

---

# **UTILIZACIÓN DE TÉCNICA ESTADÍSTICA “AGRUPAMIENTO EN DOS ETAPAS” PARA VALORACIÓN NUTRICIONAL EN COMUNIDADES RURALES DE VENEZUELA**

Rodriguez María del V.  
Asomuseo, Guanare, estado Portuguesa  
mvalle@ula.ve

## **Resumen**

Para construir un perfil de hogares con base en variables nutricionales, en este trabajo se planteó como objetivo establecer grupos de comunidades, afines según variables de valoración nutricional continuas y categóricas, a través de una técnica de agrupamiento estadístico: la técnica de agrupamiento (*Cluster*) en dos etapas; ésta técnica fue usada para obtener datos de las comunidades beneficiarias del Programa Especial Seguridad Alimentaria (PESA), ejecutado por la FAO y el Ministerio de Agricultura, en asentamientos rurales de Venezuela. Los datos provienen de la “Encuesta de Seguridad Alimentaria en el Hogar” (ESAH), aplicada a comunidades rurales en los años 2003 y 2006, la cual consideró las siguientes variables: Requerimiento calórico diario, Disponibilidad calórica diaria, Disponibilidad calórica por producción propia, y Nivel de suficiencia alimentaria. Los datos fueron procesados con el paquete estadístico SPSS, versión 12. Los resultados muestran clusters bien diferenciados: por un lado, el cluster “Bajo nivel de suficiencia alimentaria”, y por el otro, el cluster “Alto nivel de suficiencia alimentaria”. En el primero, el número de familias creció desde la primera encuesta (ESAH 1) a la segunda (ESAH 2). La conclusión es que, aun cuando el ingreso familiar crezca, la mejoría de las variables nutricionales no lo hace en la misma proporción, presentándose incluso el deterioro en estas variables.

**Palabras clave:** valoración, nutrición, clusters, suficiencia alimentaria.

## Introducción

El rol de los patrones espaciales, junto a la caracterización de grupos de hogares, basándose en medidas de bienestar, representa un campo poco estudiado. Los altos o bajos estándares de vida en determinadas áreas tienen un efecto amplificador en zonas aledañas, que derivan en la conformación grupos con características bien definidas [Hyman *et al.*, (2005)].

Partiendo de esto, se pretende incorporar el análisis estadístico a la investigación social y nutricional, a fin de obtener respuestas que ayuden al gobierno central y demás entes involucrados, a formular políticas que mejoren la situación de los grupos más desfavorecidos.

La construcción de un mapa de estudio del consumo alimentario de la población puede constituirse en una herramienta importante para diseñar políticas eficientes, en pro de mejorar la capacidad adquisitiva de los hogares pobres, o para el diseño de esquemas de intervención y localización de transferencias o subsidios. Es sabido que los recursos se usan mejor si los grupos necesitados pueden ser reconocidos más cercanamente. Es decir, si el gasto social se orienta en función de un mapa de pobreza detallado, podrían reducirse los errores de exclusión (pobres no beneficiados respecto a la población total) e inclusión (no pobres beneficiados en relación a la población total) de los programas sociales.

Una razón adicional, no menos importante, para construir mapas con estos criterios, es la posibilidad de combinarlos con otras fuentes de información, con el propósito de

capturar aspectos adicionales de la pobreza y desigualdad, que no son fáciles de medir a través de una encuesta o censo. En tal sentido, el objetivo de la investigación fue el de establecer grupos de comunidades afines, según variables de valoración nutricional, a través de una técnica de agrupamiento o cluster, en dos etapas.

## Materiales y Métodos

Como corpus de datos, se tomó en cuenta la Encuesta de Seguridad Alimentaria en el Hogar. En este sentido, se trabajó con las dos versiones aplicadas en las comunidades a estudiar: ESAH1 (aplicada en el año 2003) y ESAH 2 (aplicada en el año 2006). Este instrumento fue utilizado por el programa PESA-FAO, en las siguientes comunidades venezolanas: Los Altos y Los Ranchos (estado Monagas), Hato Arriba y La Laguna (estado Lara), La Pared (estado Nueva Esparta), Las Viejas (estado Trujillo), Maciegal (estado Mérida), Las Barbas (estado Guárico).

Por otra parte, tomamos la Hoja de consumo de alimentos, que consiste en un registro de todos los alimentos consumidos en cada comunidad intervenida, con sus respectivas frecuencias de adquisición, de consumo, lugar de adquisición, disponibilidad en gramos para cada alimento y sus respectivos aportes calóricos y de nutrientes. Esta Hoja de consumo de alimentos fue elaborada por el programa PESA-MAT-FAO.

