

EL ANALISIS DE CÚMULOS APLICADO A DATOS LONGITUDINALES DE CRECIMIENTO

GUILLERMO ESPINOSA*
JOHANNA FAULHABER

Introducción

Las diferencias individuales en el crecimiento humano normal tienen su origen, por un lado, en factores genéticos y, por el otro, en las influencias ejercidas sobre éstos por las condiciones ambientales, tanto internas o funcionales del organismo, como externas, y entre estas últimas están en primer término las nutricionales e higiénicas bajo las cuales crece el individuo. De este modo, en condiciones ambientales temporal o constantemente adversas, tanto el niño genéticamente "grande" como el "pequeño" no alcanzan el tamaño máximo al cual están predestinados por la herencia, sino que se quedan en un plano inferior a éste, mientras que bajo condiciones óptimas pasa lo contrario y tanto el primero como el segundo llegan al límite superior de su propio campo de variación heredado. Sin embargo, aún bajo las mejores condiciones ambientales, el niño genéticamente "pequeño" no llegará a alcanzar, como adulto, las dimensiones del cuerpo de los genéticamente "grandes" de la población a que pertenece.

Hemos llamado "tendencias" a los tipos de formas particulares de las curvas individuales del crecimiento, las cuales son el resultado de la interacción, a distintas edades cronológicas, de los diversos factores antes mencionados, aunque se conserva básicamente la curva sigmoidea, común al crecimiento de todos los niños.

Hipótesis

Existen diferentes "tendencias" en el crecimiento humano, observables en los datos antropométricos longitudinales obtenidos entre las edades de un mes y 13 años.

* IIMAS, UNAM.

Objetivos específicos

1. Construir una clasificación de los individuos que describa y agrupe las distintas "tendencias".
2. Estimar la utilidad de los métodos jerárquicos de clasificación para el análisis de datos longitudinales del crecimiento humano.

Selección de la muestra

Para formarnos una idea de esta utilidad, y de la manera en que habríamos de proceder en el análisis, se decidió considerar inicialmente sólo a los varones observados en el estudio longitudinal llevado a cabo por Faulhaber (1976) y se seleccionó una muestra de 40 niños para trabajarla, en esta primera etapa. Se escogió este número arbitrariamente, pero considerando que el trabajo manual y el tiempo de computadora necesarios para manejar esta información, deberían ser razonablemente reducidos. El sexo también se escogió arbitrariamente. Los individuos tomados en cuenta fueron los que asistieron con más frecuencia a las sesiones de medición.

Para determinar hasta qué punto el aumento en el número de individuos considerados alteraba los resultados obtenidos originalmente a base de los primeros 40 niños, se vio a partir de la lista de asistencia de los niños a las observaciones, que de los 250 varones 79 tenían un número aceptable de asistencias en el transcurso del tiempo.

Medidas

Las mediciones se realizaron mensualmente entre uno y 15 meses de edad, después a los 18 y 21 meses y a partir de los dos años semestralmente hasta la edad de 13 años, es decir, se cuenta con 40 grupos de edad.

Las medidas lineales se tomaron en milímetros y el peso se determinó con una aproximación de 10 gr. hasta los 15 kg. y de 100 gr. posteriormente.

De todas las medidas de que se disponía se seleccionaron las más representativas del crecimiento total del cuerpo, dejando fuera de consideración a otros diámetros, a las circun-

ferencias y a las medidas cefálicas, debido a que éstas últimas siguen un ritmo de crecimiento distinto al que caracteriza al resto del cuerpo.¹

Para este primer ensayo, por lo tanto, sólo se tomaron en cuenta la talla total y la sentada, el peso, el diámetro biacromial y el bicrestiliaco y como único índice el acromio-cristal.

Variables

La forma en que fueron obtenidos los datos, determinó dos conjuntos de variables: el de las medidas mensuales-trimestrales y el de las semestrales.

