

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL Y POSGRADO DEL C. C. H.
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

APLICACIONES DE METODOS MULTIVARIADOS EN EL CAMPO
DE LA INVESTIGACION NUTRICIONAL

TRABAJO FINAL QUE PRESENTA EL ALUMNO

MAURICIO HERNANDEZ AVILA

PARA OBTENER EL DIPLOMA EN

ESTADISTICA APLICADA

BIBLIOTECA
JUAN A. ESCALANTE N
UNIDAD ACADÉMICA DE
LOS CICLOS PROFESIONAL
Y DE POSGRADO / CCH
UNAM

AGOSTO, 1983.

MEXICO, D. F.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.



Universidad Nacional
Autónoma de México



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 esta 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.

INDICE

	Pág.
INTRODUCCION	1
MATERIAL Y METODOS	5
ANALISIS DE CONGLOMERADOS	7
ANALISIS DE COMPONENTES PRINCIPALES	9
ANALISIS DE DISCRIMINANTES	11
ANALISIS DE RESULTADOS	13
VARIABLES ANTROPOMETRICAS	13
VARIABLES SOCIOECONOMICAS	15
INTERPRETACION DE LAS COMPONENTES	17
CORRELACION ENTRE VARIABLES SOCIOECONOMICAS Y VALORES ANTROPOMETRICOS	21
CONCLUSIONES	26
BIBLIOGRAFIA	27

INTRODUCCION

La desnutrición calórico-proteica ha sido preocupación y motivo de estudio desde hace muchos años. Desde entonces los investigadores han tratado de establecer criterios diagnósticos que indiquen la presencia de esta enfermedad, así como de un sistema de clasificación que permita identificar a los sujetos afectados por ella y los diferentes factores sociales y biológicos con los que se asocia, proponiendo de esta manera alternativas de intervención.

De la gran variedad de indicadores de desnutrición que se ha propuesto, las mediciones antropométricas son las que más se han utilizado. Esto obedece a varias razones: la primera y más importante es su objetividad, adecuada y sencilla estandarización que se puede lograr para recabarlas no se requiere de aparatos sofisticados ni de personal especializado para su obtención, además pueden ser realizadas con facilidad en trabajo de campo y en situaciones de emergencia. Por otro lado, existen normas universales con las que se pueden comparar y expresar de una manera ya convenida en un porcentaje con respecto a estas normas, lo cual simplifica su manejo en el ámbito internacional y permite hacer comparaciones entre varios países.

De la gran variedad de medidas antropométricas que se han propuesto, las más usadas son peso, talla y circunferencia de brazo. De éstas surgió la primera clasificación propuesta por Gómez en 1946, que se basó en el peso para la edad (Cuadro 1). A partir de esta fe-

cha se ha desarrollado una gran cantidad de clasificaciones, aunque la mayoría carece de bases científicas para determinar los puntos de corte que definen los intervalos en que se sitúan las diferentes categorías nutricionales. La experiencia a través del tiempo es que éstos se han determinado de una manera arbitraria.

En un intento para determinar los diferentes puntos de corte, Ramos y L. C. Chen realizaron estudios en poblaciones hospitalaria y rural respectivamente, para evaluar los diferentes puntos de corte con respecto a riesgo de muerte. Ambos comprobaron que el grupo con desnutrición de tercer grado presentaba un riesgo de muerte 10 ó 20 veces mayor que el de los otros grupos. Los grupos normales, con desnutrición leve y moderada, no quedaron definidos ya que entre sí no presentan diferencias en cuanto a riesgo de muerte. El problema para definir los intervalos en que se sitúan las categorías nutricionales se presenta en todos los tipos de clasificación que se han propuesto. El criterio para determinar los puntos de corte es fundamental para la calidad de la inferencia estadística que se puede hacer en el campo de la investigación nutricional, ya que la sensibilidad* y especificidad** para definir cada grupo nutricional varía de acuerdo con la prevalencia de la desnutrición y el valor del punto de corte utilizado para definir las categorías.

En Epidemiología se pueden hacer ajustes para controlar los efectos de los errores en la clasificación. En un programa para detectar la prevalencia de una enfermedad, el principal objetivo es lo-

* Sensibilidad, se define como la probabilidad de clasificar correctamente a un individuo que tiene la enfermedad.

** Especificidad, se define como la probabilidad de clasificar correctamente a un individuo que no tiene la enfermedad

calizar o detectar los verdaderos positivos, y el programa se puede evaluar y balancear de acuerdo al costo-beneficio de la obtención de clasificaciones correctas y errores de clasificación (falsos positivos o negativos). En el campo de la desnutrición las consecuencias de un falso positivo se relacionan generalmente con un aumento en el costo de un programa de alimentación complementaria, mientras que un falso negativo puede tener consecuencias tan graves como la muerte.

Por otra parte, las implicaciones de estos errores de clasificación en la inferencia estadística al estimar diferencias de proporciones o al realizar pruebas de hipótesis no han recibido la atención que merecen, como por ejemplo, el desastroso efecto que pueden tener las variaciones en estos parámetros, que reproduce una tabla tomada del artículo de D. Quade. (Cuadro 2), en la que se calcula la potencia de la prueba: se detecta una diferencia de 0.05 en una población de 500 individuos que es evaluada en dos ocasiones después de una intervención y la confianza se fija en .05. Se aprecia claramente el efecto que las variaciones en la sensibilidad y especificidad del sistema de clasificación puede tener sobre la potencia* de la prueba; la disminución de ésta es muy importante cuando la sensibilidad o especificidad son bajas.

