

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UACPyP del CCH

IIMAS

ANALISIS MULTIVARIADO DE DATOS CATEGORICOS

TESIS CON  
FALLA DE ORIGEN

T E S I S

Que para obtener el grado de:

MAESTRA EN ESTADISTICA E  
INVESTIGACION DE OPERACIONES

Presenta la Actuaría

GUILLERMINA ESLAVA GOMEZ

México, D.F.

Septiembre 1986



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## I N D I C E

Capítulo 1 PLANTEAMIENTO DEL PROBLEMA Y METODOLOGIA DE ANALISIS	3
1.1 Planteamiento	3
1.2 Metodologia	4
Capítulo 2 ESCALAMIENTO MULTIDIMENSIONAL	6
2.1 Introduccion	6
2.2 Escalamiento metrico	9
2.3 Escalamiento no metrico	13
2.4 Comparacion del escalamiento metrico y no metrico	23
Capítulo 3 SOLUCION AL PROBLEMA	26
3.1 Descripcion de la informacion	26
3.2 Analisis preliminar	29
3.3 Construccion de matrices de similaridad	30
3.4 Analisis de valores propios de las matrices	32
3.5 Uso del Escalamiento no metrico	33
Capítulo 4 CONCLUSIONES	39
Apéndice A DISEÑO DE REGISTRO	41
Apéndice B GRAFICAS	46

## INTRODUCCION

Las metodologías aplicadas para la cuantificación de resultados en investigaciones experimentales juegan un papel muy importante, y requieren de un entendimiento claro de sus propiedades para poder hacer un uso adecuado de ellas.

El conocimiento de la conducta del neonato es un tema de interés en diversas ramas de la medicina y de ciencias afines. Para ello, se llevan a cabo estudios en los cuales es importante cuantificar mediciones que confirman o no hipótesis establecidas teóricamente; o bien son auxiliares en una etapa que es exploratoria.

En este trabajo se ilustra con un ejemplo real sobre una investigación relacionada con neonatos, el uso de una metodología estadística que no está profusamente difundida.

En un estudio realizado por investigadoras del Departamento de Psicología del Instituto Nacional de Perinatología (INPer), se evaluaron a través de un conjunto de reactivos (propuestos por Brazelton (1973) y llamados "Escala evaluativa del comportamiento neonatal" (ECN)) a un grupo de neonatos clasificados en cuatro subgrupos. Esta clasificación fue hecha de acuerdo a su estado de salud diagnosticado al momento del nacimiento.

Los reactivos (Atkin L. (1980)) que conforman la ECN se agrupan en las clases siguientes: HABITUACION, REGULACION DE ESTADO, RANGO DE ESTADO, REGULACION AUTONÓMICA y PROCESOS NOTORES. Además de medir la respuesta a estos reactivos se midió la respuesta a otro grupo de 20 reactivos que conforman el grupo REFLEJOS. Esto se hizo con el objeto de discriminar entre los diferentes grupos de neonatos clasificados de acuerdo al criterio médico como: SANOS, HIPERBILIRUBINEMICOS, TERMINO y PRETERMINO.

Era importante que esta información fuese organizada de forma tal que hiciera posible a los especialistas, identificar su estructura interna con el fin de determinar el poder discriminatorio, si es que lo hay, de los reactivos aplicados.

Un conjunto de datos como los mencionados es referido dentro de la Estadística con el nombre de "datos categóricos multivariados", ya que no son cuantificables en una escala numérica y son dos o más mediciones por individuo las que se tienen.

Se conocen varias técnicas para estudiar la estructura de datos categóricos, entre ellas se encuentran: las pruebas de independencia basadas en la estadística Ji-cuadrada, los modelos log-lineales y las medidas de asociación.

La elección de la técnica más apropiada, depende desde luego del tipo de datos y del objetivo del estudio. En el caso, objeto de este trabajo se deseaba detectar si algunas variables tenían capacidad para discriminar entre grupos de neonatos, con la finalidad de probar por un lado la consistencia interna del diagnóstico establecido por el médico y por otro el sugerir nuevas relaciones entre las variables.

Debido a la estructura de los datos se encontró que la técnica apropiada es la conocida como "escalamiento no métrico", a través de la cual se pudo discriminar entre algunos grupos de neonatos con solo uno o dos grupos de reactivos.

El trabajo se presenta en cuatro capítulos y dos apéndices en donde se incluye, en uno la descripción de los registros de cada evaluación y en el otro algunas gráficas.

En el primer capítulo se formula de manera detallada el problema y se describe brevemente la técnica utilizada en el análisis.

En el capítulo dos se describen los dos enfoques tradicionalmente conocidos como escalamiento métrico o clásico y escalamiento no-métrico, mostrando que el primero es en cierto modo un caso particular del segundo. Se dan algunas recomendaciones para su uso de acuerdo a las propiedades estructurales de la matriz de similitudes.

En el capítulo tres se presentan los resultados obtenidos del análisis de la información usando el escalamiento multidimensional obteniendo así los grupos de reactivos que discriminan entre los sujetos.

Finalmente en el capítulo cuatro se presentan en forma resumida las conclusiones del trabajo.

## Capítulo 1

### PLANTEAMIENTO DEL PROBLEMA Y METODOLOGIA DE ANALISIS

#### 1.1 Planteamiento

El problema que se desea estudiar es el de la construcción de variables numéricas, que sean capaces de discriminar entre varios grupos de individuos de manera confiable.

Se cuenta con información detallada de una investigación en un grupo de neonatos, nacidos en el Instituto Nacional de Perinatología (INPer). Este grupo está formado por cuatro subgrupos: uno de 29 neonatos que de acuerdo con el diagnóstico médico se consideran sanos, otro de 12 neonatos que al nacer se diagnosticaron sanos pero que presentaron Hiperbilirrubinemia durante el periodo en que fueron medidos y un tercero de 31 neonatos que al nacer fueron canalizados a la unidad de cuidados intensivos por alguna deficiencia determinada por el médico. El cuarto grupo de 34 neonatos que también fueron canalizados a la unidad de cuidados intensivos pero que además, a diferencia de todos los demás, nacieron a pretermino (antes de el periodo de gestación de 40 semanas). Se hace referencia a ellos, con fines puramente operativos para el análisis estadístico, como subgrupos de SANOS, HIPERSILIRRUBINEMICOS, TERMINO y PRETERMINO. Las especificaciones de los criterios de selección aparecen en Atkin(1982) y la descripción de la investigación en Atkin(1980).

Las evaluaciones de cada neonato con las que se cuenta para el análisis que aquí se reporta son: Para los dos primeros subgrupos a edad en días 01, 03, 10, 20 y 30. Para los otros dos subgrupos a edad 01, 10, 20 y 30. La evaluación de cada neonato se hizo aplicando los reactivos que aparecen en el apéndice A.

Las evaluaciones para cada neonato corresponden a mediciones de características que conforman una prueba desarrollada por el Dr. Brazelton(1973), llamada "Escala evaluativa del comportamiento neonatal" (ECN) encaminada a medir características conductuales del neonato. Estas mediciones de acuerdo a las características que reflejan se pueden agrupar en seis grupos llamados: HABITUACION, ORIENTACION, REGULACION DE ESTADO, RANGO DE ESTADO, REGULACION AUTONOMICA y PROCESOS MOTORES. Además de un

grupo de mediciones llamado REFLEJOS.

Entre los objetivos de la investigación está el de conocer si con estas mediciones o con un subgrupo de ellas es posible caracterizar a cada subgrupo de neonatos o a alguno de ellos, así como caracterizar la dinámica de crecimiento del neonato a través del tiempo. Así mismo evaluar la aplicación de la ECN como ayuda al diagnóstico médico para especificar en forma más precisa el estado de salud del neonato.

## 1.2. Metodología

Se cuenta con mediciones múltiples para cada neonato, cada una de ellas tomando valores discretos que guardan un orden. Estas se pueden identificar como variables categóricas multivariadas de acuerdo a la práctica usual en Estadística (Agresti (1980), Plackett (1981)).

Para el análisis estadístico de estos datos, se usan en este trabajo técnicas que mejoran sustancialmente a las técnicas de análisis basadas en medidas que tan solo utilizan la tendencia central y la dispersión, aplicadas tradicionalmente. El análisis estadístico que haremos es con fines exploratorios y descriptivos ya que el conocimiento que se tiene a priori de los datos no permite en principio ajustar algún modelo estadístico.

Debido a la naturaleza de los datos disponibles en este trabajo y al tipo de información que se desea obtener, se decidió utilizar el "Escalamiento multidimensional" como la técnica que ayuda a entender la estructura global de las mediciones.

El enfoque del análisis encabezado por el nombre de Escalamiento multidimensional se ha desarrollado principalmente en áreas sociales y del comportamiento. Especialmente en Psicología, pero su aplicación se ha extendido a otros campos. La idea fundamental es la representación de un conjunto de  $N$  individuos, objetos o características en un espacio  $K$ -dimensional. La configuración de los individuos representados se caracteriza por reflejar en sus "distancias interpersonales" la "estructura cualitativa" de similaridad o parecido. Es decir representaciones "cercanas" reflejan mayor grado de similaridad y representaciones "lejanas" reflejan un grado menor de similaridad.

Existen una gran variedad de metodologías desarrolladas por diversos autores así como una gran cantidad de algoritmos

computacionales. Aun cuando el escalamiento multidimensional no es una metodología que aparece en paquetes estadísticos de cómputo, existen algunos programas desarrollados para su aplicación. No es difícil instrumentar alguna de las técnicas de esta metodología si se cuenta con subrutinas eficientes para encontrar valores y vectores propios de matrices, así como subrutinas de minimización de funciones de varias variables. En el capítulo siguiente se hace una descripción más detallada de estos conceptos.

## Capítulo 2

### ESCALAMIENTO MULTIDIMENSIONAL

#### 2.1. Introducción

Se define el escalamiento multidimensional como un método de representación espacial de un conjunto de objetos. Esta definición es muy general, se da así con el fin de abarcar la gran gama de algoritmos que la satisfacen.

La "representación espacial" involucra la definición de un espacio, que puede ser el euclideo, con una dimensión específica, por ejemplo el plano  $R^2$ .

En la mayoría de los métodos que se presentan bajo el nombre de escalamiento multidimensional se señala que la información requerida para hacer uso de esta metodología consta de medidas de parecido o semejanza entre los objetos a ser representados, generalmente llamadas medidas de similaridad. Se definen estas medidas de acuerdo al contexto en el que aparecen los datos. Esto limita la aplicación, ya que en algunos casos aún cuando no se cuenta directamente con estas medidas, se cuenta con información con la cual es posible construir tales medidas de semejanza o similaridad.

Entre las características de las medidas de similaridad y de la forma en que se relacionan con las distancias interpointuales en la representación espacial se distinguen dos grandes ramas: el escalamiento métrico también conocido como clásico y el escalamiento no métrico. Dedicamos las dos secciones siguientes a cada una de estas ramas.

En relación al escalamiento multidimensional existe gran controversia respecto a su uso. Algunos autores lo menosprecian por ser una técnica puramente descriptiva, además de carecer de un modelo estadístico que permita inferir y medir en términos probabilísticos la bondad de los resultados. Otros autores sin

embargo, manifiestan la gran utilidad de la técnica como herramienta descriptiva y exploratoria al ayudar a entender la estructura de correlación de los datos, aprovechando la universalidad que en términos interpretativos posee la "distancia" y representación gráfica de los objetos bajo estudio.

La idea de esta metodología fue concebida y trabajada inicialmente en el campo de la Psicología, Thurstone en 1925 y Thomson en 1926 la usaron como técnica de escalamiento u ordenamiento de sujetos de acuerdo a puntajes obtenidos en alguna característica. Mas tarde Young y Householder en 1938, retomaron esta metodología logrando un avance considerable gracias al manejo de álgebra matricial. La idea volvió a ser trabajada apareciendo con el nombre de escalamiento métrico o clásico, por un grupo de trabajo encabezado por Torgerson en 1952 en la Universidad de Princeton. Diez años mas tarde y en forma independiente otro grupo de trabajo encabezado por Shepard en "Bell Telephone Laboratories" desarrolla la idea fundamental del escalamiento no métrico, aún cuando Kruskal da un mejor desarrollo matemático de la metodología. A partir de las ideas iniciales dadas por estos dos grupos se han desarrollado e instrumentado diversos algoritmos tanto por los iniciadores como por otros autores.

La característica fundamental de los métodos no métricos en contraste con los métricos es que los primeros toman como información únicamente relaciones de orden entre similitudes y a través de ellas proporcionan información métrica que permite operar matemáticamente con relaciones de orden, suma y producto. Esto es atractivo y útil para algunas ciencias como las Sociales y del Comportamiento, principalmente la Psicología en donde generalmente se cuenta con mediciones cualitativas mas que cuantitativas.

Se presenta con detalle la construcción de un coeficiente de similitud desarrollado por Gower (1971) ya que en el problema que se estudia en el capítulo siguiente, se requiere de estos coeficientes.

El coeficiente de similitud es apropiado para medir la similitud entre dos sujetos con base en  $K$  características que pueden ser cualitativas y cuantitativas, obteniéndose una matriz de similitud para el conjunto de los  $N$  objetos a comparar. Esta matriz tiene la propiedad de ser semidefinida positiva, propiedad necesaria para algunos análisis como el escalamiento multidimensional clásico.

Al ser comparados dos objetos (referidos indistintamente como sujetos, objetos o características)  $i, j$  al ser comparados respecto a una variable  $t$  se les asigna un puntaje  $s(i, j, t)$  con un valor cero cuando se consideran diferentes y con una fracción

positiva menor o igual a 1 cuando presentan algun grado de similitud. La forma de definir este valor depende de la escala de medición de las variables.

La posibilidad de hacer comparaciones entre los individuos  $i$ ,  $j$  respecto al caracter  $t$  lo denotamos por  $r(i,j,t)$  con valor 1 cuando la comparación es posible y cero en otro caso (casos faltantes).

De acuerdo al tipo de variable la similitud se evalua como se indica a continuación.

a) Cuando la  $t$ -ésima variable es cualitativa, evaluamos  $s(i,j,t) = 1$  si los dos individuos toman el mismo valor y  $s(i,j,t) = 0$  si toman valores distintos.

b) Cuando la  $t$ -ésima variable es cuantitativa y para los  $N$  sujetos a comparar toma los valores  $x_1, \dots, x_k$ . Entonces  $s(i,j,t) = 1 - (|x_i - x_j|) / R(t)$ . Donde  $R(t)$  es el rango de la variable, puede ser el muestral o el poblacional. Se observa que alcanza su valor máximo 1 cuando los sujetos toman el mismo valor y su valor mínimo cuando toman valores  $x_1, x_k$  opuestos (0 si se considera el rango muestral).

Una vez definidos los puntajes de comparación para cada variable, el coeficiente de similitud entre dos sujetos es el promedio de los puntajes en las  $K$  variables:

$$S(i,j) = \sum_{t=1}^K s(i,j,t) * r(i,j,t) / \sum_{t=1}^K r(i,j,t)$$

Cuando  $r(i,j,t) = 0$  para todas las variables, el coeficiente queda indefinido. Cuando esto sucede en situaciones prácticas, se dan estimaciones de estos valores (moda, media, cero, etc).

En algunos estudios las variables consideradas para la comparación pueden ser de diversos grados de interés. Esto se puede considerar al definir el coeficiente como un promedio ponderado con pesos  $w(t)$  positivos:

$$S(i,j) = \sum_{t=1}^K s(i,j,t) * r(i,j,t) * w(t) / \sum_{t=1}^K r(i,j,t)$$

Bajo estas definiciones la matriz obtenida de las comparaciones entre el grupo de sujetos tiene la propiedad de ser semidefinida positiva. Cuando hay valores faltantes no se garantiza esta propiedad. La demostración algebraica aparece en el artículo antes mencionado.

Para el caso en que la similitud se calcule con base en un

conjunto de variables con diferentes escalas de medición es posible definir un coeficiente de similitud como un promedio de los coeficientes calculados dentro de cada subgrupo (dicotómicas, cualitativas y cuantitativas) de variables. La matriz obtenida de esta forma satisface la propiedad de ser semidefinida positiva.