Análisis de cúmulos (Clasificación numérica)

Coefficiente de asociación. Inicialmente se consideraron dos coeficientes de asociación de manera tentativa, uno en el cual se expresa la divergencia en términos de la desviación estándar y el otro, en términos del rango. Sin embargo, al comparar los resultados obtenidos con ambos para la talla, se observó una gran semejanza en los resultados y se decidió trabajar con el coeficiente de Gower, que implicaba menos trabajo.

Uno de estos coeficientes fue:

$$d_{ik} = \frac{\sum_{j=1}^{27} |t_{ij} - t_{kj}|}{3 s_j w_{ik}}$$

donde d_{ik} denota la disimilaridad entre los individuos i y k ; t_{ij} la medida observada en el individuo i a la edad j ; s_j la desviación estándar de las medidas tomadas en todos los individuos de la serie a la edad j ; y w_{ik} el número de mediciones con que se cuenta tanto para el individuo i como el k .

El otro coeficiente fue:

¹ Fauhaber (1978).

$$S_{ik} = 1 - \frac{\sum_{j=1}^{27} |t_{ij} - t_{kj}|}{W_{ik} R_j}$$

donde S_{ik} denota la similaridad entre los individuos i y k ; R_j la diferencia entre los valores máximo y mínimo de las medidas observadas en todos los individuos de la serie a la edad j ; y los demás símbolos denotan lo mismo que el caso anterior.

Ambos coeficientes indicarán una gran semejanza entre individuos cuando los datos correspondientes de éstos estén muy cercanos entre sí a las diferentes edades.

Puesto que los resultados obtenidos por ambos coeficientes eran casi idénticos, se trabajó finalmente con el último citado, conocido como el coeficiente de Gower² para estados múltiples.

Métodos de clasificación. Se emplearon, simultáneamente, el de conexión simple (single link), el de conexión completa (complete link), el de promedios intergrupales (average linkage between merged groups o mean link) y el de promedios intragrupalos (average linkage within the group o mean within off diagonal).³

Evaluación de los dendrogramas. Puesto que al calcular un coeficiente de asociación en un problema dado se deforma la información original, y al calcular un dendrograma se deforma la información contenida en el coeficiente de asociación,⁴ es imprescindible evaluar, aunque sea subjetivamente, los dendrogramas obtenidos.

Tratándose de técnicas heurísticas de análisis de datos, esta evaluación subjetiva debe hacerla esencialmente el usuario de ellas.

Una manera de evaluar un dendrograma a partir del coeficiente de asociación es escribir la tabla de disimilaridad (similaridad) cambiando el orden original de los objetos por

² Sneath y Sokal (1973).

³ Anderberg (1973).

⁴ Anderberg (1973). Some Remarks on Utilization, pp. 21-24.

el que produzca el dendrograma. Los grupos deben aparecer como subtablas, cuadradas, de valores bajos (altos), que no tengan disimilitudes bajas (similitudes altas) con los otros grupos, en términos generales. La evaluación es más sencilla, si se construye una tabla, achurando de distinta manera los valores de algunos intervalos que se escojan.

Una manera de evaluar el dendrograma a partir de la información original es analizar el comportamiento de las variables en los distintos grupos con la idea de observar, si estos son realmente grupos diferentes. Con este propósito se igualan los rangos de las variables y se hacen gráficas de los "perfiles"⁵ de los grupos a manera de histogramas para evitar el traslape de las gráficas individuales.

En la Figura 1 se resume el proceso de análisis de cúmulos aplicado en este caso.

Resultados

Los resultados obtenidos para la talla ilustran bien los que se obtuvieron para las demás medidas, por lo que sólo se presentan los pasos dados para aquella.

Dendrogramas

Se analizaron los dendrogramas obtenidos con los métodos ya citados, a partir del coeficiente de Gower aplicado a las tallas observadas a las edades de un mes, 6 meses y después semestralmente hasta los 13 años. Se vio que entre ellos varía el número de grupos formados, así como la colocación de un mismo individuo dentro de estos últimos. Se decidió trabajar con el método de conexión completa por presentar con mayor claridad el agrupamiento de los individuos en sus diferentes ramas.