Si a los errores de clasificación se añaden los errores humanos en las mediciones y la dificultad para obtener datos confiables acerca de la edad del sujeto, principalmente en medios rurales, se comprende que la calidad de la inferencia es muy pobre.

* La potencia de una prueba se define como la probabilidad de detectar diferencias cuando éstas existen.

Este trabajo propone la utilización de métodos multivariados que se puedan incorporar al campo de la investigación nutricional de forma rutinaria, con el fin de mejorar la calidad de la inferencia en las investigaciones y lograr una mejor interpretación de las relaciones que existen entre la desnutrición y los diversos factores socioeconómicos. Como un ejemplo de la viabilidad de estos métodos, se incluyen los resultados de una investigación realizada por el INNSZ en el estado de Puebla en 1981.

MATERIAL Y METODOS

Los datos analizados en el presente trabajo provienen de una encuesta que se realizó en una comunidad rural de la Sierra Norte de Puebla. en 1981, como parte de un estudio piloto que se inició en esas fechas con el fin de implementar un programa de atención primaria en esa zona del país.

Una vez elegida la comunidad, se realizó un censo en el que se localizaron a las familias que tenían hijos de 1 a 5 años; a estas familias se les aplicó un cuestionario dietético y otro socioeconómico; uno de los investigadores pesó y midió a todos los niños que se encontraban entre 1 y 5 años de edad.

En este trabajo se evaluaron unicamente las variables socio-económicas las cuales se describen a continuación:

- V₁ Extensión de la parcela
- V₂ Tipo de Cosecha
- V₃ Hectáreas que dedica para sembrar café
- V₉ Total de la cosecha levantada el año pasado
- V₁₂ Miembros de la familia que trabajan
- V₁₃ Rama principal de la actividad del jefe de familia
- V₁₅ Número de dependientes
- V₁₆ Acceso a clínicas de Seguridad Social (IMSS)
- V₁₉ Gasto en transporte
- V₂₀ Gasto en alcohol
- V₂₁ Gasto Médico
- V₂₂ Gasto en producción, incluye salarios, compra de herramientas, abonos, etc.

V ₂₄	Nivel de Educación del Padre
V ₂₆	Nivel de Educación de la Madre
V ₂₇	Estado Fisiológico de la Madre
V ₂₈	Material de las paredes de la vivienda
V ₂₉	Material del techo de la vivienda
V ₃₀	Material del piso de la vivienda
V ₃₁	Fuente del agua de uso doméstico
V ₃₂	Manejo de escretas
V ₃₃	Servicio de luz eléctrica
V ₃₄	Número de cuartos en la vivienda
V ₃₅	Número de personas que duermen en la vivienda
V ₃₆	Número de personas que comen en la vivienda
V ₃₇	Ocupación principal del jefe de la familia
V ₃₈	Lugar que ocupa el preescolar en la familia
IPC	Ingreso anual per cápita
GPC	Gasto per cápita en alimentación

Todas las variables se codificaron de tal manera que los valores altos correspondieran a las características asociadas con mayor riesgo de presentar desnutrición.

Los métodos estadísticos multivariados que se utilizaron fueron los siguientes:

- A) ANALISIS DE CONGLOMERADOS
- B) ANALISIS DE COMPONENTES PRINCIPALES
- C) ANALISIS DE DISCRIMINANTES

A continuación se hace una breve descripción de estas técnicas.

ANÁLISIS DE CONGLOMERADOS

Todas las técnicas de clasificación o Análisis de Conglomerados, tienen como objetivo resolver el problema de formular un esquema de clasificación para agrupar individuos, sobre los cuales se han tomado una serie de mediciones. Se desconocen la cantidad de grupos o categorías de los cuales provienen estos individuos, por lo que se tiene que decidir:

1. Cuántos grupos es necesario considerar
2. Determinar las características de cada grupo
3. Determinar qué individuo se encuentra en cada grupo

Existen varios métodos para clasificar, los cuales se han resumido en tres grupos:

1. Métodos Jerárquicos
2. Métodos de optimización de criterios
3. Métodos basados en ajustar mezclas de densidades normales multivariadas.

En este trabajo se usó un programa de computadora, que genera los grupos de acuerdo al tercer método. La idea principal de este método es que los n individuos u objetos provienen de g poblaciones distintas y los individuos en cada una de las g poblaciones tienen una densidad de probabilidad* donde la forma de α_s es conocida pero θ_s son parámetros desconocidos.

El problema se resuelve estimando los parámetros involucrados θ_s y asignando cada individuo al grupo que tenga mayor probabilidad de pertenencia; en la práctica el método más utilizado es suponer que

* $\alpha_s (X_i | \theta_s)$

todos los grupos tienen una distribución normal multivariada con vector de medias μ_S y matriz de varianzas y covarianzas Σ_S .

Esta técnica se utilizó en el presente estudio para generar los grupos antropométricos y para explorar la existencia de grupos generados a partir de las variables socioeconómicas; en ambos casos se utilizaron nuevas variables generadas por la técnica de Componentes Principales.

ANÁLISIS DE COMPONENTES PRINCIPALES

El problema del Análisis de Componentes Principales se puede resumir de la siguiente manera:

Se ha obtenido un grupo grande de variables que se supone tienen cierta correlación y se desea reducirlas a un nuevo conjunto de variables que no estén correlacionadas entre sí.

