paraninfo, 1986
- [31] Pérez-Campanero, Juan Antonio: *Guía software de C*. Anaya Multimedia, 1993
- [32] Press, William H.; Teukolsky, Saul A.; Vetterling, William T. & Flannery, Brian P.: *Numerical Recipes in C*. 2ª Edición, Cambridge University Press, 1997
- [33] Ríos, Sixto: *Análisis Estadístico Aplicado*. Ed. Paraninfo, 1976.
- [34] Schildt, Herbert: *C: Manual de referencia, 2ª edición*. Osborne/McGraw-Hill, 1990
- [35] Shoemaker, Ch.: *Mathematical construction of ecological models. Ecosystem Modelling in theory and practice. An introduction with Case Histories*. Jr. John Wiley, 1977



Bibliografía

- [36] Spiegel, Murray R.: *Estadística: teoría y 875 problemas resueltos*. McGraw-Hill, 1974
- [37] Tabachnick, Barbara G. & Fidels, Linda S.: *Using multivariate statistics*. 2nd edn. Harper Collins Publishers, 1989
- [38] Usó-Domenech, J.L., Villacampa, Y., Stürbing, G., Karjalainen, T. & Ramo, M.: *Mariola: A model for calculating the response of mediterranean bush ecosystem to climatic variations*. Ecological Modelling, 113-129, 1995
- [39] Villacampa, Y., Cortés, M., Vives, F., Usó, J.L. & Castro, M.A.: *A new computational algorithm to construct mathematical models*. Ecosystems and sustainable development II. Advances in Ecological Sciences. Wit Press, 323-330, 1999
- [40] Villacampa, Y., Usó, J.L. & Lloret, M.: *Towards a formalization of Ecosystems and ecological models*. Ecosystems and Sustainable Development Advances in Ecological Sciences 1, Computational Mechanics Publications, 1998.
- [41] Villacampa, Y., Usó, J.L., Vives, F. & Lloret, M.: *A populational model of the reproductive behaviour of Mediterranean bushes: A case of Cistus Albidus L.* Ecosystems and Sustainable Development Advances in Ecological Sciences 1, Computational Mechanics Publications, 395-404, 1998
- [42] Villacampa, Y., Usó-Domènech, J.L., Mateu, J., Vives, F. & Sastre, P.: *Generative and recognoscitive grammars of ecologi-*



Bibliografía

- cal models.*
Ecological modelling, In press, 1999
- [43] Villacampa-Esteve, Y. & Usó-Domènech, J.L.: *Mathematical models of complex structural system: A linguistic vision.*
International Journals of General Systems, Vol. 28(1), 37-52, 1999
- [44] Vives, F., Villacampa, Y., Cortés, M. & Usó, J.L.: *An introduction to coding theory of flow equations in Ecological models.*
Ecosud 99, 1999
- [45] Walpole, Ronald E. & Myers, Raymond H.: *Probabilidad y estadística.*
4ª Edición, McGraw-Hill, 1991
- [46] Wiley, Jhone & Son: *Interactive data analysis.*
New York, Londres, 1972
- [47] Yang, Z.B.: *A new model of teneral system theory.*
Cybernetic and System. An international Journal 20, 67-76, 1989
- [48] Zeigler, B.P.: *Multifaceted modelling and discrete event simulation.*
Academic Press. London, 1984