## 2.2 Escalamiento métrico

Como se señala en la sección anterior, la característica fundamental de este enfoque es el considerar que las similitudes con las cuales se trabaja aceptan un manejo de orden, suma y multiplicación. Aún cuando la idea inicial la desarrollo Torgerson (1952) siguieron a él varios avances con un manejo más claro desde el punto de vista matemático y que, excepto por un paso previo el método corresponde a dar las coordenadas de  $N$  puntos en un espacio  $K$  dimensional a partir de una matriz de similitudes que son funciones de la distancia interpuntual. Se hace una presentación de la metodología siguiendo los lineamientos generales de Bower (1971) señalando al final, el paso previo al que se refiere Torgerson. Este último hace una presentación de la técnica como metodología de aplicación en el campo de la Psicología, Torgerson (1958).

Se parte de la suposición de que se cuenta con una matriz  $S$  cuyos elementos corresponden a similitudes entre un conjunto de  $N$  sujetos, objetos o características; en caso de no ser así se supone que se cuenta con información de la cual se puede derivar la matriz  $S$ . Dada esta matriz queremos hallar las coordenadas de  $N$  puntos  $P_1, \dots, P_N$  en un espacio euclidiano  $k$  dimensional de tal forma que se tenga una relación funcional  $f$  entre las distancias y las similitudes:

$$d(i, j) = f(s(i, j))$$

$d(i, j)$  = distancia entre el punto  $P_i$  y  $P_j$

$s(i, j)$  = similitud entre individuo  $i$  y el individuo  $j$ .

Las coordenadas de un punto  $P_i$  determinadas por el  $i$ -ésimo componente de cada uno de los  $N$  vectores propios de la matriz  $S$  satisfacen ciertas propiedades que permiten hacer una representación en subespacios del espacio generado por la matriz  $S$ . La representación en un subespacio de dimensión menor se obtiene con las proyecciones de los puntos sobre los ejes

coordenados dados por los vectores propios asociados a los valores propios de mayor magnitud.

Se requiere de algunas características para la matriz  $S$  y de ciertas restricciones sobre el escalamiento de los vectores propios además de una interpretación de las distancias interpuntuales.

Supongase que  $S$  es una matriz simétrica con  $N$  valores propios  $V_1, V_2, \dots, V_N$ , y vectores propios asociados  $C_1, C_2, \dots, C_N$ , como se muestra en la tabla siguiente:

	$V_1$	$V_2$	$\dots$	$V_N$
$P_1$	$C_{11}$	$C_{12}$	$\dots$	$C_{1N}$
$P_2$	$C_{21}$	$C_{22}$	$\dots$	$C_{2N}$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$P_N$	$C_{N1}$	$C_{N2}$	$\dots$	$C_{NN}$
	$C_1$	$C_2$	$\dots$	$C_N$

Si se toma cada renglón de la matriz de eigenvalores como las coordenadas del punto  $P_i$  en un espacio  $N$ -dimensional, la distancia entre dos puntos  $P_i, P_j$  está dada por:

$$d(i, j) = \left( \sum_{k=1}^N (c_{ik} - c_{jk})^2 \right)^{1/2} \dots \dots (1)$$

Normalizando la magnitud de los vectores propios de tal forma que la suma del cuadrado de sus elementos sea igual a su valor propio, la matriz  $S$  se puede expresar como:

$$S = C_1 C_1^T + \dots + C_N C_N^T$$

cuyos elementos corresponden a:

$$s(i, j) = \sum_{k=1}^N c_{ik} c_{jk}$$

Haciendo la sustitución adecuada la distancia interpuntual se puede expresar como:

$$d(i,j) = \{s(i,i) + s(j,j) - 2 * s(i,j)\}^{1/2}$$

Es decir la relación funcional entre los elementos de la matriz S y las distancias interpuntuales esta dada por:

$$f(s(i,j)) = 2 - 2 * s(i,j) \dots \dots (2)$$

ya que  $s(i,i) = s(j,j) = 1$  por ser S matriz de similaridad.

Cuando se cuenta directamente con la distancias interpuntuales  $d(i,j)$ , se construye la matriz S con elementos  $s(i,j) = -d(i,j)/2$  de esta forma se encuentran las coordenadas de los puntos con las distancias interpuntuales dadas.

Es factible encontrar una representación en un espacio de dimensión menor a N cuando algunos de los valores propios de la matriz S son de magnitud pequeña, ya que si  $V_k$  es de magnitud pequeña la contribución de  $(c_{ik} - c_{jk})^2$  a la distancia interpuntual como se indica en la ecuación (1) será pequeña y puede ser considerada insignificante.

En el caso que  $V_k$  sea de magnitud "mayor" puede suceder que  $(c_{ik} - c_{jk})^2$  sea pequeña. Entonces las contribuciones de mayor peso en las coordenadas de un punto son aquellas en las que  $V_k$  es grande y además con una variación "grande" en los elementos del vector propio correspondiente.

Generalmente el vector propio asociado al valor propio de magnitud mayor contiene elementos con una variación muy pequeña, estos contribuyen con una cantidad constante a los elementos de S, esto sugiere que para fines prácticos se haga una transformación en la matriz S.

Si R es la matriz S transformada, se obtienen valores y vectores propios distintos y por tanto puntos con coordenadas diferentes, sin embargo cuando esta transformación corresponde a una traslación y una rotación ortogonal las distancias interpuntuales no cambian. Se sugiere la transformación siguiente, cuya matriz transformada garantiza tener al menos un valor propio igual a cero.

$$b(i,j) = s(i,j) - s(i, \cdot) - s(\cdot, j) + s(\cdot, \cdot) \dots \dots (3)$$

$$s(i, \cdot) = \sum_{t=1}^N s(i,t) / N$$

$$s(i, j) = \sum_{t=1}^N s(t, j) / N$$

N = número de renglones de la matriz S.

observese que las distancias interpuntuales son las mismas:

$$b(i, i) + b(j, j) - 2 * b(i, j) = s(i, i) + s(j, j) - 2 * s(i, j)$$

Los elementos de la matriz transformada suman cero por renglones y por columnas, por lo que el cero es un valor propio con el vector propio (1, ..., 1) asociado y, como cualquier otro vector propio debe ser ortogonal a este entonces satisface que el producto punto con el vector de unos es cero o lo que es lo mismo la suma de los elementos de cada uno de los vectores propios asociados a la matriz transformada es cero. De esta formase encuentra que la solución que mejor se ajusta a la relación funcional (2) es la configuración de puntos en un espacio K-dimensional cuyas coordenadas corresponden a los renglones de los K valores propios asociados a los K valores propios de magnitud mayor de la matriz S transformada.

El procedimiento se puede resumir de la forma siguiente:

a) Transformar la matriz S como se indica en la ecuación (3).

b) Encontrar los vectores propios con norma igual a su valor propio asociado, (los valores propios deben estar ordenados de mayor a menor magnitud) como se muestra en la tabla siguiente:

	$V_1$	$V_2$	...	$V_K$	...	$V_N$
$P_1$	$c_{11}$	$c_{12}$		$c_{1K}$		$c_{1N}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$P_N$	$c_{N1}$	$c_{N2}$		$c_{NK}$		$c_{NN}$

c) Las coordenadas de los puntos de la configuración solución en un espacio K-dimensional, corresponden a los renglones de la tabla anterior con las K columnas elegidas.

El método presentado se conoce también con el nombre de "Coordenadas principales". Torgerson (1958), inicialmente supuso que se cuenta con una matriz B de distancias "perturbadas" y no necesariamente con una matriz semidefinida positiva, por lo tanto

estima una constante  $c$  tal que al ser sumada a cada elemento de  $B$  produce una matriz positiva semidefinida con el mayor número posible de valores propios cero. Los métodos son equivalentes cuando la matriz de la que se parte es semidefinida positiva.

Una medida empírica de la bondad del ajuste en  $K$  dimensiones es la diferencia o la razón entre la suma de los  $K$  valores propios de magnitud mayor y la suma total de los  $N$  valores propios.

La idea del método se ha generalizado, al considerar que las distancias interpuntuales y las similitudes pueden estar relacionadas por cualquier función, no necesariamente lineal:

$$d(i, j) = f(s(i, j))$$

con  $f$  una función polinomial, exponencial, o simplemente lineal. La solución son las coordenadas de los puntos en un espacio  $K$  dimensional previamente establecidos que minimizan la suma de cuadrados:

$$S = \sum_{i,j}^N (d(i, j) - f(s(i, j)))^2 / \sum_{i,j}^N d^2(i, j)$$

en donde, el mínimo se calcula sobre las configuraciones de puntos  $P_i$   $i = 1, \dots, N$  y los parámetros de la función  $f$ .

Un grupo encabezado por J. de Leeuw, F. W. Young y Y. Takane presentaron la idea, minimizando una función  $S(P, f)$  sobre el espacio de los puntos  $P$  y los parámetros de la función  $f$ . De acuerdo a la función  $S$  y a los espacios sobre los cuales se minimiza, esta técnica es llamada de diversas maneras. Ver por ejemplo de Leeuw J. y cols (1976), Young F. W. y cols. (1976).

Se han desarrollado algoritmos para esta metodología, como el paquete que permite elegir la función ligé entre distancias y similitud llamada KYST, desarrollado por Kruskal, Young, Shepard, y Torgeron. También existen subrutinas para esto en el paquete HOMALS disponible en cinta magnética en el Departamento de Estadística del IIMAS-UNAM.

### 2.3 Escalamiento no métrico

La técnica descrita en la sección anterior supone que para

encontrar la configuración espacial de N puntos que represente a los N objetos, sujetos o características, la función liga entre las similaridades  $s(i,j)$  y las distancias interpuntuales  $d(i,j)$ :

$$d(i,j) = f(s(i,j))$$

debe ser una función en donde sea válido sumar y multiplicar con las similaridades. Esta función generalmente es una polinomial, exponencial, o lineal. Otro tipo de escalamiento es aquel en el que la relación liga entre similaridades y distancias es una función monótona decreciente. En este caso la única información que se maneja es el orden de las similaridades. No se consideran directamente sus valores numéricos y por esto recibe el nombre de "no métrico". Como se menciona en la introducción de este capítulo, la idea la concibió Shepard (1962); Kruskal (1964a, 1964b) la presenta con una mayor formalización matemática. Se han hecho mejoras al algoritmo, únicamente desde el punto de vista computacional ya que la idea fundamental es la misma, esto gracias a los avances logrados en la instrumentación de subrutinas eficientes desde el punto de vista numérico.

Se describe a continuación la técnica presentada por Kruskal (1964a, 1964b) y se mencionan algunos paquetes de cómputo desarrollados para este fin; señalando además que es factible la instrumentación de la misma con cierta facilidad si se cuenta con subrutinas de minimización de funciones y de ordenamiento eficientes.

Es conveniente ver esta técnica como un problema de ajuste. Es decir dadas las similaridades deseamos encontrar la configuración cuyas distancias interpuntuales mejor ajustan a dichas similaridades. El "mejor ajuste" se refiere a la relación monótona entre similaridades y distancias medida a través de una medida que más adelante se define como "Stress". Se puede resumir la presentación del método como un proceso iterativo que consiste básicamente de las etapas siguientes:

#### 1 FORMACION DE LA CONFIGURACION INICIAL

Dado un conjunto de  $M = N(N - 1) / 2$  similaridades, se requiere de una configuración inicial que después de un proceso iterativo nos lleve a una solución satisfactoria en términos del stress. Esta configuración inicial debe contener puntos diferentes, no estar contenida en un espacio de dimensión menor al preestablecido y se puede proporcionar de tres formas distintas:

a) Si se tiene conocimiento a priori de algún candidato de acuerdo al conocimiento del comportamiento de los sujetos bajo estudio, es conveniente dar la configuración utilizando este conocimiento.

b) Si no se tiene candidato y se requiere por alguna razón práctica de "fijar" la solución final (ya que esta depende de la inicial), es usual dar una configuración inicial de puntos distribuidos uniformemente en el subespacio de dimensión K, situados alrededor del origen. Shepard (1952) recomienda dar los vértices de un simplejo unitario en K dimensiones:

$$\begin{aligned}
 P_1 & (1, 0, 0, \dots, 0, 0), \\
 P_2 & (0, 1, 0, \dots, 0, 0), \\
 P_3 & (0, 0, 0, \dots, 0, 1), \\
 P_4 & (2, 0, 0, \dots, 0, 0), \\
 P_5 & (0, 2, 0, \dots, 0, 0), \\
 P_6 & (0, 0, 0, \dots, 0, 2), \\
 & \vdots \\
 & \vdots \\
 & \text{etc.}
 \end{aligned}$$

c) Dar una configuración generada aleatoriamente.

Algunos autores sugieren dar la configuración obtenida por algún algoritmo de escalamiento métrico como configuración inicial, esto es porque proporciona una aproximación lineal a la configuración deseada. No obstante en el problema visto en este trabajo se decidió usar el simplejo unitario como configuración inicial por simplicidad en la instrumentación y porque la naturaleza de los datos no exigió un análisis detallado en el primer paso. Esto se manifiesta en la convergencia satisfactoria del método propuesto.

## II NORMALIZACIÓN LA CONFIGURACION

Se normaliza la configuración con centro en el origen y magnitud media de las coordenadas de los puntos centrados igual a 1.

## III CALCULO DE LAS DISTANCIAS INTERPUNTALES

En el cálculo de las  $M = N(N - 1) / 2$  distancias interpuntuales, la distancia o métrica usual es la euclídeana:

$$d(i, j) = \left\{ \sum_{k=1}^M (x_{ik} - x_{jk})^2 \right\}^{\frac{1}{2}}$$

Algunos autores argumentan que desde el punto de vista de sus aplicaciones ofrece una mejor interpretación. Tal vez esto se deba a que dicha métrica es intuitivamente la más aceptable. Otros autores argumentan también en su favor, su facilidad de manejo dadas sus propiedades específicas de métrica formal.

Una familia de métricas conocidas como métricas de Minkowski o métricas  $L_p$  corresponde a:

$$d(i, j) = \left\{ \sum_{k=1}^M (x_{ik} - x_{jk})^p \right\}^{\frac{1}{p}}$$

La mayoría de los paquetes de cómputo tienen instrumentada esta familia de métricas, es común que cuando no se especifica el valor de  $p$  se asigne automáticamente  $p = 2$ , lo que da la distancia euclídeana.

La descripción del método se hace suponiendo una métrica  $L_p$ , excepto cuando se indique lo contrario. En la aplicación descrita en el capítulo siguiente, los cálculos se hicieron con la distancia euclídeana.

#### IV REGRESION MONOTONA

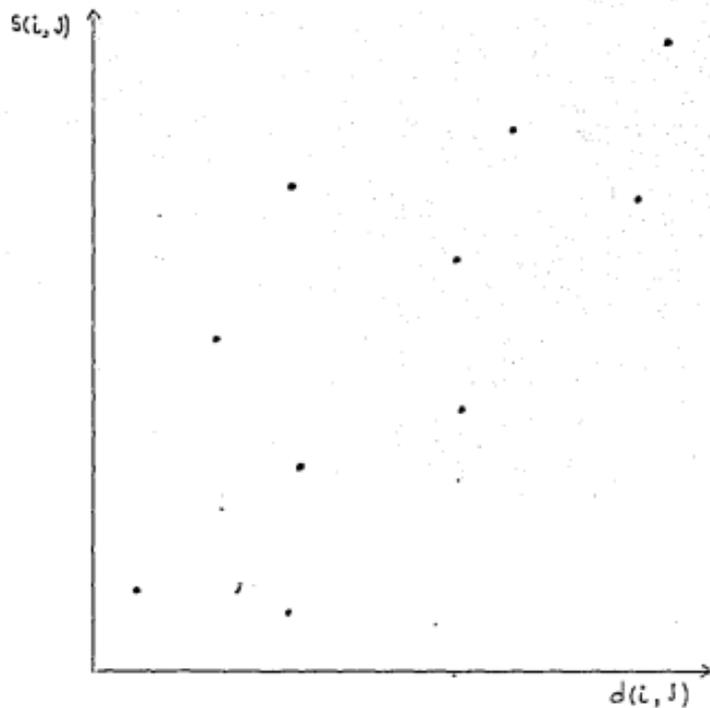
El cálculo de una regresión monótona entre distancias y similitudes es la parte fundamental que refleja las propiedades específicas del escalamiento no métrico. La regresión monótona también conocida como regresión isotónica aparece ampliamente descrita en Barlow R. E. (1972) en donde se señalan aplicaciones diversas e incluso se hacen descripciones del algoritmo que usa Kruskal (1964b) que es el que se usa en los cálculos para el problema de este trabajo.

El problema en esta etapa se reduce al problema de encontrar, dado un conjunto de similitudes ordenadas, un conjunto de distancias interpuntuales que guarden el mismo orden que las similitudes correspondientes.