Estas nuevas variables reciben el nombre de componentes; son funciones lineales de las variables observadas y están definidas de tal forma que:

La primera componente tiene varianza mayor que todos los índices construidos, es decir el primer componente representa el mayor porcentaje de la variabilidad del conjunto inicial de variables .

La principal idea base de este procedimiento, es que las primeras dos o tres componentes representen gran parte de la información observada en las variables originales; de esta manera se reduce el número de variables, con la ventaja de que estas nuevas variables o componentes pueden sustituir parcial o totalmente a las originales en los subsiguientes análisis que se planteen.

Es importante hacer notar que la técnica de Análisis de Componentes principales, no involucra ninguna suposición acerca de las variables observadas y sólo se requiere que las escalas de medición en las variables respuesta sean equiparables, ya que se trata simplemente de una forma más compacta de presentar la información.

En el presente estudio se realizó esta técnica para reducir la información de dos tipos de variables, las Antropométricas y las Socioeconómicas.

ANÁLISIS DE DISCRIMINANTES

El Análisis de Discriminantes es la extensión del análisis de varianza que se utiliza frecuentemente en la estadística univariante.

El problema de análisis de discriminantes se puede resumir de la siguiente forma:

Se tiene un conjunto de "n" individuos repartidos en "g" grupos, los cuales son mutuamente exclusivos; para cada individuo se tienen registradas "p" características las cuales se representan por X_1, X_2, \dots, X_p .

Se desea investigar la relación de los individuos con el conjunto de variables o características registradas.

Los objetivos del análisis de discriminantes son:

1. Determinar si existe diferencia estadísticamente significativa entre el promedio de valores asociados a grupos definidos a priori.
2. Establecer procedimientos de clasificación entre grupos en base a las variables nuevas.
3. Determinar cuál de las variables independientes es responsable de la mayor parte de las diferencias entre los grupos. Cuál de las variables independientes tiene más peso en la discriminación entre grupos.

El problema se resuelve de la siguiente manera:

Se genera una combinación lineal con las variables independientes, de tal manera que se maximice la varianza intergrupo y se minimice la va

rianza intragrupo; esta combinación lineal se deriva de una ecuación que toma la siguiente forma:

$$Z = W_1X_1 + W_2X_2 + \dots + W_pX_p$$

donde: Z es el valor del discriminante

W es el peso de discriminación asociado a cada variable

X es el valor de la variable independiente.

En el presente estudio se utilizó la técnica de Análisis de Discriminantes para establecer cuáles de las variables socio-económicas tiene mayor poder de discriminación entre los grupos nutricionales, es decir se utilizó el objetivo No. 3.

ANÁLISIS DE RESULTADOS

VARIABLES ANTROPOMÉTRICAS

Como se mencionó en el capítulo correspondiente a material y métodos, se efectuó un análisis de componentes principales, con el fin de obtener un índice antropométrico que resumiera la información obtenida por las tres variables originales (talla, peso y circunferencia de brazo) y que las substituyera en los análisis posteriores.

El análisis mostró que el 80% de la varianza total era explicada por la primera componente, por lo cual se consideró que ésta resumía de manera adecuada la varianza antropométrica y que podría substituir a las variables originales.

Con el objeto de definir grupos que representaran diferentes categorías nutricionales y que se formaran a partir de criterios matemáticos, se realizó un análisis de conglomerados con el nuevo índice antropométrico que resultó del análisis de componentes principales.

El análisis de conglomerados se hizo mediante el programa de computadora BINNORMIX. Este programa describe las características de cada uno de los grupos que genera y las probabilidades de que existan. Esto lo hace de una manera iterativa, permitiendo evaluar cuantos grupos considere necesarios el investigador. Dado el tamaño de la muestra ($n=113$) se decidió evaluar hasta el límite de seis grupos; los resultados aparecen resumidos en el Cuadro 3.

De los grupos formados decidimos trabajar únicamente con el grupo 5 ya que presentaba la mejor distribución. Posteriormente se hizo un reagrupamiento de éste, de lo que resultaron finalmente tres categorías nutricionales:

- Grupo 1: individuos con valores de antropometría altos
(media 1.4261)
- Grupo 2: individuos con valores de antropometría intermedios
(0.8048 y 0.0305)
- Grupo 3: individuos con valores de antropometría bajos
(-1.0305 y -2.1107)

Es importante recalcar que no existe ningún criterio con bases matemáticas que sea del todo satisfactorio para decidir el número de grupos; generalmente queda en manos del investigador, en su experiencia y en un análisis cuidadoso de los grupos formados.

Para sustentar la validez de los grupos nutricionales se hicieron las siguientes pruebas:

a) análisis de varianza (prueba de F) con los valores de talla, peso, circunferencia de brazo y peso para talla, expresados en porcentajes de adecuación. El resultado de la prueba y el rango de los grupos se resume en el cuadro 4; donde se puede apreciar que en todos los grupos los valores de F fueron significativos. La prueba de comparaciones múltiples mostró que todos los grupos presentaban diferencias estadísticamente significativas entre sí, lo cual verifica las diferencias entre éstos en cuanto a sus características antropométricas.

b) pruebas de J_i^2 entre los grupos nutricionales y factores de riesgo. Los resultados se resumen en el cuadro 5, en donde se observa que la asociación con los diferentes factores de riesgo nutricional es más consistente en la clasificación hecha con el análisis de conglomerados; estas diferencias seguramente son el resultado de los diferentes valores de sensibilidad y especificidad de cada una de las clasificaciones.