Asimismo, se consideraron algunos datos referidos a *Requerimientos alimentarios*, valores que fueron calculados por el subcomponente

“Nutrición” del Programa PESAMAT-FAO. Los datos se referían a los valores de disponibilidad calórica, requerimientos calóricos, porcentaje de autoabastecimiento, disponibilidad calórica, y producción propia para cada familia de las comunidades intervenidas. En relación a las técnicas estadísticas a emplear, se trata de técnicas cuya principal virtud es la de permitir clasificar numerosos elementos en función de criterios pre establecidos. Un *cluster* es una colección de objetos o datos que son similares a otros, y que pueden ser tratados colectivamente como un grupo. Un buen método de agrupamiento produce grupos con gran similaridad de los elementos dentro de un grupo, así como baja similaridad entre grupos. Ya que el objetivo de estas técnicas es agrupar objetos similares, se necesita alguna medida para evaluar las diferencias y similitudes entre objetos. La estrategia más común consiste en medir las similitudes en términos de la distancia entre pares de objetos. Los objetos con distancias reducidas entre ellos, son más parecidos entre sí que aquellos que tienen distancias mayores. Existen varias formas de calcular estas distancias, siendo una de ellas el agrupamiento en dos etapas (Banfield y Raftery, 1993). Se trata de un modelo de distancias que permite el uso de variables categóricas y continuas, además de que permite el procesamiento de gran cantidad de datos, cosa que no es posible con el empleo del cluster jerárquico. Este agrupamiento tiene la ventaja de determinar el número óptimo de clusters.

En una primera etapa, se usa una

aproximación al agrupamiento secuencial y se comprime la base de datos para obtener subclusters, donde este agrupamiento revisa cada registro y decide si debe unirse con el registro anterior y formar un cluster, o bien comenzar uno nuevo con ese registro como primera observación.

En una segunda etapa, los subclusters resultantes de la primera etapa son usados como insumos, agrupándose en un número óptimo de clusters.

Así pues, este estudio empleará la técnica de clusters en dos etapas, a partir de ciertas variables de valoración nutricional, a saber: a) Continuas: requerimiento neto diario; Disponibilidad calórica diaria; Disponibilidad calórica por producción propia; Nivel de suficiencia alimentaria; b) Categóricas: Comunidad y Nivel de suficiencia alimentaria.

Estas variables forman parte de la Encuesta de Seguridad Alimentaria en el Hogar (ESAH), aplicada por el Programa Especial de Seguridad Alimentaria, ejecutado por la FAO con el apoyo del Ministerio de Agricultura y Tierras en diversas comunidades del país.

**Tabla 1.** Estadísticas descriptivas ESAH 1 Variables de valoración nutricional

	N	Mín	Máx	Media	Desv.	Sesgo	Curtosis
	*	*	*	*	*	*	**
Requerimiento Neto Diario	184	0	54010	12153,81	7526,686	1,786	0,179 5,400 0,356
Disponibilidad Calórica Diaria	184	0	56870	9289,2577	6832,537	2,964	0,179 15,276 0,356
Disponibilidad Calórica Prod propia	184	0	17139,64	1043,0064	2230,924	4,165	0,179 20,595 0,356
Nivel Suficiencia alimentaria	184	0	357,51	88,9727	57,11080	1,474	0,179 3,698 0,356

Fuente: Datos Programa Especial Seguridad Alimentaria. Sub componente Nutrición

\* Estadístico; \*\* Desviacion

**Tabla 2.** Estadísticas descriptivas ESAH 1: variables de valoración nutricional**ESTADÍSTICAS DESCRIPTIVAS. ESAH1. VARIABLES DE VALORACION NUTRICIONAL**

	N	Minimum	Maximum	Mean	Std.	Skewness	Std. Error	Kurtosis	Std. Error
	Statistic	Statistic	Statistic						
Requerimiento Neto Diario	184	,00	54010,00	12153,81	7526,686	1,786	,179	5,400	,356
Disponibilidad Calorica diaria	184	,00	56870,00	9289,2577	6832,537	2,964	,179	15,276	,356
Disponibilidad Calorica Propia	184	,00	17139,64	1043,0064	2230,924	4,165	,179	20,595	,356
Nivel de Suficiencia Alimentaria	183	,00	357,51	88,9727	57,11080	1,474	,180	3,698	,357
Valid N (listwise)	183								