El dendrograma correspondiente a la talla total se representa en la Fig. 2 y la correspondiente tabla de asociación en la Fig. 3. Se observa en ellas que se obtuvieron 6 grupos. Al comparar entre si las gráficas individuales de los niños asignados a cada uno de éstos, se pudo observar que cada grupo comprende a aquellos que se asemejan por su situación en los campos establecidos por la media menos y más una y

⁵ Hartigan (1975), Cap. 1, Profiles, pp. 32-34.

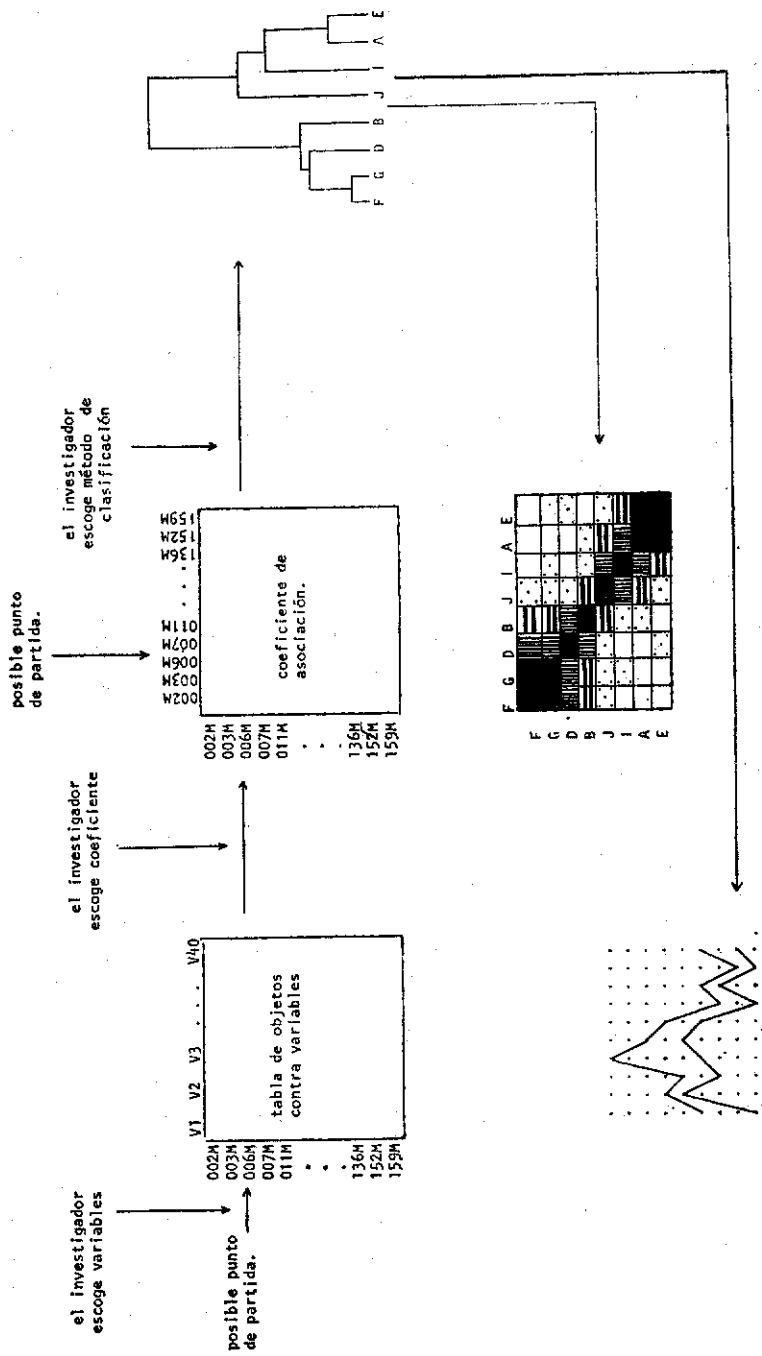


Fig. 1 EL proceso de análisis de cúmulos

dos desviaciones estándar del total de niños observados para cada edad. Por consiguiente se enumeraron los grupos de 1 a 6, comprendiendo el primero los individuos más altos, hasta llegar en orden decreciente al 6 que contiene los niños de estatura menor.