VARIABLES SOCIOECONOMICAS

Debido a que las variables socioeconómicas están muy correlacionadas y que un número tan grande de variables resulta difícil de manejar e interpretar, realizamos un análisis de componentes principales, con el objeto de reducir el número de variables y estudiar las relaciones que existen entre éstas.

De las 47 variables iniciales se seleccionaron 23 para ser incluidas en el análisis; el resto de ellas se eliminaron, ya que presentaban variabilidad mínima (los individuos de la muestra respondieron en la misma forma), porque mostraron comunalidades* muy bajas en fases posteriores del análisis. En la mayoría de los casos la eliminación se debió a que presentaban variabilidad mínima.

No existen criterios con fundamentación teórica que permitan calcular el tamaño de muestra adecuado para este tipo de análisis, ni para decidir cuál de las variables observadas aparecen representadas en cada uno de los componentes no excluidos del análisis; los criterios usados en este trabajo son producto de las inclinaciones

* Se define comunalidad como el % de varianza que comparte esta variable con el resto de variables incluidas en el análisis.

de tipo práctico que varios autores sugieren; de ellas, se utilizaron las que se consideraron las más razonables por su grado de rigor estadístico. Para mayor detalle ver (2,3).

En el caso del tamaño de la muestra, Hair recomienda que el número de individuos exceda en radio de 2 ó 3 al número de variables seleccionadas para el análisis (3).

Para decidir qué variables están representadas en la componente estudiada, nos basamos en el siguiente criterio: se consideraron como significativas todas aquellas cuya correlación fuese al menos la mitad de la máxima correlación observada en dicha componente.

Existen varios criterios para seleccionar el número de componentes que se deben extraer de el análisis; los más utilizados son los siguientes:

1. Incluir todas las componentes necesarias para que expliquen en forma conjunta el 85% de la suma de las varianzas observadas.
2. Excluir todos los componentes que tengan varianza estandarizada menor a 1.
3. Excluir componentes que expliquen menos del 5% de la variación total.

Para decidir el número de componentes nos basamos principalmente en el tercer criterio y junto con una observación detallada de las variables que aparecen representadas en cada una de las componentes se decidió trabajar con las primeras cuatro.

INTERPRETACION DE LAS COMPONENTES

En el cuadro 6 aparecen representadas las variables con su me dia y su desviación estandar. Las variables se codificaron de mane- ra ordinal, dando mayor peso a las características que se pensó se asociaban con un mayor riesgo de padecer desnutrición, por lo que los valores altos representan en general a las peores característi- cas.

En el cuadro 7 aparece el porcentaje de varianza explicado. Por cada componente se puede observar que las primeras cuatro repre- sentan al 64.2% de la varianza observada, lo que se considera, por algunos autores, un valor adecuado para este tipo de estudios (3). En el cuadro 8 aparece la matriz de los factores rotados (Varimax), en la cual se basó la interpretación de la relación entre las varia- bles.

Primera componente:

En ésta aparecen todas las variables que tienen relación con las actividades agrícolas.

TENENCIA DE LA TIERRA Y VARIABLES AGRICOLAS

Variables observadas:	Correlación observada:
Extensión de la parcela	.97245
Tipo de cosecha	.93151
Hectáreas para sembrar café	.95085
Total de la cosecha	.9437
Categoría de la ocupación	.76226
Gasto en producción	.42417

Era de esperar que éstas variables coincidieran en la misma componente, ya que se encuentran correlacionadas; el hecho de que la ocupación aparezca representada junto con éstas variables, se explica debido a que ésta en el medio rural gira en torno a la agricultura. Se puede considerar a esta componente como una nueva variable que resume la información que se relaciona con la tenencia de la tierra, tipo y cantidad de cosecha y ocupación.

Segunda componente:

CONDICIONES DE VIVIENDA Y NIVEL DE EDUCACION DE PADRE Y MADRE

VARIABLES OBSERVADAS:	CORRELACION OBSERVADA:
Educación del padre	.66408
Educación de la madre	.58688
Material de las paredes	.509046
Material del piso	.64977
Manejo de escretas	.76872
Electricidad	.63770
Material del techo	.33338
Número de cuartos	.38206
Rama de actividad	.34007

Se puede observar que las variables de mayor peso en esta componente son la educación del padre y de la madre y las condiciones de la vivienda.

La correlación que guardan la educación y las condiciones de vivienda reviste gran importancia y nos da la pauta para explicar el peso fundamental que esta componente puede tener en la génesis de la desnutrición. Lo que la componente evidencia es que el nivel educa-

BIBLIOTECA
 JUAN A. ESCOBANTE B.
 UNIDAD ACADÉMICA DE
 LOS CICLOS PROFESIONAL
 Y DE POSGRADO / CCH
 U N A M

tivo está determinando de alguna manera las condiciones de la vivienda. Así, al cuantificar las condiciones de habitación, indirectamente se está cuantificando también el grado de educación de sus moradores.

Tercera componente:

TAMAÑO FAMILIAR

VARIABLES OBSERVADAS:	CORRELACIÓN OBSERVADA:
Número de gentes que trabajan	.46446
Número de dependientes socioeconómicos	.90389
Número de personas que comen en la casa	.95025
Número de personas que duermen en la casa	.94923
Número de cuartos de la vivienda	-.40626

Sin duda esta componente representa el tamaño de la familia. La correlación inversa que se observa entre "número de cuartos" y el resto de las variables se puede explicar por la codificación, ya que valores bajos en la variable "número de cuartos" corresponde a casas con mayor número de cuartos, por lo que la correlación inversa en este caso es lógica.