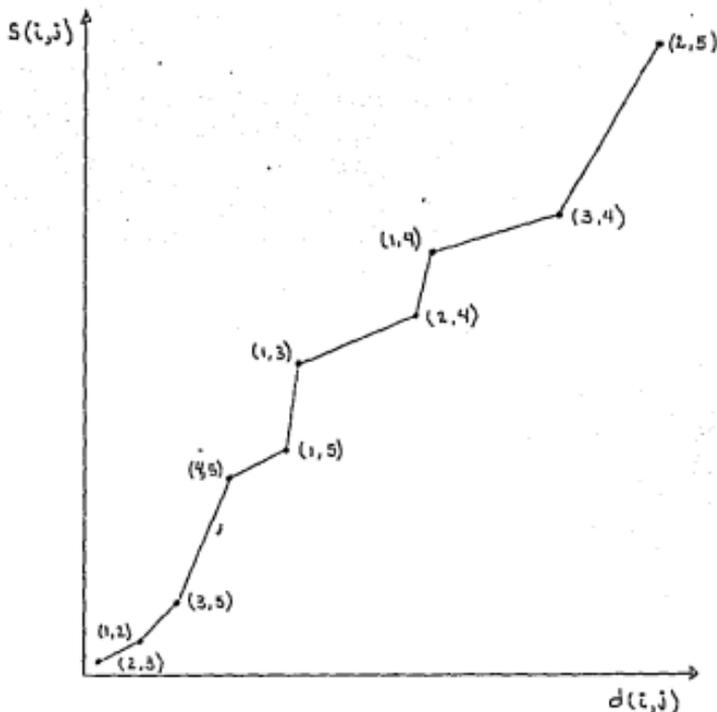
La configuración cuyas distancias tengan el orden de las similitudes puede no existir, y por esto se dan aproximaciones a estas distancias. El grado de estas aproximaciones se mide por el stress. Se da la idea de regresión monótona en términos geométricos, de esta manera se puede establecer la similitud con la técnica común del análisis de regresión.

Dado el conjunto de  $M$  similitudes ordenadas y el conjunto de distancias interpuntuales de una configuración, se pueden

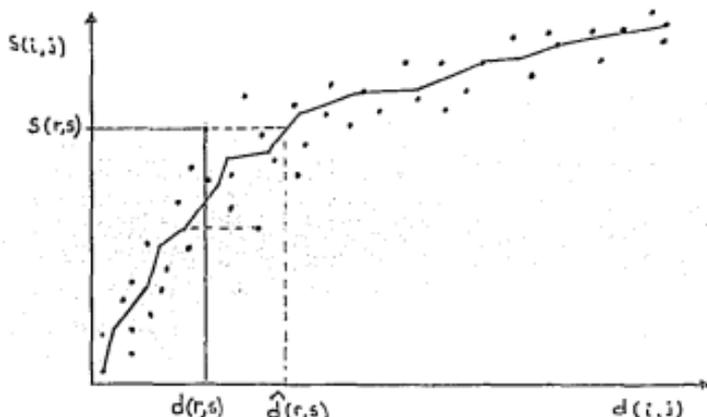
graficar en el plano de ejes coordenados como sigue:



Cada punto  $(i,j)$  tiene abscisa  $d(i,j)$  y ordenada  $s(i,j)$ , un ajuste perfecto se da cuando el orden en las distancias es el mismo que el de las similitudes. Gráficamente lo anterior implica que la poligonal que resulta de unir los puntos se obtiene "moviéndose" siempre a la derecha y hacia arriba como se muestra en la gráfica siguiente



Como no siempre sucede que se pueda tener una poligonal como la descrita en el párrafo anterior es necesario trazar una curva que satisfaga el "moverse" siempre hacia la derecha y hacia arriba y la medida de cuánto difiere del ideal resulta importante. Gráficamente:



El camino para medir las diferencias entre la curva ajustada y los puntos, se calcula con las diferencias sobre el eje de las distancias y no sobre el de las similitudes ya que la información que están proporcionando es únicamente de orden. En el ajuste se pueden localizar dos conjuntos de puntos; los originales  $(s(i,j), d(i,j))$  y los ajustados  $(s(i,j), \hat{d}(i,j))$ .

Las abscisas de los puntos de la curva ajustada, a las que por abuso de lenguaje se llaman "distancias" ajustadas  $\hat{d}(i,j)$ , no es otra cosa más que una secuencia de números que satisfacen el orden de las similitudes asociadas y que son además, los más parecidos a la secuencia de las distancias dadas. Estos números generalmente no representan distancias, o lo que es equivalente no existe una configuración de puntos cuyas distancias interpuntuales correspondan a la secuencia de "distancias" ajustadas.

Evidentemente existen una infinidad de secuencias de números que satisfacen la restricción de orden de las similitudes. Se elige para el ajuste, aquella secuencia que se parece más a la de las distancias dadas. Este parecido lo medimos como usualmente se hace en estadística, con la suma de cuadrados de las diferencias, estandarizada por la suma del total de los cuadrados de las distancias (para hacerla independiente de la escala en que estas distancias se expresan). A esta medida se le llama

stress:

$$\text{stress} = \frac{\sum_{i,j=1}^N (d(i, j) - \hat{d}(i, j))^2}{\sum_{i,j=1}^N d^2(i, j)}$$

El stress mide que tan bien se relacionan monótonamente las distancias de la configuración propuesta con las similitudes observadas. Como se señala anteriormente, esta medida refleja las diferencias en el eje de las distancias y el ajuste se elige de una familia no paramétrica de curvas monótonas ascendentes.

#### V MINIMIZACION SOBRE EL ESPACIO DE PUNTOS

Una vez que se tiene una forma cuantitativa para medir la bondad del ajuste de una configuración se desea encontrar la configuración que mejor ajuste a las similitudes, es decir aquella con stress mínimo. Redefiniendo esta medida como una función de las distancias expresadas en términos de sus coordenadas:

$$S(x_1, x_2, \dots, x_n, \dots, x_m)$$

La idea del método es mover los puntos de la configuración actual de forma tal que sus distancias tengan una mejor relación monótona con las similitudes y por tanto un stress menor.

El problema matemático corresponde a un problema de minimización esto es, el encontrar los valores de las coordenadas de los puntos que minimicen la función stress:

$$\text{Min}_{\{x_{ij}\}} \text{stress}(x_1, \dots, x_m)$$

Actualmente existen subrutinas, instrumentadas que pueden usarse para hallar el mínimo de una función. Algunas se encuentran disponibles en la biblioteca del Departamento de Análisis Numérico del IIMAS, UNAM. En Kruskal (1964b) se sugiere el método de gradientes para hallar el mínimo; este método ha sido modificado por varios autores obteniéndose algunos más eficientes.

#### VI ITERACION Y CONVERGENCIA

Una vez obtenida la nueva configuración dada por las coordenadas que minimizan el stress como función de ellas, se recalcula la regresión monótona con las nuevas distancias interpuntuales, como se indica en la etapa IV. Se continúa con el

proceso iterativamente hasta hallar una configuración satisfactoria.

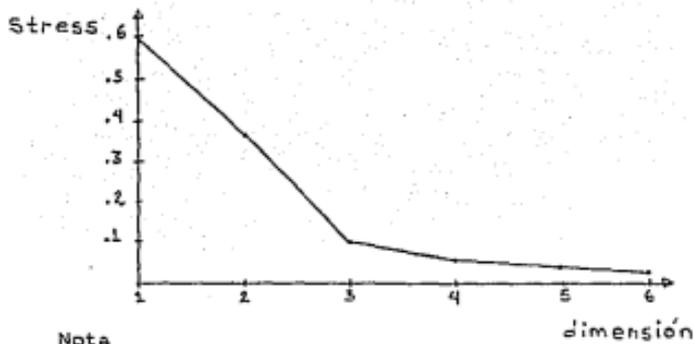
El criterio para establecer la cota mínima del stress que indique la configuración satisfactoria no está determinado por límites cuantitativos. Kruskal (1964) dió un criterio empírico en el que establece la siguiente escala:

STRESS	BONDAD DE AJUSTE
.20	Pobre
.10	Regular
.05	Bueno
.025	Excelente
0.00	Perfecto

Este criterio, establecido con base en simulaciones, se aleja mucho de el valor interpretativo de las configuraciones con stress mayor a 0.20, Shepard (1974) señala que este criterio de evaluación dirige a los usuarios de la técnica a dar un peso determinante al stress como elemento para evaluar la solución final, esta medida es parte de la evaluación que debe ser considerada junto con la información que en términos interpretativos pueda proveer la configuración. Es factible obtener configuraciones utiles en términos interpretativos, que reflejen el comportamiento de la estructura de los datos, aún cuando su stress sea mayor al 0.20. Por lo anterior se sugiere que el stress, aunque es una medida importante, no debe ser considerada como una medida global de la bondad de la solución del problema.

En el desarrollo de la técnica se supone fija la dimensión del espacio. La elección de esta dimensión depende básicamente del conocimiento a priori que se tenga de la investigación de la cual proceden los datos y a posteriori, de la interpretación y el stress asociado a la configuración solución.

Respecto al conocimiento a posteriori del stress se han hecho algunas sugerencias útiles para determinar la dimensión, una de ellas es aplicar la técnica con distintos valores de la dimensión, hacer una gráfica del stress asociado contra la dimensión y elegir la dimensión cercana al lugar en el que la curva se estabiliza; en otras palabras la abscisa correspondiente al punto en donde la curva presenta el "codo" mas pronunciado, como se ilustra en la gráfica siguiente:



En la etapa IV correspondiente al ordenamiento de las similaridades y cálculo de la regresión monótona, puede suceder que existan similaridades iguales. Se han planteado dos caminos para tratarlas en el ajuste de la regresión monótona. En el primero las distancias ajustadas correspondientes a las similaridades iguales, son diferentes. En el segundo las distancias ajustadas correspondientes a las similaridades iguales también son iguales. En Kruskal (1964a, 1964b) a las similaridades iguales se les llama "empates" y los dos caminos de tratamiento para el ajuste les llama enfoque Primario y enfoque Secundario, a continuación damos una descripción breve de cada uno de ellos.

a) Al ordenar las similaridades; se forman los bloques donde toman un valor constante, en cada bloque se asigna un orden interno de acuerdo a las distancias correspondientes; se continúa con el algoritmo usual. De esta forma las distancias ajustadas dentro de cada bloque toman valores diferentes, expresado formalmente se establece:

$$s(i, j) < s(r, s) \text{ implica } d(i, j) < = d(r, s)$$

Es decir la regresión monótona ajusta una función monótona no decreciente.

b) Dentro de cada bloque se asigna como distancias el promedio de las distancias correspondientes a las similaridades dentro del grupo; cada bloque se considera para fines de ajuste, como el valor constante de la similaridad con su correspondiente distancia promedio dentro del bloque. Así, las distancias ajustadas dentro de cada uno de ellos toman el mismo valor, expresado formalmente se establece:

$s(i, j) < s(r, s)$  implica  $d(i, j) < d(r, s)$

y además

$s(i, j) = s(r, s)$  implica  $d(i, j) = d(r, s)$ .

La regresión monótona ajusta una función estrictamente creciente.

Finalmente el considerar que la función liga sea una función monótona no decreciente, permite captar relaciones lineales y no lineales entre las similaridades y la distancia, lo que permite obtener una configuración en un subespacio de dimensión menor que el que se obtiene por el procedimiento métrico. Sin embargo este es un método más caro computacionalmente.

#### 2.4 Comparación del escalamiento métrico y no métrico

Para comparar estas dos técnicas de escalamiento multidimensional, se hace una presentación desde un punto de vista geométrico. De esta forma se ve claramente que el escalamiento métrico es un enfoque más restrictivo que el no métrico. Por esta razón este último resulta ser más general, aplicable a matrices de similaridad no necesariamente simétricas ni semidefinidas positivas.

Recuérdese que el objetivo de la técnica es representar un conjunto de  $N$  sujetos por una configuración de puntos  $P_1, \dots, P_N$  en un espacio de dimensión  $K$ , de tal suerte que las distancias interpuntuales guarden una relación funcional con las similaridades correspondientes. Para elegir los puntos y la relación funcional se define la función stress que depende de las coordenadas de los puntos y de la relación expresada como función, como una suma de cuadrados de diferencias:

$$\text{stress} = \sum_{\substack{i,j=1 \\ i < j}}^N [d(i, j) - f(s(i, j))]^2 / \sum_{\substack{i,j=1 \\ i < j}}^N d^2(i, j)$$

El problema se reduce a determinar los puntos  $P_1, \dots, P_N$  en un espacio de dimensión  $K$  y la función  $f$ , que minimizan a  $S(P_1, \dots, P_N, f)$  sobre las configuraciones en el espacio  $K$  dimensional y un espacio de funciones. Desde el punto de vista matemático se hacen las siguientes observaciones:

a) La función stress S, específicamente llamada funcional por depender de otra función, es derivable excepto cuando  $P_i = P_j$  para algún  $i \neq j$ , es decir sobre los planos  $P_i = P_j$  en el espacio K dimensional y cuando  $P_1 = \dots = P_p$  la funcional toma el valor infinito. Esta observación debe ser considerada en el algoritmo de minimización.

b) Una alternativa de solución es especificar el espacio de funciones por una familia paramétrica, por ejemplo

$$f(s) = a + b * s$$

$$f(s) = a + \exp(b * s)$$

con a, b, c parametros reales y s la similaridad entre sujetos.

c) En el enfoque de escalamiento métrico el problema se resuelve en forma aproximada, al buscar las coordenadas principales con  $K = N$  o cuando K son los valores propios distintos de cero; la función específica corresponde a:

$$f(s(i, j)) = d(i, j) = 2 - 2 * s(i, j)$$

En cuyo caso la funcional toma el valor mínimo cero.

Cuando K corresponde a los K valores propios de magnitud mayor y es estrictamente menor a N, la solución corresponde a una configuración muy cercana a la configuración donde la funcional alcanza el mínimo.

Sin embargo cuando los valores propios de la matriz son muy parecidos y los K primeros (los de magnitud mayor) no representan la mayor parte de la variación total, la solución no es una buena aproximación a la solución que minimiza la funcional. En este caso la función lineal ya no es satisfactoria y debe buscarse otra diferente.

El método no métrico resuelve el problema de minimizar la funcional S sobre las configuraciones en el espacio K dimensional y sobre el espacio de las funciones monótonas. Comparativamente el espacio de las funciones monótonas es menos restrictivo a un espacio de funciones paramétricas, por esta razón el enfoque no métrico es más general.

Una forma de solucionar el problema de minimización en el espacio de funciones monótonas es usando el método del gradiente. Eligiendo un punto de partida se calcula la funcional S y se va moviendo por pasos en la dirección del gradiente para hallar un mínimo. En cada paso se mueve tanto en el espacio de puntos como en el espacio de funciones monótonas.

Una forma mas simple de aplicar el método del gradiente para minimizar la funcional  $S$ , es como se describio en la seccion de escalamiento no métrico. Es decir minimizar alternativamente: en un paso moverse en el espacio de las funciones (calcular la regresion monótona) dejando fija la configuración y en otro paso dejar fija la función y moverse en el espacio de los puntos en el espacio  $K$  dimensional.

Finalmente puede decirse que para elegir el tipo de enfoque debe de tomarse en cuenta la estructura de la matriz de similaridades. Cuando esta es semidefinida positiva y ademas los  $K$  primeros valores propios (ordenados en orden decreciente de magnitud) representan la mayor parte de la variación, es recomendable usar el escalamiento métrico; ya que en este caso se obtiene una buena aproximación al mínimo de la funcional. En cualquier otro caso es recomendable usar el escalamiento no métrico ya que garantiza una solución mejor en terminos de un stress mas pequeno.

\*\*\*\*\*

## Capítulo 3

### SOLUCION AL PROBLEMA

#### 3.1 Descripción de la información

Como se señaló en el primer capítulo, se desea analizar la información de forma tal que permita entender la estructura de correlación de las variables. Para esto es necesario examinar los datos a fin de que estas variables se vean claramente. En este capítulo aplicamos la técnica de escalamiento multidimensional con fines descriptivos y exploratorios que permitan sugerir nuevas hipótesis, así como entender la estructura de los datos para poder aplicar metodología estadística que requiere de suposiciones específicas.

Se comienza por describir la información disponible y el análisis preliminar al escalamiento multidimensional.