**Tabla 3.** Estadísticas descriptivas ESAH 2: variables de valoración nutricional**ESTADÍSTICAS DESCRIPTIVAS. ESAH2. VARIABLES DE VALORACION NUTRICIONAL**

	N	Minimum	Maximum	Mean	Std.	Skewness	Std. Error	Kurtosis	Std. Error
	Statistic	Statistic	Statistic						
Requerimiento Neto Diario	129	1975,00	29210,26	11266,19	5477,263	,969	,213	,941	,423
Disponibilidad Calorica diaria	129	,00	43092,86	9433,1674	6066,788	2,057	,213	7,205	,423
Disponibilidad Calorica Propia	129	,00	8034,57	905,5624	1577,404	2,674	,213	7,773	,423
Nivel de Suficiencia Alimentaria	129	,00	294,75	92,0693	53,38478	1,331	,213	1,967	,423
Valid N (listwise)	129								

Dicha encuesta se aplicó en dos versiones para esas comunidades. En la primera versión, se encuestaron 200 familias, distribuidas tal como se muestra en la Tabla anterior, mientras que en la segunda versión, se encuestaron 160 familias. No se realizó un muestreo, pues sólo se trabajó con aquellas comunidades que eran beneficiarias del Programa Especial de Seguridad Alimentaria para ese momento; por lo tanto, las conclusiones estadísticas sólo serán aplicadas a los

hogares encuestados y, de ninguna manera, a la generalidad de ellos.

En relación a las variables nutricionales analizadas, es preciso decir que las tres primeras (Requerimiento calórico neto diario, Disponibilidad calórica diaria, Disponibilidad calórica por producción propia) están expresadas en calorías / día, mientras que el Nivel de suficiencia alimentaria (1) es un porcentaje derivado de:

$$\text{Nivel Sufic} = \frac{\text{Disponibilidad Diaria} \times 100}{\text{Requerimiento Calorico}} \dots \dots \dots \quad (1)$$

Todas estas variables están medidas para una unidad de análisis: Hogar. Las variables de valoración nutricional no se estandarizan, pues el agrupamiento en dos etapas (Cluster en dos etapas) no lo amerita; sin embargo, deben cumplirse los supuestos de independencia entre variables continuas y de normalidad de éstas. Además, de que se asume que las variables categóricas tienen una distribución multinomial.

La variable REQUERIMIENTO CALÓRICO NETO DIARIO (2) está referida a las calorías requeridas por una familia, para un buen desenvolvimiento de sus actividades físicas y mentales. Estos requerimientos van a variar de acuerdo a la edad, nivel de actividad física y a si hay mujeres embarazadas y lactantes en el hogar. Esta variable se expresa de la siguiente manera:

$$\text{Requerimiento K cal Netas/dia} = \frac{\text{Requerimiento K cal/dia} \times \text{Participacion comidas al dia}}{\text{Nro. integrantes x familia}} \quad (2)$$

La variable Nivel de Suficiencia alimentaria está presente en esta investigación en forma numérica continua y en forma categórica. La

primera de ellas, el Nivel de suficiencia alimentaria numérica continua (3) es el valor resultante de

$$\text{Nivel Suficiencia Alimentaria} = \frac{\text{Disponibilidad Calorica Familiar/diaria} * 100}{\text{Requerimiento Neto Diario}} \quad (3)$$

Esta es una variable que tiene 3 niveles, y esto es lo que da origen a la variable Nivel de suficiencia alimentaria categórica. Los valores o puntos de corte estructuran la variable en su forma categórica, tal como se indica a continuación:

Nivel suficiencia alimentaria menor de 90% = por debajo de la norma.

Nivel suficiencia alimentaria entre 90 y 110% = normal.

Nivel suficiencia alimentaria mayor de 110% = por encima de la norma.

En cuanto a la variable **DISPONIBILIDAD CALÓRICA DIARIA**, Abreu y Ablan (2004) señalan que se ha determinado que en países de ingresos bajos existe una relación directa entre el ingreso real per cápita y su consumo de energía alimentaria (disponibilidad energética/persona/día). Por otro lado, algunos investigadores, en vez de enfocarse en la variable meramente nutricional (disponibilidad calórica diaria), han avanzado e incorporado una dimensión económica a ésta, resultando en la variable como “Disponibilidad energética para consumo humano al nivel de venta al detal”, variable “que ha sido considerada como una buena aproximación a la magnitud de la variable consumo”, (Abreu y Ablan, 2004).