Se pudo apreciar además, que la asignación individual a dichos grupos se basó más bien en el promedio de las diferencias absolutas entre las medidas del niño y las de los miembros del grupo. Más adelante se precisa este punto.

Histogramas

Los histogramas de los 6 grupos, representados en la Fig. 4 señalan el rango, dividido para cada edad en 10 unidades, que ocupan los niños comprendidos en cada uno de los 6 grupos.

Se observa que los grupos extremos están "bien formados" en el sentido de que hay uno de individuos altos, otro de individuos bajos.

Además se constata que las fronteras de los grupos se traslapan sobre todo en los grupos centrales, lo que hace suponer que hay casos individuales que bien pueden pertenecer a cualquiera de los dos grupos. Esto también sugiere que la tabla de asociación podría llevarse a la forma de Robinson y por lo tanto sería más indicado buscar una seriación que una clasificación.⁶

Los histogramas muestran diferencias generales entre los grupos y en la Fig. 5 se representa el punto medio del rango de cada uno ellos a las diversas edades. Por las diferentes formas que presentan las curvas, sobre todo las de los grupos centrales, cabe preguntarse si esto indica "tendencias" distintas. Para responder a esta pregunta se hicieron las siguientes consideraciones:

1. Las diferencias mayores se dan tanto en las edades bajas como en las altas, que son en las que se tienen más faltas de asistencia a las observaciones, de manera que podría ser que pocos individuos, determinaron el aspecto general del histograma del grupo.

⁶ López y Espinosa (1976).

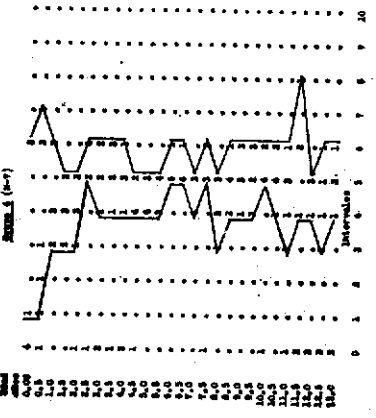
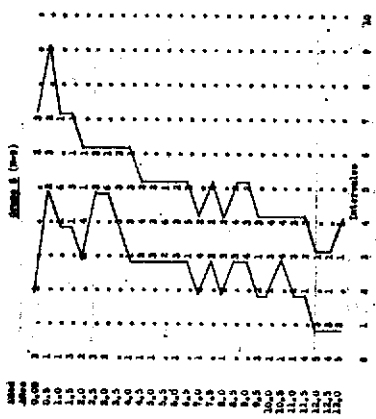
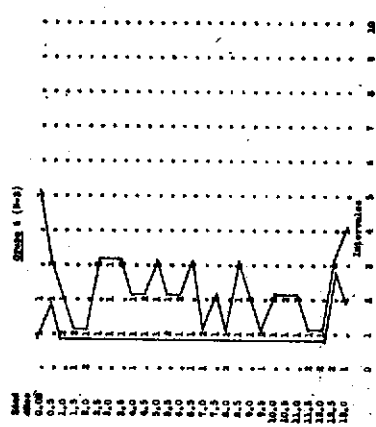
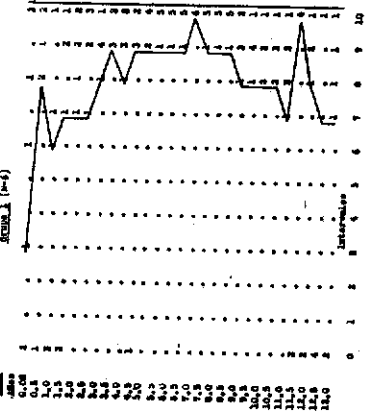
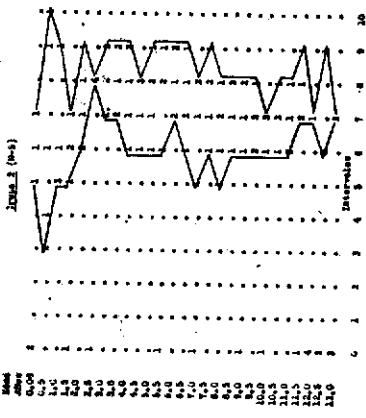
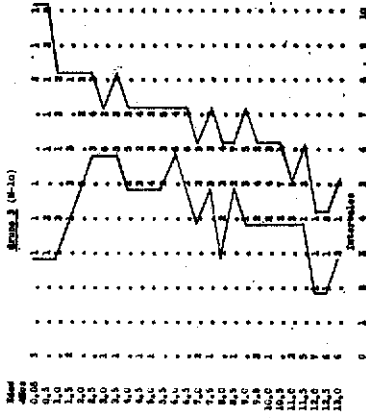