La información disponible para el análisis se encuentra en cinta magnética, la descripción de los registros aparece en el apéndice A. Esta información fue recolectada en los años 1981-1982. El número de neonatos aparece en el cuadro siguiente, clasificados de acuerdo al subgrupo al que pertenecen y de acuerdo a la edad expresada en días:

Edad (días)	01	03	10	15	20	30	45
SANDS	29	29	29	-	29	29	29
HIPER	12	12	12	-	12	12	12
TERMINO	16	-	-	27	-	30	-
PRETERMINO	24	-	-	29	-	30	-

El grupo de neonatos evaluados a edad 01 son los mismos que se evalúan hasta edad 45, además de algunos neonatos que se evalúan adicionalmente al grupo inicial.

Con el fin de manejar en forma resumida y de acuerdo a interpretaciones clínicas, las mediciones (reactivos) se han dividido en 6 grupos: HABITUACION, ORIENTACION, REGULACION DE ESTADO, RANGO DE ESTADO, REGULACION AUTONOMICA y PROCESOS MOTORES, además de los registros correspondientes a identificación de los neonatos. Para algunos de los reactivos se tomó registro de dos calificaciones: la modal y la óptima. La primera corresponde a la respuesta mas frecuente en el neonato al aplicarle varias veces el reactivo. La segunda a la respuesta considerada óptima por la persona que evaluó al neonato (Atkin(1980)). Existe controversia sobre el tipo de calificación que mejor refleja la conducta del neonato. Sería interesante hacer un análisis con las dos calificaciones para verificar si aportan información diferente. Sin embargo el análisis se hace con las calificaciones registradas como óptimas, dejando abierta la posibilidad de trabajar con las calificaciones modales en análisis posteriores.

Los grupos de reactivos (variables) están conformados de la forma siguiente:

#### HABITUACION

- V41 Decremento luz
- V42 Decremento sonaja
- V43 Decremento campana
- V44 Decremento pinchazo

#### ORIENTACION

- V45 Orientación pelota
- V47 Orientación sonaja
- V49 Orientación visual y aud. inanimado
- V51 Orientación cara
- V53 Orientación voz
- V55 Orientación cara y voz
- V57 Alerta

V80 Calidad de respuesta

REGULACION DE ESTADO (REGESTADO)

V61 Acunamiento

V64 Consolarse con intervencion

V78 Actividad para autocalmarse

V79 Facilidad mano-boca

RANGO DE ESTADO (RANGO)

V66 Punto máximo de excitacion

V67 Rapidez para excitarse

V68 Irritabilidad

V77 Cambios de estado

V82 Irritabilidad general

REGULACION AUTONÓMICA (REGAUTON)

V74 Temblores

V75 Sobresaltos

V76 Cambios de color

FUNCIONAMIENTO MOTOR (MOTOR)

V58 Tono general

V59 Madurez motriz

V60 Esfuerzo para sentarse

V63 Movimientos defensivos

V71 Actividad

El grupo de REFLEJOS comprende 17 reflejos de tipo neurológico registrados de la V13 a la V32

El número que aparece como etiqueta de las variables corresponde al número de variable en la descripción de los registros en el apéndice A.

Las variables toman valores enteros como calificación: de 1 a

6 para las variables de REFLEJOS; entre 1 y 10 para las variables comprendidas entre V45 y V55; entre 1 y 9 para el resto de las que se mencionan.

### 3.2 Análisis preliminar

Con el fin de depurar la información, se calcularon las frecuencias de cada variable, con esto se detectaron algunos errores de registro, que una vez identificados en algunos casos fue posible recuperar la información correcta y en otros se declaró valor faltante (registros incorrectos o ausentes).

Se calcularon el número de valores faltantes para cada subgrupo de neonatos y para cada edad. Se observó lo siguiente:

En las variables del grupo HABITUACION el porcentaje de valores faltantes fue mayor al 50%, se decidió eliminar del análisis a este grupo, cualquier estimación para estos casos sería muy burda.

La variable con un porcentaje que fluctúa entre el 5 y el 39% de valores faltantes es V64 de REGULACION DE ESTADO, se decidió sustituir los valores faltantes por sus correspondientes valores modales por cada edad y por cada subgrupo y así no eliminarla del análisis.

Para el resto de las variables los porcentajes de valores faltantes no superaron el 15%, excepto para el subgrupo HIPERBILIRRUBINEMICOS en la edad 20 donde no se evaluaron a tres de doce, es decir el 25% de valores faltantes. En estos casos se decidió sustituirlos por valores modales, esto con el fin de no reducir el tamaño de la muestra.

Se calcularon los valores modales dentro de cada edad y cada subgrupo; se observó lo siguiente:

En las variables del grupo REFLEJOS no se observaron cambios de edad a edad. Para los subgrupos SANDS e HIPERBILIRRUBINEMICOS, todos los reflejos tienen 2 como valor modal, excepto para V15 con valor modal 4, y V25 con valor 1; para los de PRETERMINO y TERMINO todos los reflejos tienen valor modal 2, excepto en V15 y V24 con valor modal 4 y V25 con 1. No se detectaron neonatos con Reflejos que indicaran "anormalidad", considerando esta información se decidió analizar únicamente las variables que conforman la Escala de Conducta Neonatal.

Las diferencias que se pueden ver para el resto de variables son poco marcadas, para el grupo ORIENTACION aparece diferencia si se comparan SANOS e HIPER contra TERMINO y PRETERMINO solamente en la edad 30. Entre estos dos pares de subgrupos tambien se observan diferencias en las edades que se comparan para las variables V63 del grupo MOTOR; V67 de RANGO ; V61, V64, V79 de REGESTADO; V74 de REGAUTON.

Este análisis sugiere trabajar con cinco grupos de variables: ORIENTACION, REGESTADO, RANGO, REGAUTON y MOTOR con calificaciones "óptimas" y sustituyendo los valores faltantes por valores modales dentro de cada edad y cada subgrupo muestral.

### 3.3 Construcción de matrices de similitud

Con el fin de representar a los neonatos como puntos en un espacio euclideo de dimensión K se requieren medidas de similitud entre individuos. No se cuenta directamente con ellas pero es posible derivarlas a partir de las variables disponibles, para esto es necesario considerar lo siguiente:

a) La configuración de puntos proporcionada por el Escalamiento multidimensional esta centrada en el origen, lo que sugiere aplicar la técnica al grupo total para fines de comparación y no a cada subgrupo como podría suponerse.

b) Se desea asignar puntajes a los individuos respetando los agrupamientos a priori de las variables. Esto implica construir medidas de similitud por cada grupo de variables para la muestra total.

c) El escalamiento multidimensional se va a aplicar al grupo total; en las edades 01 y 30 se tienen mediciones para el total. En las edades restantes no se pueden comparar porque las evaluaciones se hicieron a edades distintas, sin embargo de acuerdo a comentarios de la Dra. L. Atkin (investigadora del INPer) es de interés comparar los subgrupos a edad 15 con un lado con los de edad 10 y por otro con los de edad 20.

En total se construyeron 20 matrices de similitud, una para cada grupo de variables y para cada edad.

El número de neonatos disponibles para cada edad determina en cada caso la dimensión de las matrices. En el cuadro siguiente se presenta el número de neonatos clasificados por subgrupo y por edad.

	EDAD 01	EDAD 10 o 15	EDAD 20 o 15	EDAD 30
SANDS	29	29	29	29
HIPER	12	12	12	12
TERMINO	16	27	27	30
PRETERMINO	24	29	29	32
Total	81	97	97	103

Cada matriz contiene los coeficientes de similaridad entre  $i$  y el neonato  $j$  ( $S(i,j)$ ). Este coeficiente se construyó de acuerdo al coeficiente de similaridad propuesto por J. Gower, descrito en el capítulo anterior. Para cada edad y cada subgrupo de variables se obtiene como sigue:

$$s(i,j,k) = 1 - |x(i,k) - x(j,k)| / R(k)$$

donde  $x(i,k)$  corresponde al valor de la variable  $k$  en el neonato  $i$ .

$R(k)$  corresponde al rango teórico de la variable  $k$ .

Tomando el promedio sobre todas las variables ( $K$ ), de cada subgrupo se tiene:

$$S(i,j) = \sum_{r=1}^K s(i,j,r) / K$$

Así por ejemplo la matriz correspondiente a la edad 01 y al grupo ORIENTACION, es una matriz de 81 renglones y 81 columnas; cada elemento  $s(i,j)$  denota la similaridad entre el neonato  $i$  y el neonato  $j$ , respecto al grupo ORIENTACION. Recuerdese que estas matrices tienen la propiedad de ser simétricas definidas semipositivas.

### 3.4 Análisis de valores propios de las matrices

Como un análisis preliminar de las matrices para decidir el enfoque apropiado del escalamiento multidimensional, se calcularon los valores propios de cada matriz de similitud centrada en el origen, es decir de la matriz transformada:

$$b(i,j) = S(i,j) - S(i, \cdot) - S(\cdot, j) + S(\cdot, \cdot)$$

$$\text{donde } S(i, \cdot) = \sum_{j=1}^N S(i,j) / N$$

$$S(\cdot, j) = \sum_{i=1}^N S(i,j) / N$$

$$S(\cdot, \cdot) = \sum_{i,j=1}^N S(i,j) / N * N$$

N = número de renglones = número de columnas de la matriz.

Al analizar la magnitud relativa de cada valor propio, se observó que en todas las matrices la magnitud relativa del valor propio mayor, fue menor a 0.45. La suma de los dos valores propios mayores fue de 0.55. Es decir con una o con dos dimensiones se expresa a lo mas el 55% de la variabilidad en las matrices de similitud. En la tabla siguiente se presenta el porcentaje de variabilidad que explican los i primeros valores propios, i = 1, 2, 3.

EDAD 01	ORIENTACION	REGESTADO	RANGO	REGAUTON	MOTDR
1	33	24	40	28	18
2	39	36	50	43	34
3	45	48	57	57	54
/					
EDAD 10 o 15					
1	33	23	45	34	22
2	41	42	52	50	36
3	49	55	59	64	48
EDAD 20 o 15					
1	31	23	32	32	19

2	38	51	47	55	33
3	44	52	54	64	45
EDAD 30					
1	35	26	47	34	21
2	42	44	59	53	35
3	48	55	59	64	46

Como puede observarse de acuerdo a las magnitudes de los valores propios mas grandes una aproximación en tres dimensiones es muy pobre. Dado que se desea reducir la dimensión para tener una gráfica interpretable, es recomendable aplicar el enfoque no métrico.

Para tener idea de las configuraciones obtenidas con el escalamiento métrico, se obtuvieron las configuraciones en una y dos dimensiones, por ser estas las dimensiones que nos permiten manejar un valor "representante" de cada grupo de variables, y a lo mas dos.

Los puntos obtenidos aparecen muy proximos (debido a la estructura de los valores propios), la distribución espacial de ellos proporciona información muy pobre, solamente algunos casos aparecen separados. En el apéndice se presenta una gráfica típica de estas configuraciones (gráfica B.1).

Como sugiere la estructura de los valores propios de las matrices de similitud y lo confirman las representaciones obtenidas, el escalamiento métrico no es adecuado para la información que se maneja. Por ello se aplica el enfoque no métrico.

### 3.5 Uso del Escalamiento no métrico

Para el cálculo de estas soluciones se usó un programa desarrollado por el Act. Manuel Iraizos Lopez (Iraizos L. M. (1984)), disponible en el Departamento de Estadística del IIMAS, UNAM. Fue necesario hacer algunas modificaciones para mejorar su eficiencia en terminos numéricos y operativos, así

como para solucionar otros problemas no previstos cuando fue elaborado.

Recordando que entre los objetivos, esta el de tener representaciones espaciales en la menor dimension posible, se obtuvieron representaciones con cada matriz en una y dos dimensiones. Procediendo como se indico en el capítulo dos se obtuvieron las configuraciones solución. En la tabla siguiente se muestra el valor del stress asociado a cada una de ellas.

#### STRESS ASOCIADO A LA CONFIGURACION FINAL

##### DIMENSION UNO

Edad	ORIENTACION	REGESTADO	RANGO	REGAUTON	MOTOR
01	.2038	.3696	.3500	.3892	.4706
10 y 15	.3988	.4135	.1959	.3852	.4218
20 y 15	.2242	.4819	.5516	.3619	.4189
30	.2754	.4949	.4193	.3962	.4214

##### DIMENSION DOS

01	.1390	.2188	.1195	.1621*	.2120*
10 Y 15	.1148*	.2203	.1276	.1391*	.1911*
20 Y 15	.1210	.1980	.1636	.1439	.2209
30	.1206	.2018	.0952	.1266	.2169

Al interpretar los resultados recuérdese que la funcional Stress depende tanto de las coordenadas de los puntos como de una función liga. La configuración solución es aquella que minimiza esta funcional junto con una función monótona, el proceso de minimización se lleva a cabo alternando dos etapas hasta converger u obtener una solución satisfactora, en terminos interpretativos.

Una de las etapas en el proceso de minimización consiste en aplicar la regresión monótona entre similaridades y distancias interpuntuales; la otra en minimizar el Stress sobre el espacio de las coordenadas de los puntos; esto usando el método del gradiente y, como se menciona en el capítulo anterior, cuando dos puntos son iguales la función no es diferenciable. Esto trae como consecuencia que en el proceso de minimización cuando dos puntos coinciden, se deben perturbar adecuadamente sus posiciones para continuar con el proceso a través de las zonas no diferenciables.

La solución que se busca como la óptima corresponde a la de mínimo global, esta solución raras veces se obtiene (de hecho no se conoce ningún método para resolver el problema de mínimo global). Las soluciones que se obtuvieron, seguramente en la mayoría de los casos corresponden a soluciones donde el Stress alcanza un mínimo local, no obstante proporcionan información que resulta ser directamente interpretable.

Los valores marcados con un asterisco en la tabla anterior corresponden a valores del stress cuya configuración asociada son puntos donde la función no es diferenciable. Desde el punto de vista teórico esta configuración puede estar alejada de un mínimo, sin embargo consideramos que la aproximación es satisfactoria por su interpretación.

De acuerdo a los valores del stress observamos lo siguiente:

1) Para el grupo ORIENTACION, el stress toma los valores más pequeños y en el grupo MOTOR el stress toma los valores más grandes.

2) Considerando dos dimensiones, el stress se reduce proporcionalmente en un 50% respecto al que se obtiene con una dimensión.

3) para RANGO el stress es donde tiene una variación mayor de edad a edad, sobre todo en una dimensión.

Considerando una dimensión se obtienen cinco variables para cada neonato,  $A_1, A_2, A_3, A_4, A_5$ , correspondiendo a los subgrupos ORIENTACION, REGESTADO, RANGO, REGAUTON, MOTOR respectivamente. Se graficaron por pares, etiquetando a los puntos de acuerdo al grupo al que pertenece el neonato que representa.

\* = SAND

2= HIPER

3= TERMINO

#### 4º PRETERMINO

En el apéndice B se presentan las gráficas a las que se hace referencia en este capítulo.