Cuarta componente:

INGRESO Y GASTO EN ALIMENTACION

VARIABLES OBSERVADAS:	CORRELACIÓN OBSERVADA:
Ingreso per cápita	.80258
Gasto en alimentación	.69710
Seguridad social	.62883
Material del piso	.48590
Material del techo	.53641
Material de las paredes	.56610
Número de cuartos	.50698

Las variables que más peso tienen en esta componente son las de "ingreso" y "gasto en alimentación". Existe una alta correlación con otros índices de vida, como características de la vivienda y seguridad social. Lo que se interpreta de esta componente es que las condiciones de vivienda no están determinadas únicamente por los ingresos familiares. Esto es interesante ya que existen familias con malas condiciones de vivienda que no necesariamente corresponden a familias con bajos ingresos, pero que probablemente correspondan a familias con bajo nivel de educación de los padres (ver segunda componente).

La interpretación que se puede hacer con estas componentes es interesante, pero se limita a describir las correlaciones entre las variables originales; para evaluar su participación como factores de riesgo se analizaron por medio de otras técnicas multivariadas:

a) Se hizo un análisis de discriminantes, en el cual se incluyó como variable dependiente a los valores de la clasificación nutricional, que se obtuvo por medio de el análisis de conglomerados con el índice antropométrico; como variables independientes se usaron las 4 primeras componentes que resultaron del análisis de componentes principales, en el que se incluyeron las 27 variables socio-económicas.

El objetivo de este análisis fue el de evaluar el peso de cada una de las variables como factor de riesgo para desarrollar des

nutrición; esto se evalúa de una manera indirecta con el poder discriminativo de cada variable.

b) Se realizó un análisis de conglomerados usando las 4 componentes obtenidas, para explorar la hipótesis de que dentro de la comunidad existieran grupos estadísticamente diferentes en cuanto a los factores socioeconómicos se refiere y que estuvieran expuestos a diferentes riesgos de padecer desnutrición.

CORRELACION ENTRE VARIABLES SOCIOECONOMICAS Y VALORES ANTRÓPOMETRICOS

ANÁLISIS DE DISCRIMINANTES:

Los resultados del análisis de discriminantes se resumen en el cuadro 9. Se aprecia que las componentes que describen significativamente entre los grupos nutricionales son: la 2 (condiciones de vivienda y nivel de educación de padre y madre) y la 4 (ingreso y gasto en alimentación). Las componentes 1 y 3 no discriminaron significativamente; esto resulta un tanto contradictorio ya que en la literatura se menciona a los factores agrícolas como importantes en la génesis de la desnutrición. Sin embargo, sólo lo es en apariencia, puesto que la mayoría de los individuos de la muestra no cultivan para el autoconsumo, situación en la que las variables agrícolas sí tienen un peso determinante.

ANÁLISIS DE CONGLOMERADOS:

Para este análisis se realizó el mismo procedimiento que se describió en la sección correspondiente a antropometría. Se utilizaron las nuevas variables socioeconómicas (las 4 componentes) para formar los grupos; los resultados mostraron claramente la existencia de 4 grupos (esto se decidió con base en las probabilidades* de que exista un número determinado de grupos contra otro número de éstos). Las características de los diferentes grupos se describen en el cuadro 10.

Se distingue claramente que el grupo 4, ya que es el que presenta los valores más altos**, después le siguen el 2 y 3; aquél presenta valores altos en la primera y tercera componentes, éste presenta valores altos únicamente en la cuarta, mientras que el número 1 se asocia con valores bajos en todas las componentes (hay que recordar que, por codificación, los valores altos representan los valores más bajos en la escala cualitativa). Las características de los grupos de acuerdo a lo que representa cada componente, se podrían resumir de la siguiente manera:

Grupo 1: Esta formado por familias pequeñas, en las que los padres tienen un mejor nivel de educación***, poseen tierras y obtienen buenas cosechas, y un nivel de ingresos medio.

* Probabilidad de la H_0 de que existan 4 contra 3 grupos .00002
 Probabilidad de la H_0 de que existan 5 contra 4 grupos 1.00000

** Los valores altos se refieren a valores de riesgo de padecer desnutrición.

*** Las comparaciones están definidas únicamente para la comunidad, al mencionar a altos o bajos, se hace referencia al promedio dentro de la comunidad.

Grupo 2: Está formado por familias pequeñas, sin tierras, con los mejores índices de vivienda y educación, y los ingresos más altos de la comunidad.

Grupo 3: Está formado por familias con tierras, pero que en conjunto perciben los ingresos anuales más bajos.

Grupo 4: Está formado por familias sin tierra, con vivienda en malas condiciones, bajos índices de educación e ingresos anuales bajos.

Para sustentar la existencia de los grupos socioeconómicos se calcularon las probabilidades de pertenencia a cada uno de los diferentes grupos; se encontraron probabilidades altas ($P > .40$) de pertenecer a dos; sólo en 3 de los 113 casos; ésta confusión fue entre los grupos 3 y 4; el resto de los casos mostraron probabilidades muy bajas ($P < .01$) de pertenecer a un segundo grupo, por lo que se consideró que la clasificación fue muy buena, lo cual traduce que la probabilidad de que existan clases sociales dentro de esta comunidad es muy alta.