Fig. 4 Histogrammes de los 6 Grupos

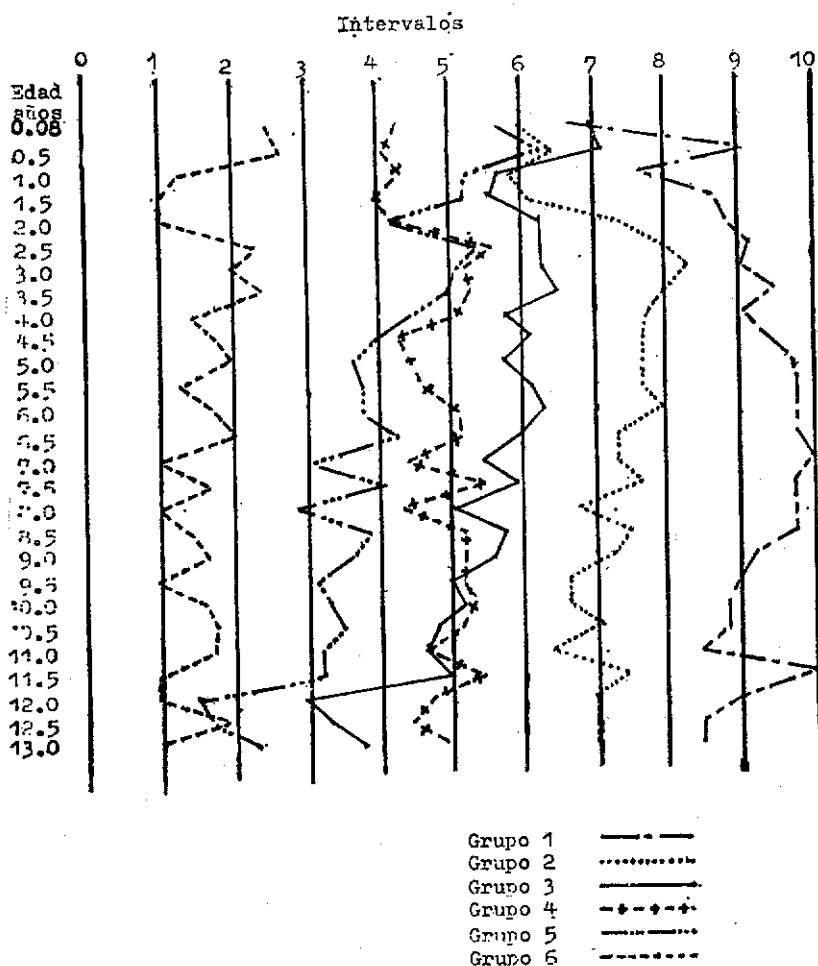


FIG. 5. Curva central de los histogramas de los 6 grupos establecidos para la talla.

2. Los histogramas no reflejan las curvas individuales y podría ser que la tendencia global no describiera la forma de las gráficas individuales.

Para aclarar el punto 1, se repitió el análisis considerando como variables a las tallas desde 1 año hasta los 11 años de edad. Los grupos fueron los mismos, esencialmente, y los histogramas mostraron tendencias semejantes, aunque menos marcadas, como era de esperarse.