Para el grupo de neonatos a edad 01 se observa lo siguiente:

1) En las gráficas donde aparece la variable ORIENTACION las poblaciones se confunden, hay dos casos, uno de TERMINO y otro de PRETERMINO en donde la variable toma valores muy diferentes. (gráfica B.2)

2) La variable representante de RANGO al igual que la anterior no tiene poder discriminatorio. (gráfica B.3)

3) Las variables REGESTADO y REGAUTON separan a los subgrupos SANOS y PRETERMINO. Los grupos HIPER y TERMINO no se distinguen y aparecen algunos casos aislados del grupo total. (gráfica B.4)

4) La variable MOTOR al ser graficada con REGESTADO separa a los grupos SANOS de PRETERMINO; el grupo HIPER se mezcla con el grupo sanos (excepto por dos casos que aparecen con los PRETERMINO. (gráfica B.5)

Para el grupo de edad 10 y 15 se observa:

1) La variable ORIENTACION no discrimina a los subgrupos de neonatos, sin embargo aparecen cinco neonatos del grupo TERMINO en donde la variable toma valores en un extremo y dos neonatos del grupo HIPER donde la variable toma valores en el otro extremo. (gráfica B.6)

2) La variable REGESTADO discrimina a los subgrupos SANOS e HIPER de los subgrupos PRETERMINO y TERMINO. (gráfica B.7)

3) Las variables RANGO y REGAUTON no tienen poder discriminativo. (gráfica B.8)

4) En la variable MOTOR hay dos neonatos del grupo HIPER en donde la variable toma valores extremos. (gráfica B.9)

Para el grupo de edad 20 y 15

1) La variable ORIENTACION no discrimina, sin embargo aparecen seis neonatos del grupo TERMINO que se aíslan del grupo total. (gráfica B.10)

2) La variable RANGO discrimina SANOS de PRETERMINO, el grupo HIPER se mezcla con el de SANOS y el de TERMINO aparece disperso en la población total. (gráfica B.11)

3) La variable RANGO y MOTOR son las que mejor discriminan al

subgrupo SANOS de PRETERMINO, el grupo HIPER se mezcla con el de SANOS. (grafica B.12)

4) Las variables REGESTADO y REGAUTON discriminan, con menos rigor que las dos anteriores SANOS de PRETERMINO, el grupo HIPER se mezcla con el de SANOS; el grupo de TERMINO se dispersa en el grupo total. (grafica B.13)

Para el grupo de edad 30 observamos:

1) La variable ORIENTACION no discrimina a los subgrupos de neonatos; aparecen dos neonatos del grupo TERMINO en donde la variable toma valores extremos y uno del de HIPER donde la variable toma un valor en el otro extremo. (grafica B.14)

2) La variable RANGO no tiene poder discriminante, sin embargo al ser graficada contra la variable ORIENTACION aparece un subgrupo de ocho neonatos del grupo PRETERMINO que se aísla del grupo total. (grafica B.15)

3) La variable REGESTADO discrimina SANOS de PRETERMINO, el subgrupo de HIPER se agrupa con el de SANOS, pero el de TERMINO aparece disperso en el grupo total. (grafica B.16)

4) Las variables MOTOR y REGAUTON no discriminan a los subgrupos de neonatos, aunque existe tendencia a separarse SANOS e HIPER de PRETERMINO; los del grupo TERMINO se dispersan en el grupo total. (grafica B.17)

5) El grupo TERMINO aparece disperso en todas las graficas, ninguna variable lo discrimina. (grafica B.18)

De este analisis gráfico podemos decir que las variables MOTOR, REGESTADO, REGAUTON y RANGO son las que mejor discriminan entre los subgrupos SANOS e HIPER. La variable ORIENTACION no muestra poder discriminatorio pero muestra algunos neonatos aislados que es importante identificarlos. El subgrupo TERMINO aparece disperso en el grupo total es decir ninguna variable lo discrimina. El grupo HIPER tiende a mezclarse con el grupo de SANOS.

A este grupo de cinco variables para la edad 01 se le aplico un analisis de componentes principales, los resultados confirman lo anterior, las variables MOTOR, REGESTADO y REGAUTON tienen mayor representacion en el primer componente principal; ORIENTACION aparece representado en el segundo componente; RANGO aparece representada con pesos similares en los dos componentes. B.19

Al hacer las graficas de los cinco componentes en parejas, se observa que solamente el primer componente tiene poder discriminatorio entre SANOS y PRETERMINO, los HIPER se mezclan en su mayoría con los SANOS y el grupo TERMINO se mezcla en su

mayoría con el grupo PRETERMINO.

Se hizo el mismo análisis gráfico para las diez variables obtenidas al considerar dos dimensiones en el análisis multidimensional. De esta forma cada grupo de variables ORIENTACION, REGESTADO, RANGO, REGAUTON y MOTOR quedan representadas por dos variables. Del análisis se observó que la información que refleja cada uno de estos grupos con un solo representante se refleja también con uno de los dos representantes; la otra variable representante no parece reflejar a primera vista información adicional, esto mismo se ve reflejado al aplicar análisis de componentes principales a las diez variables.

Es posible que las diez variables obtenidas al considerar dos dimensiones proporcionen, información adicional a la que proporcionan las cinco que se obtuvieron al considerar una dimensión; esto a priori no se sabe. Se requiere de un mayor grado de conocimiento del y de un análisis estadístico minucioso. El grupo de las cinco permite un manejo más sencillo de los neonatos por ser un grupo más reducido.

# ESTA TESIS NO DEBE SALIR DE LA BIBLIOTECA

## Capítulo 4

### CONCLUSIONES

a) Desde el punto de vista de la metodología de análisis, se puede decir que el enfoque métrico es recomendable cuando la matriz de similitudes es simétrica, semidefinida positiva y el número de valores propios que representan "la mayor parte" de la variación es comparable con la dimensión requerida para la representación; también cuando la configuración asociada es útil en términos interpretativos. En estas circunstancias el uso del análisis no métrico proporciona resultados similares, con la desventaja de ser computacionalmente más caro. En otro caso como se explicó en el capítulo dos, es recomendable usar el enfoque no métrico.

b) Desde el punto de vista médico, del análisis gráfico se observa que los resultados reflejan la situación siguiente:

i) Los grupos de variables que mejor discriminan o distinguen a los subgrupos de neonatos son MOTOR, REGULACION DE ESTADO Y REGULACION AUTONOMICA.

ii) El grupo ORIENTACION no tienen poder discriminatorio intragrupal, pero sí distingue algunos casos aislados del subgrupo al que pertenecen.

iii) El grupo RÁNGO en la edad 10 o 15 tiene poder discriminatorio; para el grupo a edad 30 aísla a un subgrupo de neonatos que pertenecen al subgrupo TERMINO.

iv) Los grupos SANOS y de alto riesgo PRETERMINO son los que gráficamente aparecen separados, como se esperaba ya que son los que más difieren de acuerdo al diagnóstico médico.

v) El subgrupo HIPERBILIRRUBINEMICOS tiende a agruparse con el subgrupo SANOS sobre todo en los neonatos a edad 30.

vi) El subgrupo TERMINO se dispersa entre el grupo total.

vii) Se identifican algunos neonatos que se aíslan del subgrupo al que pertenecen.

Del análisis gráfico se sugiere lo siguiente:

a) Identificar y examinar a los neonatos que se presentan aislados del subgrupo al que pertenecen o del grupo total.

b) Considerar información adicional de los neonatos del grupo HIPER, posiblemente edad exacta en la que presentaron hiperbilirrubinemia, con el fin de poder caracterizar a los neonatos que se agrupan con los del grupo SANOS y a aquellos que tienden a mezclarse con los de PRETERMINO.

c) Considerar información adicional de los neonatos del grupo de TERMINO, ya que como se observo es un grupo heterogeneo que aparece disperso en el grupo total, posiblemente sea recomendable considerar su periodo de gestación y su peso al nacer.

Las representaciones gráficas pueden aportar información adicional a las investigadoras que conocen el aspecto médico de los neonatos, ya que poseen información a priori de los neonatos así como del problema en general.

Desde el punto de vista de análisis estadístico y con fines descriptivos y exploratorios, se sugiere hacer un análisis gráfico con las diez variables obtenidas con el escalamiento multidimensional al considerar dos dimensiones, podrían aportar información adicional; hacer un análisis de componentes principales con el grupo de cinco variables y otro con el de diez y hacer un análisis gráfico de los componentes principales; aplicar el escalamiento multidimensional unicamente a los grupos de SANOS y PRETERMINO, con el fin de explorar su interrelación.

Finalmente se puede decir que en este trabajo se obtuvieron basicamente dos resultados: El primero es el de confirmar la consistencia del diagnostico en los casos claramente discriminables en la practica medica usual. El segundo sugerir nuevas relaciones no aparentes en el diagnostico original.

Apéndice A

DISEÑO DE REGISTRO.

ARCHIVO E.C.N.

No. VAR	COLS.	CONTENIDO (DESCRIPCION)	CLAVE	RANGO	DATOS FALTANTE
1	1-5	Número de expediente	IDEN		-
2	6-7	Sexo	SEXO		-
3	8-9	Edad Clave (días)	EDAD	01-60	-
4	10	Renglón	RENG	1-2	-
5	11-16	Fecha de examen	FECHA		000000
6	17-20	Hora de inicio	HORIN	0001-2359	9999
7	21-22	Duración de Exámen (minutos)	DUR	00-99	-
8	23-26	Hora de Alimentación	HORAL	0001-2359	9999
9	27-28	Examinador	EXAM	01-04	99
10	29	Estado Inicial	EDDIN	1-5	9
11	30	Estados Predominantes (1)	EDD01	1-6	9
12	31	Estados Predominantes (2)	EDD02	1-6	9
13	32	Presión Plantar	PRENPLA	1-6	9
14	33	Presión Palmar	PRENPAL	1-6	9
15	34	Clonus	CLON	1-6	9
16	35	Babinski	BABIN	1-6	9

17	36	Parado	PARAD	1-6	9
18	37	Marcha	MARCHA	1-6	9
19	38	Apoyo	APOYO	1-6	9
20	39	Incurvación	INCURV	1-6	9
21	40	Movimientos de arrastre	MOVARR	1-6	9
22	41	Glabella	GLAB	1-6	9
23	42	Desviación Tónica de Ojos y Cabeza	DESVTD	1-6	9
24	43	Nistagmus	NISTAG	1-6	9
25	44	Reflejo Tónico Cervical	RETOCER	1-6	9
26	45	Moro	MORO	1-6	9
27	46	Reflejo de búsqueda	BUSO	1-6	9
28	47	Reflejo de chupeteo	CHUP	1-6	9
29	48	Movimientos pasivos brazo derecho	MPBRAD	1-6	9
30	49	Movimientos pasivos brazo izquierdo	MPBRAI	1-6	9
31	50	Movimientos pasivos pierna derecha	MPPID	1-6	9
32	51	Movimientos pasivos pierna izquierda	MPPII	1-6	9
33	52	Atractivo	ATTRACT	0-8	9
34	53*	Variables interferentes	VARINT	0-3	9
35	54	Necesidad de estimulación	NECEST	0-8	9
36	55	Actividades para autocalmarse (mano-boca)	ACTA01	0-1	9
37	56	Actividades para autocalmarse (chupeteo sin nada en la boca)	ACTA02	0-1	9
38	57	Actividades para autocalmarse (estímulos visuales y/o auditivos)	ACTA03	0-1	9

39	58	Actividades para autocalmarse (con cambios posturales)	ACTA04	0-1	9
40	59	Actividades para autocalmarse (por razones no observables)	ACTA05	0-1	9
41	60-61	Decremento luz	DECLUZ	01-10	00
42	62-63	Decremento sonaja	DECSON	01-10	00
43	64-65	Decremento campana	DECCAM	01-10	00
44	66-67	Decremento pinchazo	DEPIN	01-10	00
45	68-69	Orientación pelota (optima)	PELOP	01-10	00
46	70-71	Orientación pelota (modal)	PELMD	01-10	00
47	72-73	Orientación sonaja (optima)	SONOP	01-10	00
48	74-75	Orientación sonaja (modal)	SONMD	01-10	00
49	76-77	Orientación visual y aud. inanimado (optima)	VAUDOP	01-10	00
50	78-79	Orientación visual y aud. inanimado (modal)	VAUDMD	01-10	00
RENGLON 2					
51	11-12	Orientación cara (optima)	CARADP	01-10	00
52	13-14	Orientación cara (modal)	CARAMD	01-10	00
53	15-16	Orientación voz (optima)	VOZDP	01-10	00
54	17-18	Orientación voz (modal)	VOZMD	01-10	00
55	19-20	Orientación cara y voz (optima)	CVOZDP	01-10	00
56	21-22	Orientación cara y voz (modal)	CVOZMD	01-10	00
57	23	Alerta	ALERT	1-9	0
58	24	Tono general	TONGEN	1-9	0
59	25	Madurez motriz	MADMDT	1-9	0
60	26	Esfuerzo para sentarse	ESFSENT	1-9	0

61	27	Acunamiento (optimo)	ACUNDP	1-9	0
62	28	Acunamiento (modal)	ACUNMO	1-9	0
63	29	Movimientos defensivos	MOVDEF	1-9	0
64	30	Consolarse con intervención (optimo)	CONSDP	1-9	0
65	31	Consolarse con intervención (modal)	CONSMO	1-9	0
66	32	Punto máximo de excitación	PMAXEX	1-9	0
67	33	Rapidez para excitarse	RAPAEX	1-9	0
68	34	Irritabilidad	IRRIT	1-9	0
69	35	No. de est. que produjeron irritabilidad	NIRRIT	0-8	9
70	36	No. de est. aplicados	ESTAP	1-8	9
71	37	Actividad	ACT	1-9	0
72	38	Actividad espontánea	ACTESP	0-3	9
73	39	Actividad provocada	ACTPRO	0-3	9
74	40	Temblores	TEMBL	1-9	0
75	41	Sobresaltos	SOBRES	1-9	0
76	42	Cambios de color	CAMBCOL	1-9	0
77	43	Cambios de estado	CAMEDO	1-9	0
78	44	Actividad para autocalmarse	AUTCAL	1-9	0
79	45	Facilidad mano-boca	FACMABO	1-9	0
80	46	Calidad de respuesta.	CALRESP	1-9	0
81	47	Persistencia del examinador	PERS	1-9	0
82	48	Irritabilidad general	IRRITGE	1-9	0
83	49	Valor reforzante	VALREF	1-9	0
84	50-51	Sonrisas	SONRIS	00-99	-

B5 B0 Variable indicadora

VARIND 0-4\*

- \* 0 sano
- 1 sano hiperbilirrubinemicos
- 2 nosano término
- 3 nosano pretérmino