Para evaluar si estas clases sociales están expuestas a diferentes riesgos, se correlacionaron éstas con la clasificación nutricional obtenida por el análisis de conglomerados; la asociación fue altamente significativa ($P < .000$). Ver cuadros 11, 12 y 13.

En el cuadro 10, se puede observar que el 77% (27) de los niños con valores bajos de antropometría pertenece al grupo social más bajo de la comunidad, mientras que el 78% de los niños con valores altos se sitúa en los dos estratos más altos de la comunidad, lo

que indica que estos grupos sociales reflejan diferentes riesgos de desnutrición.

En la figura 1 se aprecia la representación geométrica de los grupos sociales; cada número corresponde a una familia e indica al grupo que pertenece, la gráfica está construida en base a los valores obtenidos en las diferentes componentes; en las ordenadas aparecen los valores de educación, ingreso e índices de calidad de vivienda y en las absisas los valores de tenencia de tierra y calidad y cantidad de las cosechas.

Se pueden observar dos grandes conglomerados, que están definidos por los valores de tenencia de la tierra; éstos están formados por los grupos 2 y 4 que no tienen tierra y el 1 y 3 que sí tienen; las diferencias dentro de estos dos grandes grupos está condicionada por los valores que tienen los sujetos en ingreso, educación y vivienda, indicador que define de manera importante el riesgo, ya que los grupos que tenían los valores bajos en estas categorías presentan mayor prevalencia de desnutrición.

En resumen, se puede concluir que existen dentro de las comunidades clases sociales que están expuestas a diferentes riesgos de desnutrición, esto es muy importante para la planeación de estudios en un futuro, ya que los factores sociales, si no se controlan, pueden actuar como factores de confusión interfiriendo de manera drástica con la evaluación de proyectos de intervención. Por otro lado, se podrían utilizar estos grupos para la optimización de nuevos programas, incidiendo únicamente sobre los grupos de alto riesgo.

Se comprueba una vez más que la variable "condiciones de vivienda" es un excelente indicador de riesgo nutricional por su estrecha asociación a ingreso y a educación de los padres.

CONCLUSIONES

Del análisis de resultados se desprende la gran utilidad que los métodos multivariados pueden tener en el campo de la investigación nutricional.

Usando estos métodos se puede lograr una mejor clasificación de los individuos que se encuentran afectados por la desnutrición. Permiten construir grupos con varios indicadores del estado nutricional discriminando de una mejor manera a los sujetos afectados.

Por otro lado, con su uso se logra un mejor entendimiento de los fenómenos biológico-sociales como es la desnutrición; en los que se interrelacionan un gran número de factores causales o de riesgo. Esto permite controlar los factores de confusión inherentes a los individuos estudiados aumentando la calidad de la inferencia que se pueda obtener de las investigaciones nutricionales.

Su inclusión en los paquetes estadísticos de las computadoras facilita grandemente su uso por cualquier investigador a costo muy bajo.

Este estudio continuará; se harán evaluaciones con un mayor número de factores de riesgo en un proyecto longitudinal con duración de varios años, lo que permitirá demostrar la gran utilidad de estos métodos en la investigación para describir de una manera más integral el fenómeno de la desnutrición, a los individuos afectados por ella y a los que se encuentran en alto riesgo de padecerla.

BIBLIOGRAFIA

1. Gómez, F. Desnutrición. Boletín Médico del Hospital Infantil. Vol. 3, 1946.
2. Hernández, A., Pardiñas, F., Ruiz, F. Cortadores de Caña en México, Vol. IV. 1983. Fideicomiso para obras sociales a campesinos cañeros de escasos recursos.
3. Joseph, F., Hair, Jr. R., Anderson, E. Multivariate Data Analysis, 1979. Petroulium Publishing Company
4. Robert, G.D., Steel, J. Torrie, H. Principles and Procedures of Statistics. A Biomedical Approach, Second Edition, 1980. Mc. Graw Hill International Book Company
5. Neil, H. T. Multivariate Analysis with Aplications in Education and Psychology. 1975. Wadsworth Publishing Company.
6. Brian Everit. Cluster Analysis, 1977. Herineman Educational Book.
7. Norman H. Nie; Hadlai Hull, C. et.al. Statistical Package for the social Sciences, 1970. McGraw-Hill Book Company, Sencond Ed.
8. Hernández, A. Problems in cluster analysis (1979) Tesis doctoral Oxford University.
9. Donald F. Morrison. Multivariate Statistical Methods, 1976. Mc Graw-Hill International Book Company, Second Edition.

CUADRO 1

CLASIFICACION DE GOMEZ

GRADO DE DESNUTRICION	% DEL PESO IDEAL PARA LA EDAD
NORMAL	91 a 100%
I	76 a 90%
II	61 a 75%
III	60% ó menos

CUADRO 2

EFFECTO DE LAS VARIACIONES DE SENSIBILIDAD
Y ESPECIFICIDAD SOBRE LA POTENCIA

SENSIBILIDAD	ESPECIFICIDAD	DIF. ESPERADA	POTENCIA
	.90	.0250	.3098
.60	.95	.0275	.447
	1.0	.0300	.7419
	.90	.0350	.4586
.80	.95	.0375	.6113
	1.0	.0400	.8479
	.90	.0450	.6041
1.0	.95	.0475	.7460
	1.0	.05	.9147

Cuadro 1. Tomado de "QUADE" variaciones en la diferencia esperada y la potencia de la prueba en relación a la sensibilidad y especificidad.