Para aclarar el punto 2, se expresaron, en términos de la desviación estándar correspondiente, las divergencias que cada uno de los 40 niños considerados en este ensayo presentaba a las diversas edades con respecto a la media aritmética de la serie total de los individuos estudiados. La Fig. 6 representa estas curvas para el grupo 1, es decir, para uno de los grupos más claramente definidos como tal, siendo representativo de lo que sucede en los demás grupos, sólo que en ellos el cuadro se complica por el entrecruzamiento de las curvas individuales de niños clasificados en grupos distintos. Se confirma, por lo tanto, que las curvas individuales no determinan la forma general de los grupos.

Aumento en el número de individuos en el cálculo

Al aumentar el número de individuos considerados en el cálculo de 40 a 79 varones, se pudo constatar que los 6 grupos establecidos inicialmente para el menor, se redujo a 5. Consecuentemente, aunque los grupos extremos de niños más altos y más bajos quedaron idénticos, hubo un agrupamiento distinto de los individuos que constituyen los grupos centrales.

Algunos aspectos colaterales de interés antropológico

Para indagar si el agrupamiento de los individuos según su desarrollo durante los 15 primeros meses de la vida nos podría dar un indicio de la posición que ocupa dentro de los grupos establecidos para las edades posteriores, se realizaron los cálculos correspondientes a la talla total y el peso a base de las observaciones mensuales.

Al comparar el lugar que el mismo individuo ocupa en la clasificación mensual y la semestral se vio que en ambas características algunos niños se mantienen en los grupos correspondientes, pero que la mayoría de ellos cambian indistintamente de grupo con el transcurso del tiempo.

Nos interesó igualmente determinar la coincidencia o la discrepancia en la clasificación de un mismo individuo en cuanto al total de los caracteres métricos tomados en cuenta. En otras palabras, nos planteamos la pregunta hasta qué punto los individuos clasificados según alguna de las características como grandes o pequeños lo sean también en cuanto a las