Apéndice B

GRAFICAS



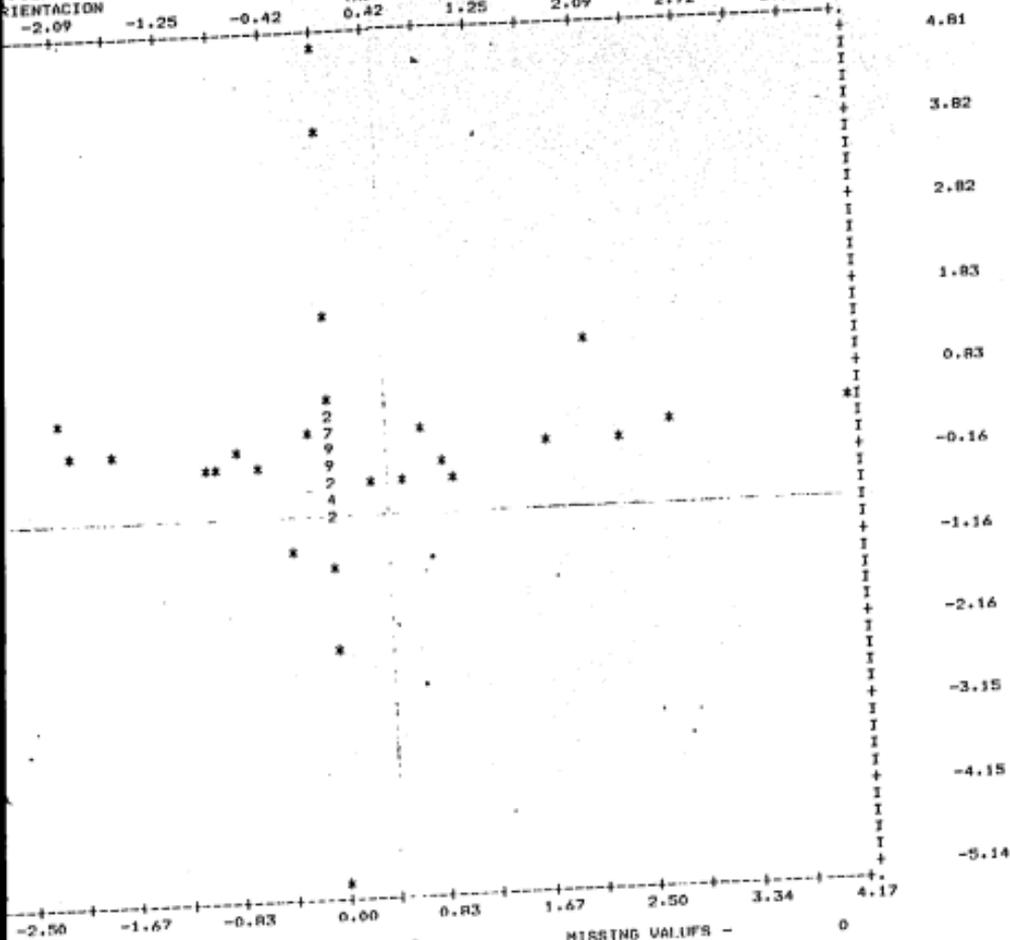
7/05/86)  
ORIENTACION  
-2.09

(ACROSS) A3  
0.42 1.25

RANBO  
2.09

2.92

3.76



EXCLUDED VALUES-

MISSING VALUES -

0

0

ORDINAL 31

B.2

09/04/86

PAGE 4

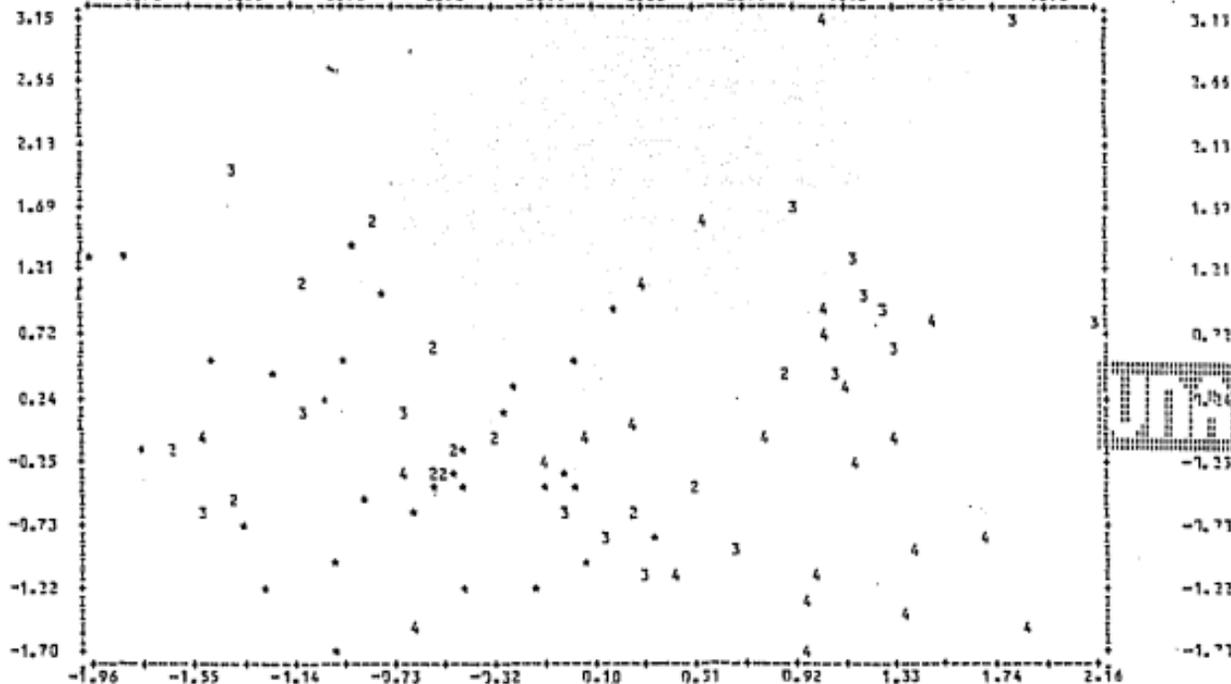
FILE NAME  
SCATTERGRAM OF

CREATION DATE = 09/04/86  
(COUNT) A1 ORIENTATION  
-1.76 -1.35 -0.93

(ACROSS) A2  
0.30 0.71

REGISTADO  
1.15

1.54 1.95



PLOTTS) VALUES - 197

EXCLUDED VALUES - 0

MISSING VALUES - 0

ORDINAL 31

B.3

09/04/86

PAGE 5

37

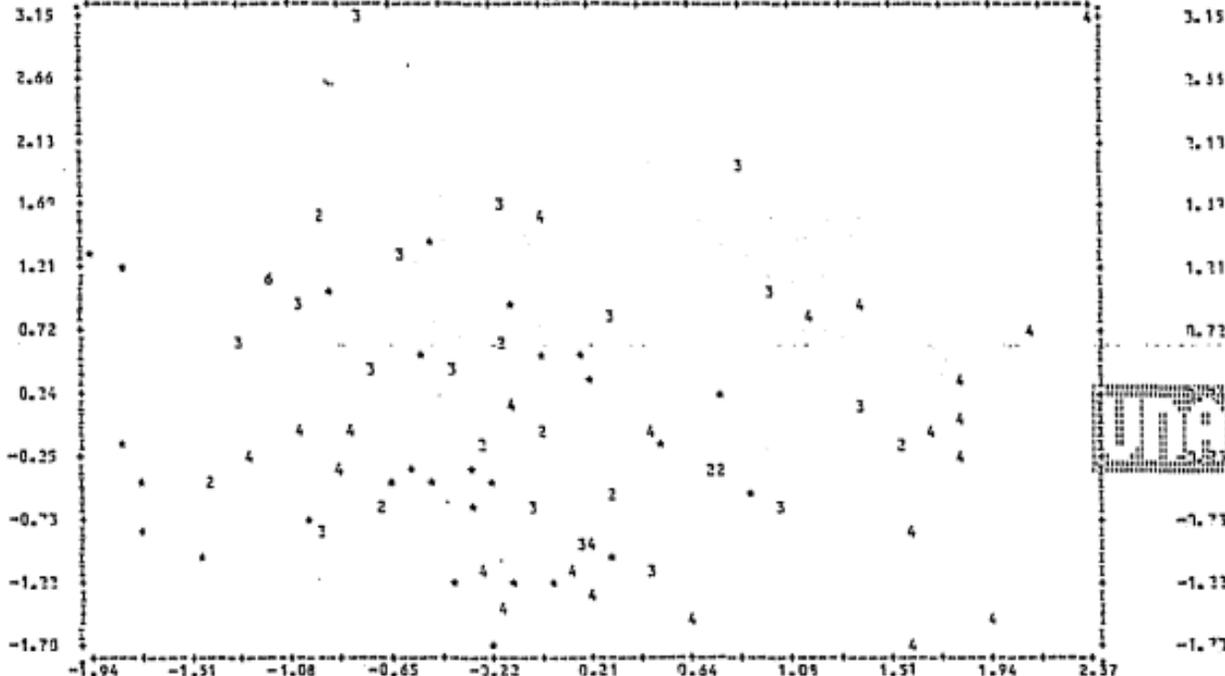
FILE NAME  
SCATTERGRAM OF

CREATION DATE = 09/04/86  
EQUIMS #1 ORIENTATION

(ACROSS) A3

RANGE

-1.73 -1.30 -0.86 -0.43 -0.00 0.43 0.86 1.29 1.72 2.15



0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24
25	26	27	28	29
30	31	32	33	34
35	36	37	38	39
40	41	42	43	44
45	46	47	48	49
50	51	52	53	54
55	56	57	58	59
60	61	62	63	64
65	66	67	68	69
70	71	72	73	74
75	76	77	78	79
80	81	82	83	84
85	86	87	88	89
90	91	92	93	94
95	96	97	98	99
100	101	102	103	104
105	106	107	108	109
110	111	112	113	114
115	116	117	118	119
120	121	122	123	124
125	126	127	128	129
130	131	132	133	134
135	136	137	138	139
140	141	142	143	144
145	146	147	148	149
150	151	152	153	154
155	156	157	158	159
160	161	162	163	164
165	166	167	168	169
170	171	172	173	174
175	176	177	178	179
180	181	182	183	184
185	186	187	188	189
190	191	192	193	194
195	196	197	198	199
200	201	202	203	204
205	206	207	208	209
210	211	212	213	214
215	216	217	218	219
220	221	222	223	224
225	226	227	228	229
230	231	232	233	234
235	236	237	238	239
240	241	242	243	244
245	246	247	248	249
250	251	252	253	254
255	256	257	258	259
260	261	262	263	264
265	266	267	268	269
270	271	272	273	274
275	276	277	278	279
280	281	282	283	284
285	286	287	288	289
290	291	292	293	294
295	296	297	298	299
300	301	302	303	304
305	306	307	308	309
310	311	312	313	314
315	316	317	318	319
320	321	322	323	324
325	326	327	328	329
330	331	332	333	334
335	336	337	338	339
340	341	342	343	344
345	346	347	348	349
350	351	352	353	354
355	356	357	358	359
360	361	362	363	364
365	366	367	368	369
370	371	372	373	374
375	376	377	378	379
380	381	382	383	384
385	386	387	388	389
390	391	392	393	394
395	396	397	398	399
400	401	402	403	404
405	406	407	408	409
410	411	412	413	414
415	416	417	418	419
420	421	422	423	424
425	426	427	428	429
430	431	432	433	434
435	436	437	438	439
440	441	442	443	444
445	446	447	448	449
450	451	452	453	454
455	456	457	458	459
460	461	462	463	464
465	466	467	468	469
470	471	472	473	474
475	476	477	478	479
480	481	482	483	484
485	486	487	488	489
490	491	492	493	494
495	496	497	498	499
500	501	502	503	504
505	506	507	508	509
510	511	512	513	514
515	516	517	518	519
520	521	522	523	524
525	526	527	528	529
530	531	532	533	534
535	536	537	538	539
540	541	542	543	544
545	546	547	548	549
550	551	552	553	554
555	556	557	558	559
560	561	562	563	564
565	566	567	568	569
570	571	572	573	574
575	576	577	578	579
580	581	582	583	584
585	586	587	588	589
590	591	592	593	594
595	596	597	598	599
600	601	602	603	604
605	606	607	608	609
610	611	612	613	614
615	616	617	618	619
620	621	622	623	624
625	626	627	628	629
630	631	632	633	634
635	636	637	638	639
640	641	642	643	644
645	646	647	648	649
650	651	652	653	654
655	656	657	658	659
660	661	662	663	664
665	666	667	668	669
670	671	672	673	674
675	676	677	678	679
680	681	682	683	684
685	686	687	688	689
690	691	692	693	694
695	696	697	698	699
700	701	702	703	704
705	706	707	708	709
710	711	712	713	714
715	716	717	718	719
720	721	722	723	724
725	726	727	728	729
730	731	732	733	734
735	736	737	738	739
740	741	742	743	744
745	746	747	748	749
750	751	752	753	754
755	756	757	758	759
760	761	762	763	764
765	766	767	768	769
770	771	772	773	774
775	776	777	778	779
780	781	782	783	784
785	786	787	788	789
790	791	792	793	794
795	796	797	798	799
800	801	802	803	804
805	806	807	808	809
810	811	812	813	814
815	816	817	818	819
820	821	822	823	824
825	826	827	828	829
830	831	832	833	834
835	836	837	838	839
840	841	842	843	844
845	846	847	848	849
850	851	852	853	854
855	856	857	858	859
860	861	862	863	864
865	866	867	868	869
870	871	872	873	874
875	876	877	878	879
880	881	882	883	884
885	886	887	888	889
890	891	892	893	894
895	896	897	898	899
900	901	902	903	904
905	906	907	908	909
910	911	912	913	914
915	916	917	918	919
920	921	922	923	924
925	926	927	928	929
930	931	932	933	934
935	936	937	938	939
940	941	942	943	944
945	946	947	948	949
950	951	952	953	954
955	956	957	958	959
960	961	962	963	964
965	966	967	968	969
970	971	972	973	974
975	976	977	978	979
980	981	982	983	984
985	986	987	988	989
990	991	992	993	994
995	996	997	998	999
1000	1001	1002	1003	1004

PLOTTED VALUES - 197 EXCLUDED VALUES - 0 MISSING VALUES - 0

ORIGINAL 01

B. 4

09/04/86

PAGE 9

36

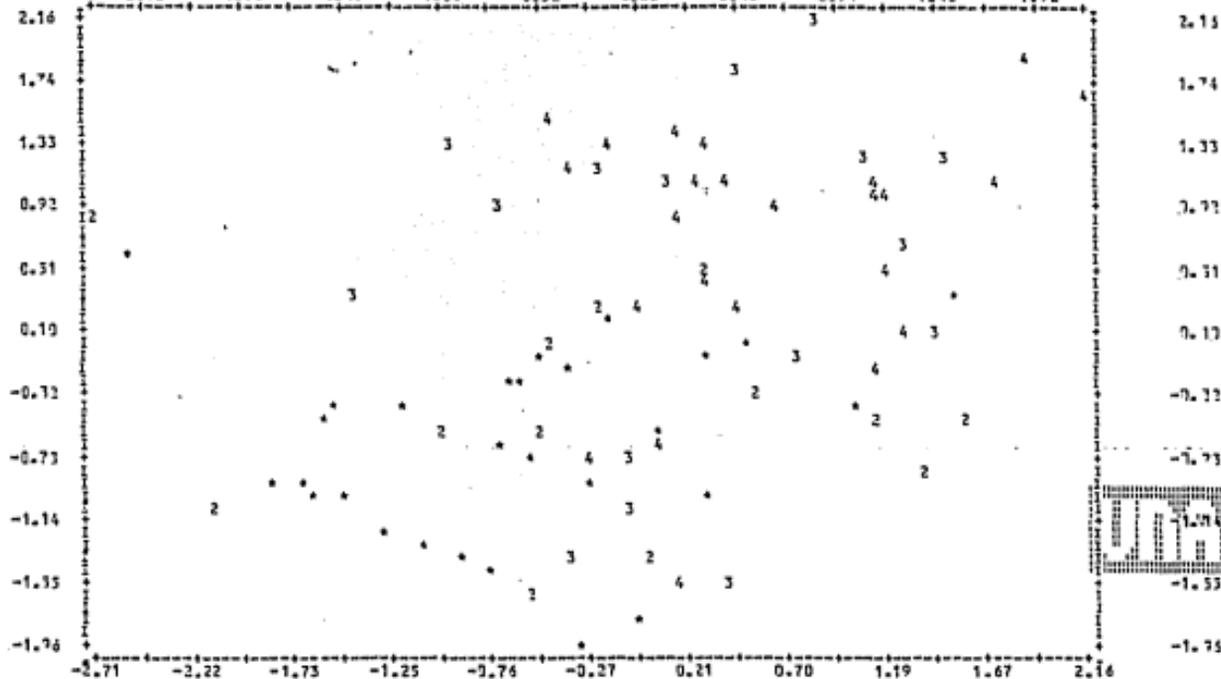
FILE NOYANG  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(COLUMN) A2 REGISTADO  
-2.46 -1.98 -1.49

(ACROSS) A4  
-0.03 0.46

REGAUNTON  
0.94

1.43 1.92



PLOTED VALUES - 197

EXCLUDED VALUES -

0

MISSING VALUES - 0

ORDINAL 01

B.15

09/04/86

PAGE 10

51

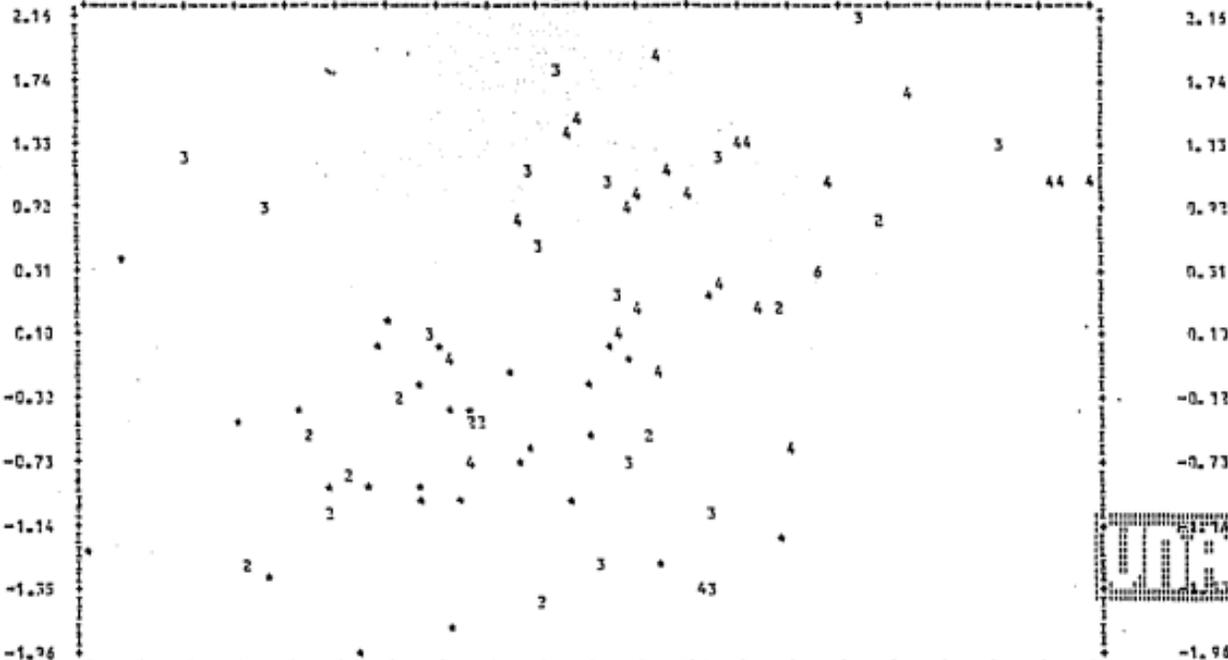
FILE SQUARE  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(COUNT) A2 REGETADO