CUADRO 3

MEDIAS Y TAMAÑOS DE LOS GRUPOS GENERADOS
EN EL ANALISIS DE CONGLOMERADOS

n	2 GRUPOS	n	3 GRUPOS	n	4 GRUPOS	n	5 GRUPOS	n	6 GRUPOS
51	.6401	31	1.1871	31	1.2463	16	.8440	17	.8440
62	-.7481	53	.0790	47	.0804	18	1.4921	18	1.4921
		29	-1.0869	31	-1.0161	44	.0799	40	.0789
				4	-2.0925	31	-.7708	23	-.7708
						4	-1.6138	11	-1.4449
								4	-2.1883

CUADRO 4

DEFINICION DE LOS GRUPOS ANTROPOMETRICOS OBTENIDOS EN ANALISIS DE CONGLOMERADOS, EXPRESADOS EN RANGOS DE PORCENTAJE DE ADECUACION

	TALLA	PESO	C. BRAZO	TALLA-PESO	n
F Prob.	.0000	.000	.0000	.0000	n
GRUPO 1	96 a 100	92 a 100	96 a 100	91 a 100	18
GRUPO 2	84 a 100	75 a 94	87 a 100	83 a 100	60
GRUPO 3	82 a 95	62 a 80	78 a 97	79 a 100	35

CUADRO 11

CORRELACION ENTRE CLASIFICACIONES

GRUPOS FORMADOS CON VARIABLES ANTROPOMETRICAS	GRUPOS FORMADOS CON VARIABLES SOCIOECONOMICAS			
	1	2	3	4
1	4	10	0	4
2	6	13	8	33
3	1	4	3	27

P < .0001

CUADRO 5

PRUEBA DE J₁ CON LOS DIFERENTES FACTORES DE RIESGO

	CLUSTER	PESO	TALLA	C.B.
HECTAREAS DE CAFE	*	*	NS	NS
TOTAL CAFE	*	*	NS	NS
NO. DEPENDIENTES	NS	NS	NS	NS
SEGURIDAD SOCIAL	*	NS	*	NS
GASTO EN TRANSPORTE	*	NS	*	NS
GASTO EN ALCOHOL	*	*	NS	NS
EDUCACION DEL PADRE	NS	NS	NS	*
EDUCACION DE LA MADRE	*	NS	NS	*
MATERIAL DE TECHO	*	*	*	*
MATERIAL DE PAREDES	*	*	*	
MATERIAL DE PISO	*	*	*	*
FUENTE DE AGUA	*		*	*
EXCRETAS	*	*	*	*
ELECTRICIDAD	*	*	NS	*
NO. DE CUARTOS	*	*	*	*
OCUPACION	NS	NS	*	NS
NO. DE PERSONAS QUE COMEN	NS	*	*	NS

VARIABLES INCLUIDAS EN EL ANALISIS

VARIABLE	MEDIA	DESVIACION ESTANDAR
V ₁	1.8053	0.3977
V ₂	2.7168	0.6192
V ₃	2.7257	0.6013
V ₉	3.6283	0.8259
V ₁₂	1.9292	1.0498
V ₁₃	3.3009	1.4008
V ₁₅	4.2832	1.8345
V ₁₆	1.8584	0.3502
V ₂₂	3.1150	1.1554
V ₂₄	2.9505	0.8170
V ₂₆	3.3186	0.5711
V ₂₈	2.0708	0.8835
V ₂₉	2.0619	0.7354
V ₃₀	1.6106	0.4898
V ₃₁	1.3451	0.4775
V ₃₂	1.6106	0.4898
V ₃₃	1.4956	0.5022
V ₃₄	3.1239	0.8466
V ₃₅	6.3363	2.0205
V ₃₆	6.3363	2.0161
V ₃₇	2.4956	0.7920
IPC	4.3540	2.1995
GPC	2.9115	0.9020

CUADRO 7

FACTORES Y VARIANZA EXPLICADA

FACTOR	% DE VAR.	% ACUM.
1	26.4	26.4
2	16.1	42.5
3	13.0	55.5
4	6.3	61.8*
5	5.0	66.8
6	4.7	71.5
7	4.1	75.6
8	3.7	79.2
9	3.3	82.5
10	2.9	85.4
11	2.8	88.2
12	2.3	90.5
13	1.9	92.3
14	1.6	93.9
15	1.4	95.3
16	1.1	96.4
17	1.0	97.3
18	0.8	98.2
19	0.6	98.8
20	0.5	99.3
21	0.4	99.7
22	0.1	99.9
23	0.1	100.0

MATRIZ DE COMPONENTES ROTADA (VARIMAX)

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
V ₁	0.97245	0.06268	0.02391	-0.01734
V ₂	0.93151	0.03504	-0.03392	-0.07399
V ₃	0.95081	0.08428	0.00536	0.04686
V ₉	0.94437	0.09329	0.05559	0.05697
V ₁₂	0.24850	0.27116	0.46446	-0.29645
V ₁₃	-0.30216	0.34007	-0.01224	0.37305
V ₁₅	0.94792	0.08252	0.90389	0.04937
V ₁₆	-0.20775	0.19520	-0.15152	0.62883
V ₂₂	0.42417	0.38793	-0.21011	0.16267
V ₂₄	-0.01899	0.66408	0.16535	-0.03450
V ₂₆	0.06872	0.58688	0.11274	-0.01868
V ₂₈	0.14142	0.50946	-0.00242	0.56610
V ₂₉	0.36495	0.33358	-0.00007	0.53641
V ₃₀	0.18365	0.64977	-0.09273	0.48590
V ₃₁	0.2657	0.28865	0.00714	0.28481
V ₃₂	0.10305	0.76872	0.04530	0.28494
V ₃₃	0.15162	0.63770	0.14278	0.17416
V ₃₄	0.21385	0.38206	-0.40626	0.50698
V ₃₅	-0.09795	0.12175	0.95025	0.13118
V ₃₆	-0.09838	0.11207	0.94923	0.13151
V ₃₇	0.76226	0.19631	-0.06006	0.18366
IPC	0.22199	0.16274	0.25819	0.80258
GPC	-0.01792	-0.17552	0.26780	0.69710