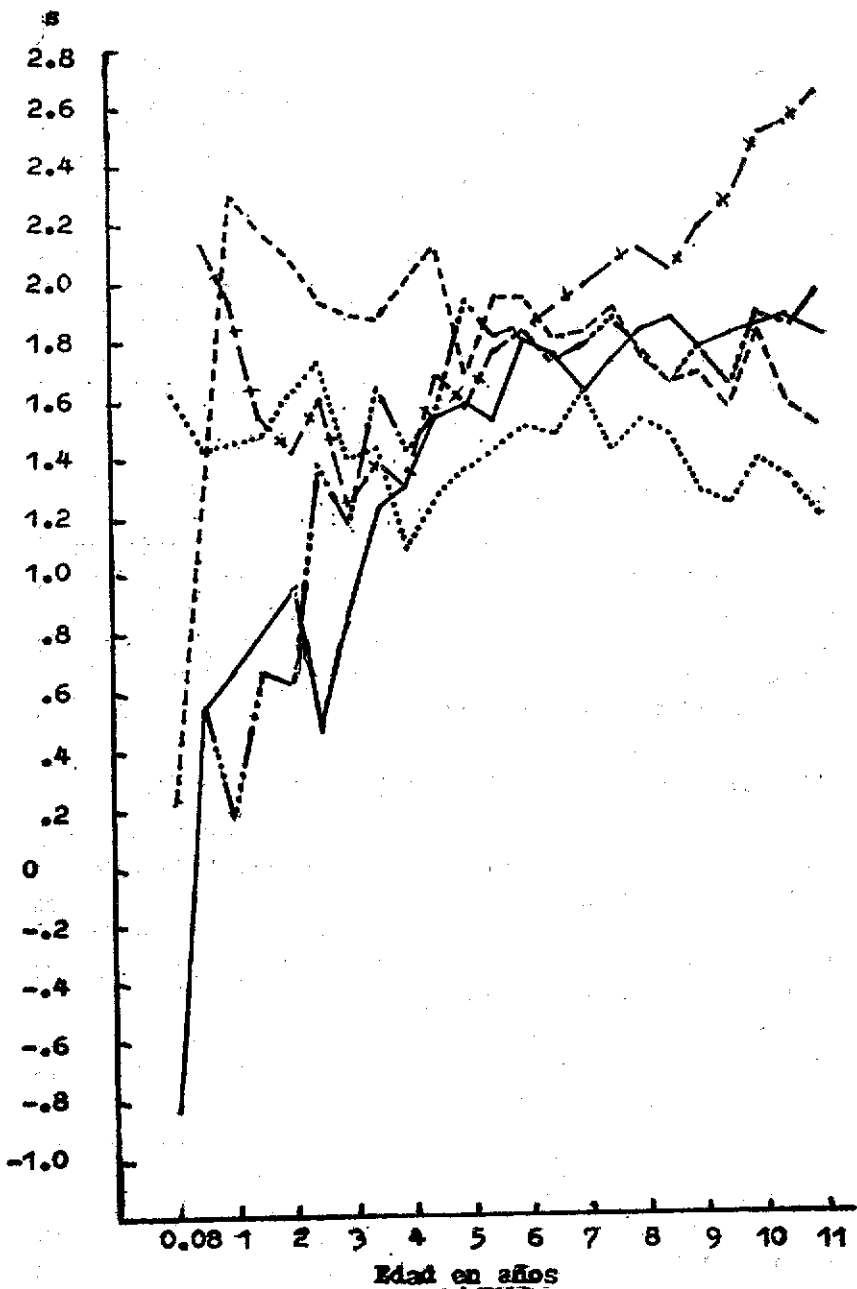


FIG. 6. Desviaciones de algunos individuos del grupo 1 en términos de la s.

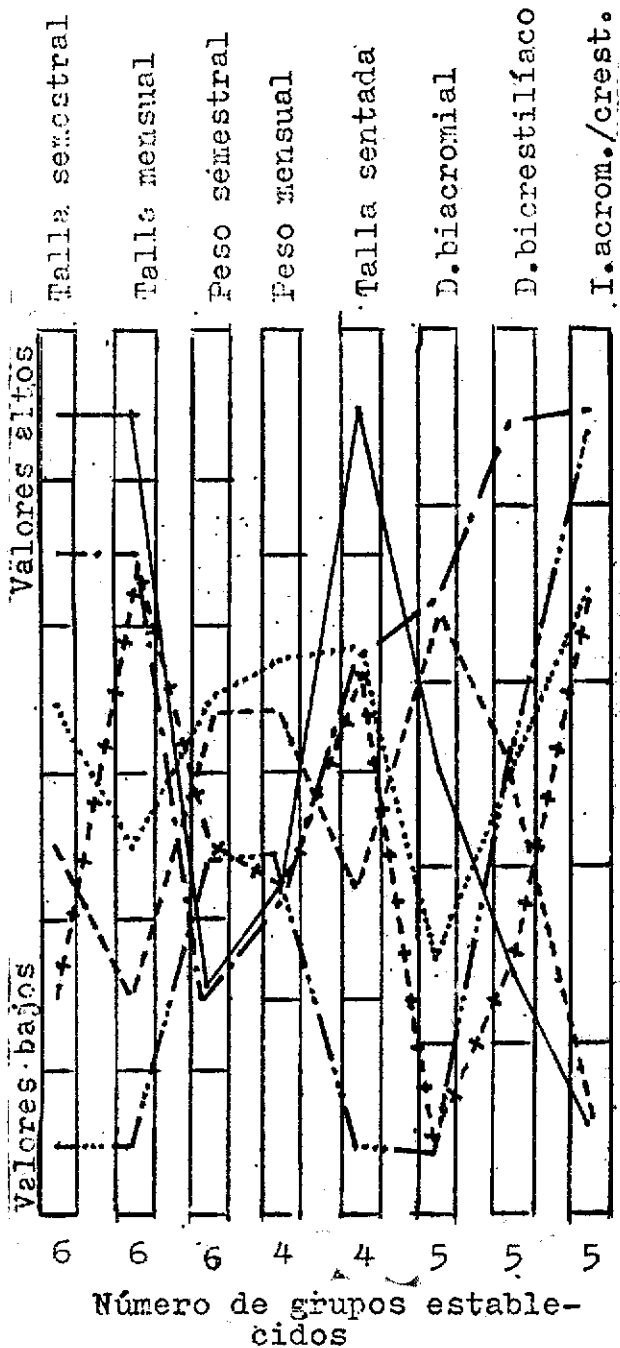


FIG. 7. Grupo que seis individuos de diferente talla ocupan en las diversas medidas.