(ACROSS) AS

MOTOR

-2.15 -1.66 -1.15 -0.67 -0.18 0.33 0.81 1.30 1.80 2.29



PLOTTED VALUES -

197

EXCLUDED VALUES-

0

MISSING VALUES -

0

-1.96

ORDINAL 1015

B.6

09/04/86

PAGE 6

21

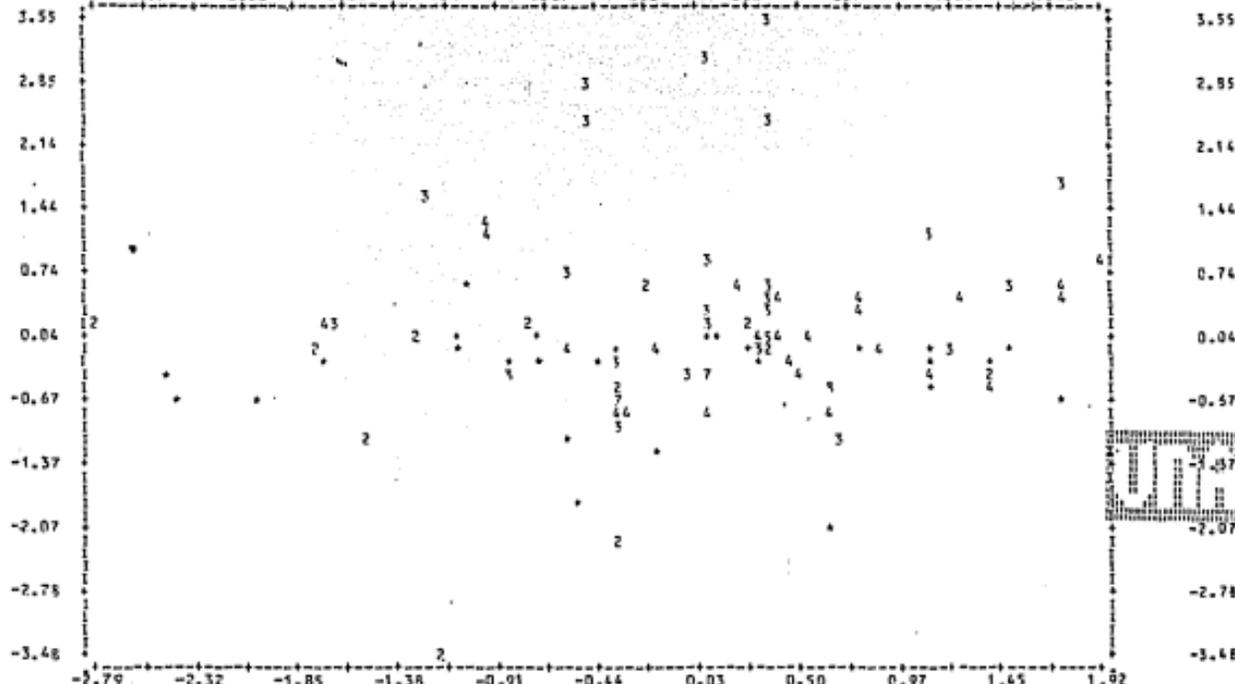
FILE NONAME  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DPHH) A1 ORIENTATION  
-2.55 -2.08 -1.61

(ACROSS) A6  
-0.20 0.27

REGAUTON  
0.74

1.21 1.68



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

PLOTTED VALUES - 250 EXCLUDED VALUES - 0 MISSING VALUES - 3

ORDINAL 1015

B.7

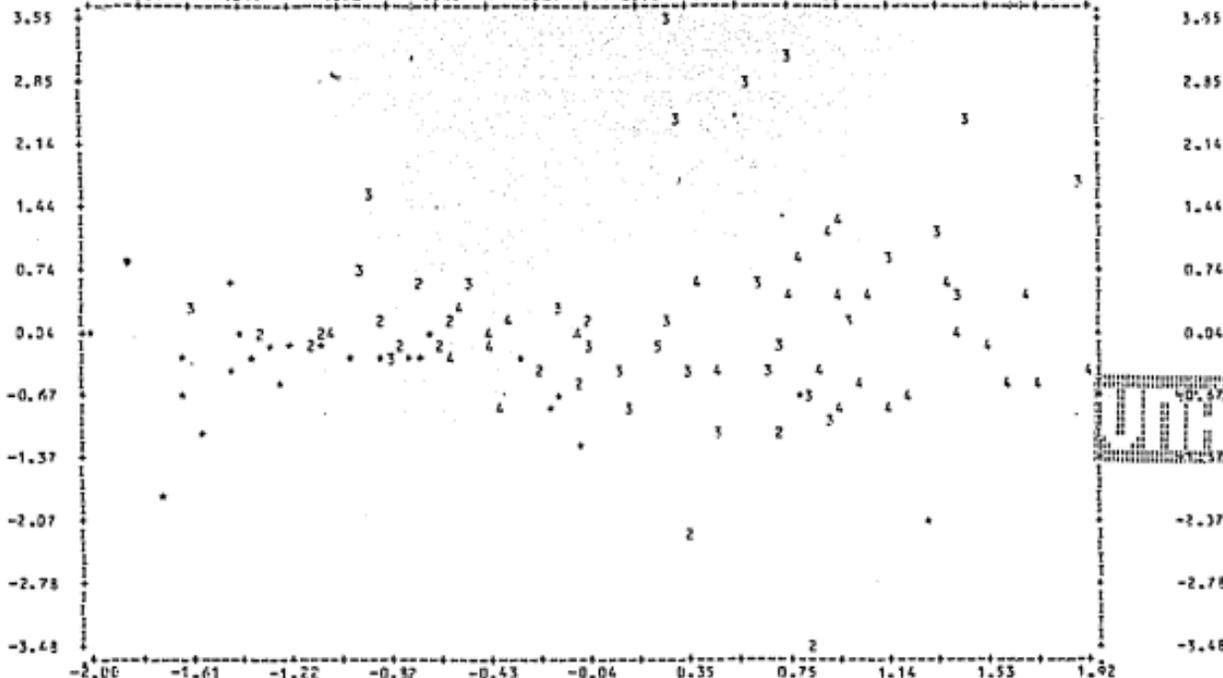
00/04/56

PAGE 4

FILE NOVAH  
SCATTERGRAM OF (CREATION DATE = 00/04/56)  
(DOWN) A1 ORIENTATION

(ACROSS) A2 REGESTADO

-1.81 -1.41 -1.02 -0.63 -0.24 0.16 0.55 0.94 1.34 1.73



PLOTTED VALUES -

2:0

EXCLUDED VALUES-

0

MISSING VALUES -

0

00/04/56

ORDINAL 1015

B. B

09/04/86

PAGE 11

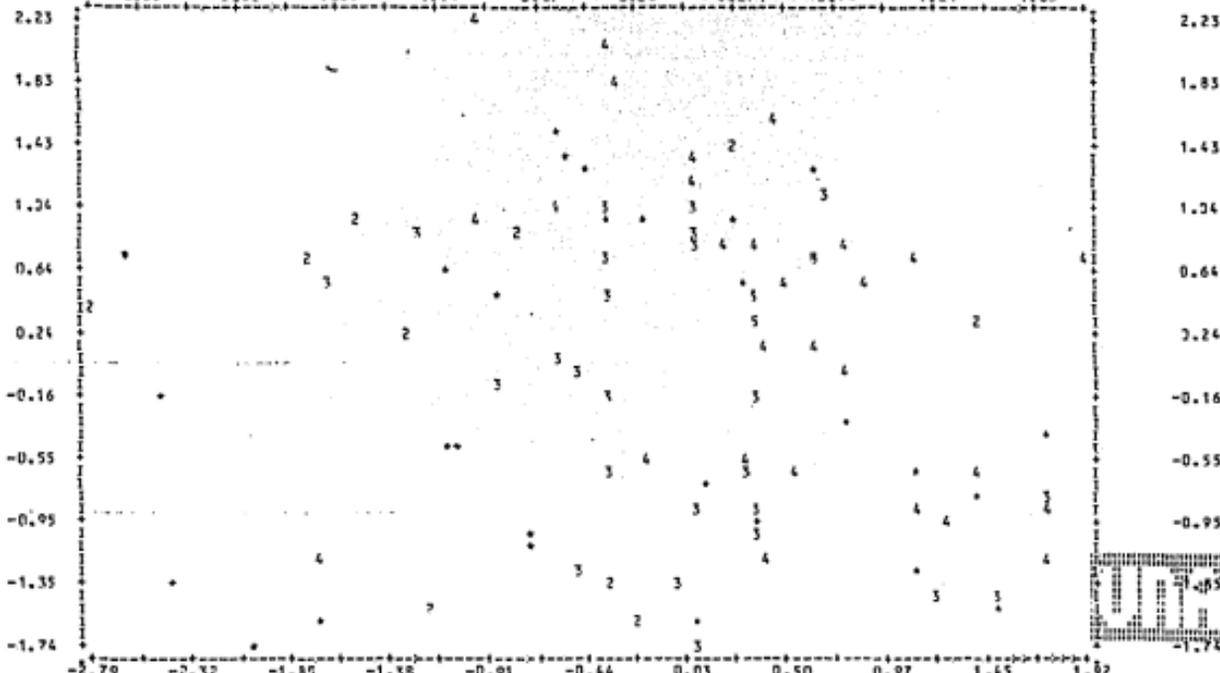
FILE NAME  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DOWN) A3 RANGE

(ACROSS) A4

REGAUTN

-2.55 -2.02 -1.61 -1.14 -0.67 -0.20 0.27 0.74 1.21 1.69



PLOTTED VALUES -

250

EXCLUDED VALUES -

0

MISSING VALUES -

0











B.14

09/04/86

PAGE 7

ORIGNAL 30

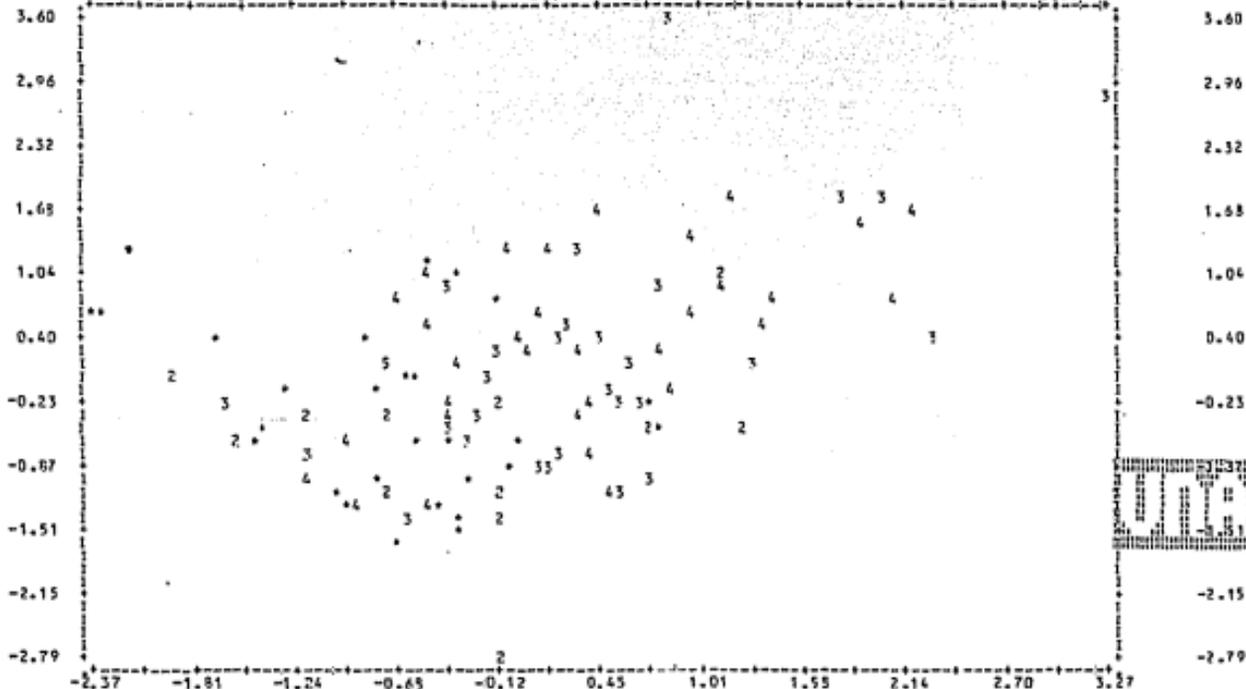
FILE NAME SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DOWN) A1 ORIENTATION  
-2.09 -1.52 -0.98

(ACROSS) A5

MOTOR

1.86 2.42 2.99



PLOTTED VALUES -

271

EXCLUDED VALUES -

0

MISSING VALUES -

0

ORIGNAL 30

B.15

09/04/86

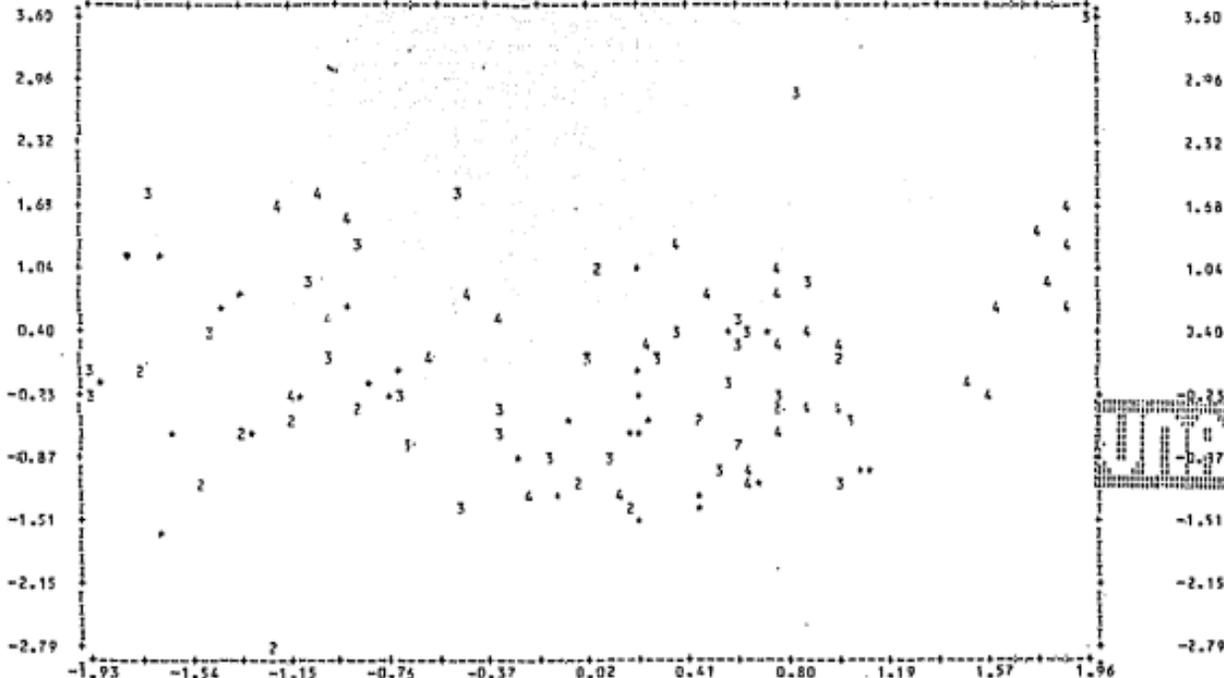
PAGE 5

FILE HONAME  
SCATTERGRAM 37

(CREATION DATE = 09/04/86)  
(DOWN) A1 ORIENTATION  
-1.73 -1.34 -0.96

(ACROSS) A5  
0.23 0.60

RANGE 0.90 1.38 1.77



0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

PLOTTED VALUES - 271

EXCLUDED VALUES - 0

MISSING VALUES - 3

ORDINAL 30

B.16

09/01/86

PAGE 4

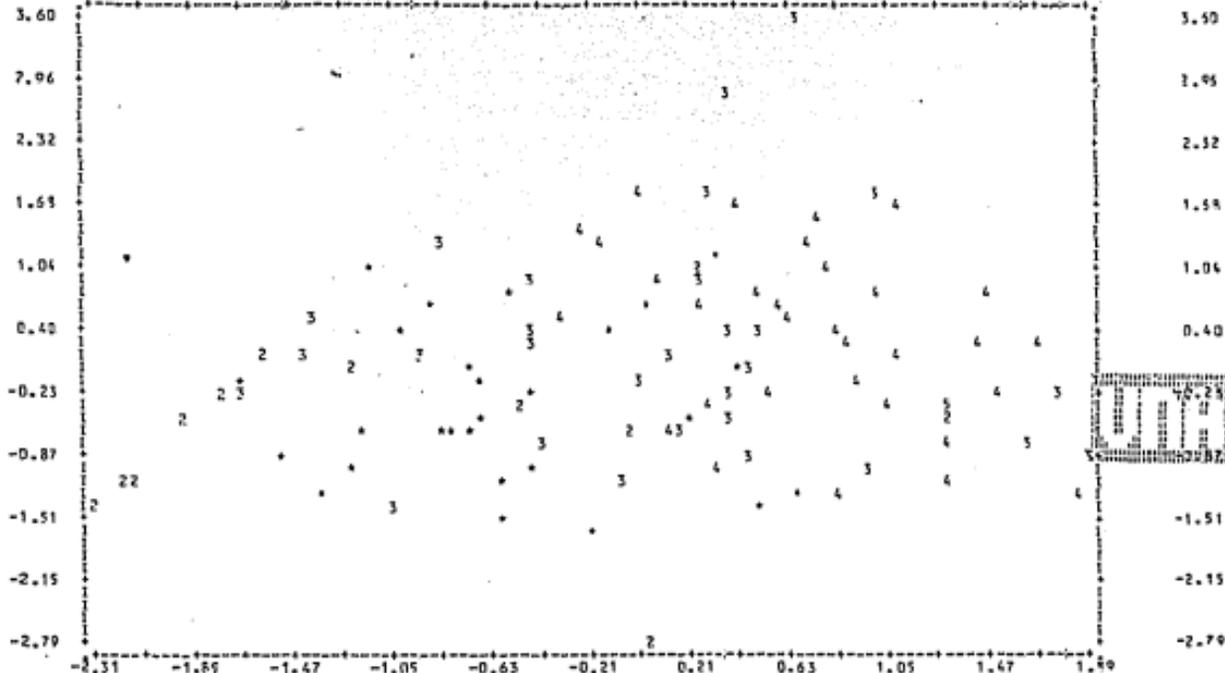
FILE NO NAME  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DOWN) A1 ORIENTATION  
-2.10 -1.68 -1.26