CUADRO 9

RESULTADOS DE ANALISIS DE DISCRIMINANTES

VARIABLES	F	SIGNIFICANCIA
V ₁ TENENCIA DE TIERRA Y VARIABLES AGRICOLAS	1.52	.3257
V ₂ INDICE DE VIDA Y EDUCACION DE LOS PADRES	3.582	.0311**
V ₃ TAMAÑO FAMILIAR	.5487	.5793
V ₄ INGRESO E INDICADORES DE VIDA	13.56	.00000***

CUADRO 10

GRUPOS FORMADOS CON VARIABLES SOCIOECONOMICAS

VARIABLES	GRUPOS SOCIOECONOMICOS FORMADOS			
	1	2	3	4
Variables agropecuarias	-2.2790	0.579	-1.6713	.4632
Indices de vida y educación	-0.2135	-1.0220	-0.0410	.4574
Tamaño familiar	-0.0544	.2477	-0.0425	-0.0841
Ingreso	-0.3804	-0.9565	0.4324	.3765

Los valores altos indican mayor riesgo de desnutrición.

CUADRO 12

DISTRIBUCION DE LOS GRUPOS ANTROPOMETRICOS	VARIABLES SOCIO-ECONOMICAS			
	1	2	3	4
ANTROPOMETRIA				
1	22.5	55	0	22.5
2	10	21	14.5	56
3	3	11	9	77

CUADRO 13

DISTRIBUCION DE LOS GRUPOS DE RIESGO DE DESNUTRICION
(SOCIOECONOMICOS)

ANTROPOMETRIA	VARIABLES SOCIO-ECONOMICAS					
	1	2	3	4	5	6
1	36	37	0	6		
2	54	48	72	51		
3	10	5	27	43		

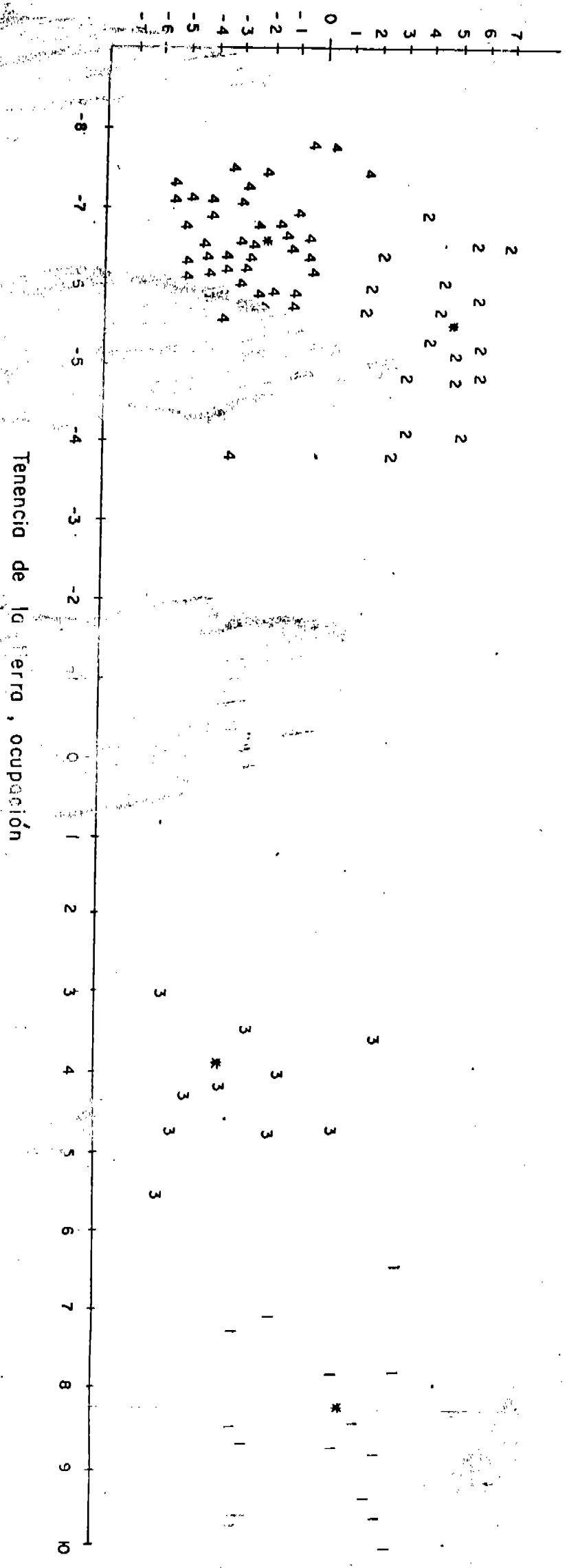


FIGURA 1