demás. La Fig. 7 contiene las curvas correspondientes para algunos individuos clasificados según su talla y la posición que ocupan en cuanto a los grupos que se establecieron para los demás caracteres considerados. Como era de esperarse, dichas curvas muestran considerables variaciones, las cuales son, precisamente, la base de los diversos tipos morfológicos individuales, aunque los coeficientes de correlación entre las medidas en cuestión son muchas veces estadísticamente significativas.⁷

Conclusiones

Por lo que se refiere a la utilidad de los métodos jerárquicos de Análisis de Cúmulos para el análisis de datos longitudinales de crecimiento humano, cabe decir que la razón de ser de estas técnicas es, precisamente, determinar si puede construirse una clasificación a partir de los datos a los que se aplique y construirla en caso de que esto sea posible. En este sentido, llevó a determinar que no es posible hacerlo para los datos en cuestión, en términos de los valores absolutos de las medidas observadas, lo cual no significa que sea imposible transformar esas variables de manera que se pueda construir una clasificación que refleje las distintas tendencias.

Además, puede concluirse que al aplicar esas técnicas a datos obtenidos al medir una población animal de una sola especie se debe buscar una seriación de los objetos, en el sentido de Robinson, para detectar si realmente pueden definirse grupos distintos. El cálculo de un dendrograma implica una tabla de asociación en la forma de Robinson,⁸ y hay indicios de que una clasificación puede obtenerse a partir de una o más seriaciones.⁹ Esta seriación también puede obtenerse intuitivamente.¹⁰

BIBLIOGRAFÍA

ANDERBERG, Michael R.
1973 *Cluster Analysis for Applications*. Academic Press,
New York.

⁷ Faulhaber (1979).

⁸ Chávez Cano.

⁹ López, Espinosa y Carvajal (1976).

¹⁰ Espinosa, López y Yankelevich (1977).

CHÁVEZ CANO, M.

fesional: Facultad de Ciencias, UNAM. En preparación.

s.f. *Los conceptos de clasificación de Lerman*. Tesis Pro-

ESPINOSA, G., A. LÓPEZ Y G. YANKELEVICH

1977 Análisis de Cúmulos: Salud Pública. *Gaceta Médica de México*. Vol. 113, No. 11, pp. 525-531.

FAULHABER, J.

1976 *Investigación longitudinal del crecimiento*. Colección Científica No. 26, Antropología Física. Instituto Nacional de Antropología e Historia. México.

1978 Algunos cambios morfológicos durante el crecimiento. *Anales de Antropología*. Vol. XV :323-340. Instituto de Investigaciones Antropológicas, UNAM, México.

1979 El crecimiento diferencial en algunas regiones del cuerpo. *Anales de Antropología*. Vol. XVI en prensa. Instituto de Investigaciones Antropológicas, UNAM. México.

HARTIGAN, John A.

1975 *Clustering Algorithms*. John Wiley and Sons, Nueva York.

LÓPEZ, A. y G. Espinosa

1976 Escalamiento multidimensional, seriación y taxonomía numérica. *Comunicaciones Técnicas del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas*. Serie Azul: Monografías, Vol. 3, No. 20. UNAM, México.

LÓPEZ, A. G. ESPINOSA y R. CARVAJAL

1971 A Relationship Between Seriation and Cluster Analysis. *Comunicaciones Técnicas del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas*. Serie Naranja: Investigaciones. Vol. 7, No. 128. UNAM, México.

SNEATH, P. H. A. y R. R. SOKAL

1973 *Numerical Taxonomy*. W. H. Freeman, San Francisco.