(ACROSS) A2  
0.00 0.42

REGESTADO  
0.84

1.26 1.68



PLOTTED VALUES -

271

EXCLUDED VALUES -

0

MISSING VALUES -

0

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44	45
46	47	48	49	50
51	52	53	54	55
56	57	58	59	60
61	62	63	64	65
66	67	68	69	70
71	72	73	74	75
76	77	78	79	80
81	82	83	84	85
86	87	88	89	90
91	92	93	94	95
96	97	98	99	100

ORDINAL 30

B.17

09/04/86

PAGE 13

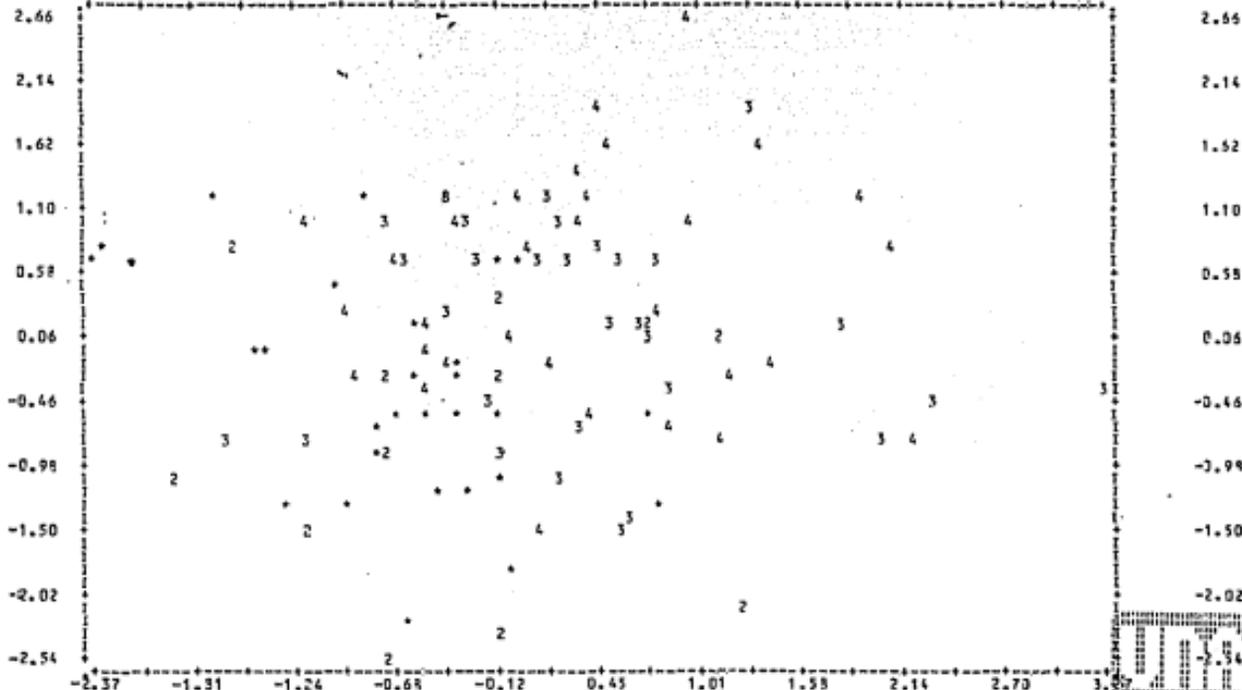
FILE NAME  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DOWN) A4 REGAULTON  
-2.09 -1.52 -0.96

(ACROSS) A5  
0.73 1.29

MOTOR  
1.86

2.42 2.09



PLOTTED VALUES -

271

EXCLUDED VALUES -

0

MISSING VALUES -

3



ORDINAL 30

B.18

09/04/86

PAGE 11

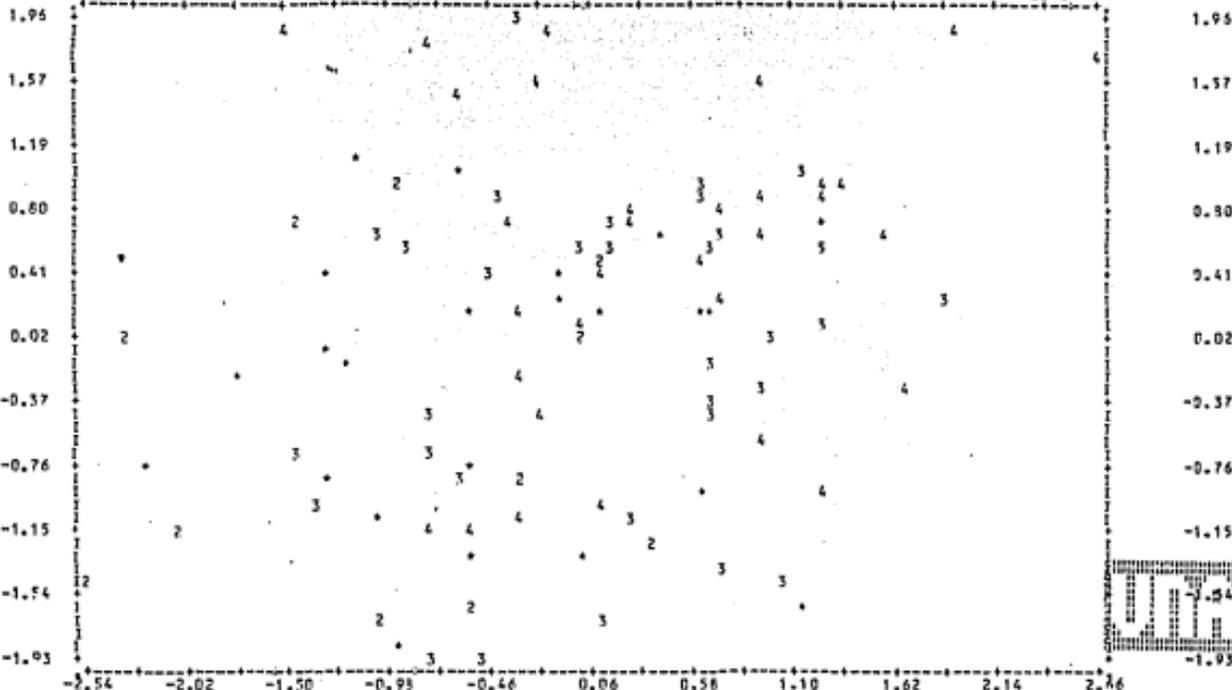
6

FILE NAME  
SCATTERGRAM OF

(CREATION DATE = 09/04/86)  
(DOWN) A3 RAHSO

(ACROSS) A4  
REGAUTON

-2.28 -1.76 -1.24 -0.72 -0.20 0.32 0.84 1.36 1.88 2.40



PLOTTED VALUES -

271

EXCLUDED VALUES -

0

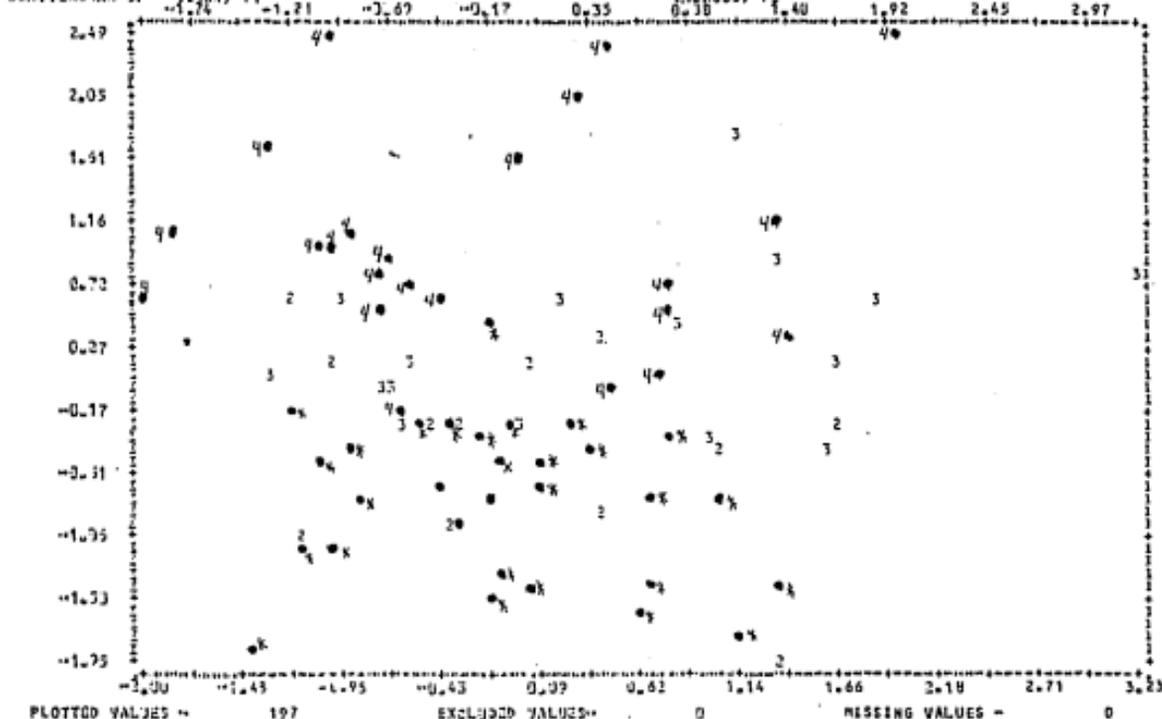
MISSING VALUES -

0

FILE SCATTER PLOT

(CREATED BY DATE = 07/21/85)

(VERSION) F2



1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8
5	6	7	8	9
6	7	8	9	10
7	8	9	10	11
8	9	10	11	12
9	10	11	12	13
10	11	12	13	14
11	12	13	14	15
12	13	14	15	16
13	14	15	16	17
14	15	16	17	18
15	16	17	18	19
16	17	18	19	20
17	18	19	20	21
18	19	20	21	22
19	20	21	22	23
20	21	22	23	24
21	22	23	24	25
22	23	24	25	26
23	24	25	26	27
24	25	26	27	28
25	26	27	28	29
26	27	28	29	30
27	28	29	30	31
28	29	30	31	32
29	30	31	32	33
30	31	32	33	34
31	32	33	34	35
32	33	34	35	36
33	34	35	36	37
34	35	36	37	38
35	36	37	38	39
36	37	38	39	40
37	38	39	40	41
38	39	40	41	42
39	40	41	42	43
40	41	42	43	44
41	42	43	44	45
42	43	44	45	46
43	44	45	46	47
44	45	46	47	48
45	46	47	48	49
46	47	48	49	50
47	48	49	50	51
48	49	50	51	52
49	50	51	52	53
50	51	52	53	54
51	52	53	54	55
52	53	54	55	56
53	54	55	56	57
54	55	56	57	58
55	56	57	58	59
56	57	58	59	60
57	58	59	60	61
58	59	60	61	62
59	60	61	62	63
60	61	62	63	64
61	62	63	64	65
62	63	64	65	66
63	64	65	66	67
64	65	66	67	68
65	66	67	68	69
66	67	68	69	70
67	68	69	70	71
68	69	70	71	72
69	70	71	72	73
70	71	72	73	74
71	72	73	74	75
72	73	74	75	76
73	74	75	76	77
74	75	76	77	78
75	76	77	78	79
76	77	78	79	80
77	78	79	80	81
78	79	80	81	82
79	80	81	82	83
80	81	82	83	84
81	82	83	84	85
82	83	84	85	86
83	84	85	86	87
84	85	86	87	88
85	86	87	88	89
86	87	88	89	90
87	88	89	90	91
88	89	90	91	92
89	90	91	92	93
90	91	92	93	94
91	92	93	94	95
92	93	94	95	96
93	94	95	96	97
94	95	96	97	98
95	96	97	98	99
96	97	98	99	100

## REFERENCIAS

- Atkin L. Programa de seguimiento psicológico en el INPer-DIF. Servicio de psicología y desarrollo infantil INPer, Comunicación interna, 1980.
- Atkin L. C., Givaudan M. M., Landeros V. G., Montiel G. G., Olivera O. M. C. y Valdes B. A. M. Perfil psicológico del niño de alto riesgo en seguimiento. Boletín médico del Hospital Infantil de México, 1986, 43, 90-98.
- Atkin L. C. y cols. Estudio inicial del uso de la escala de conductas neonatales (Brazelton) en un grupo de neonatos sanos nacidos en el INPer-DIF: análisis descriptivo. Servicio de psicología y desarrollo infantil INPer, Comunicación interna, 1982.
- Atkin L. C., Givaudan M. M. y Olivera O. M. C. Estudio comparativo de la conducta de neonatos hiperbilirubinémicos y sanos a través del primer mes de vida. Departamento de crecimiento y desarrollo INPer, Comunicación interna, 1984.
- Agresti A. Analysis of ordinal categorical data. Nueva York: Wiley, 1984.
- Barlow R. E., Bartholomew D. J., Bremner J. M. y Brunk H. D. Statistical inference under order restrictions. Londres: Wiley, 1972.
- Brazelton T. B. Neonatal Behavioral Assessment Scales. Clinics in developmental medicine no. 50, 1973, London: William medical books.
- de Leeuw J., Young F. W. y Takane Y. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika, 1976, 41, 471-503.
- Gower J. C. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 1966,

53, 315-328.

- Gower J. C. A general coefficient of similarity and some of its properties. *Biometrics*, 1971, 27, 857-871.
- Iraizos Lopez M. Implantación de programas software para realizar técnicas de escalamiento multidimensional. Tesis de licenciatura, Facultad de Ciencias UNAM, 1984.
- Kruskal J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, 29, 1-27. (a)
- Kruskal J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, 29, 28-42. (b)
- Shepard R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, 27, 125-140. (a)
- Shepard R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 1962, 27, 219-246. (b)
- Shepard R. N. Representation of structure in similarity data: problems and prospects. *Psychometrika*, 1974, 39, 373-421.
- Shepard R. N., Romney K. A. y Nerlove S.B. *Multidimensional scaling*. Nueva York: Seminar Press, 1972.
- Sneath P. H. y Sokal R. R. *Numerical taxonomy*. San Francisco: Freeman, 1973.
- Torgerson W. S. *Theory and methods of scaling*. Nueva York: Wiley, 1958.
- Young F. W., de Leeuw J. y Takane Y. Regression with qualitative and quantitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, 1974, 41, 505-529.