En cuanto al cálculo de la Disponibilidad calórica, éste se hizo para cada familia de las comunidades estudiadas, tomando como referencia los valores promediados a nivel nacional, los cuales se especifican en la Tabla de Valores de Referencia de Energía y Nutrientes del INN.

La variable Disponibilidad calórica por producción propia se obtuvo de la encuesta ESAH, y se calculó igualmente

para cada familia. Esta variable está referida a los aportes calóricos derivados de algún proceso de producción doméstico.

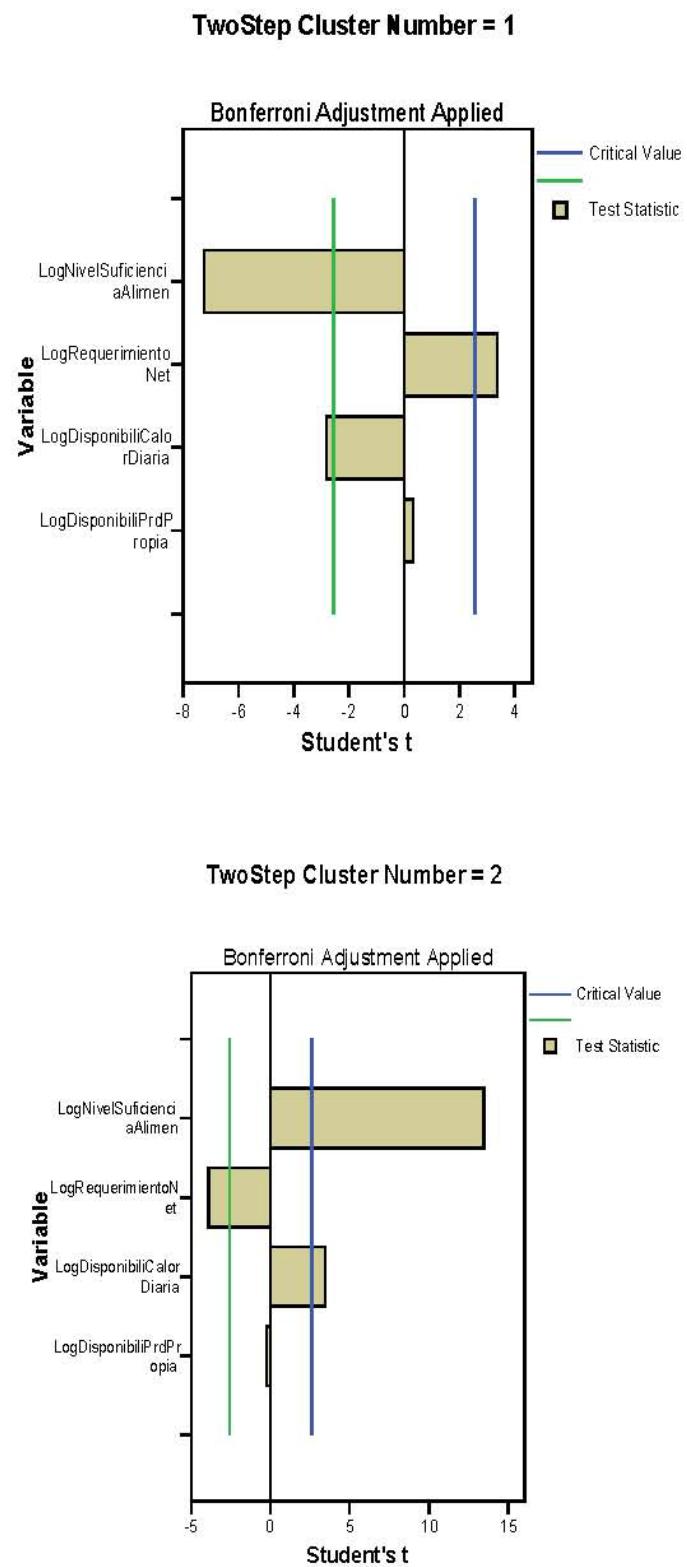
## Resultados y Discusión

### **Agrupamiento en dos etapas: ESAH 1.**

Con los resultados obtenidos, se conformaron dos grupos o clusters, tomando como criterio la variable Nivel de suficiencia alimentaria, resultando en: Cluster 1, “Nivel de suficiencia alimentaria baja” y Cluster 2, “Nivel de suficiencia alimentaria alta”.

Al cluster 1 lo llamaremos “Bajo Nivel de Suficiencia Alimentaria”; sus valores promedios de Requerimiento neto son mayores que la Disponibilidad calórica diaria, mientras que la Disponibilidad calórica derivada de la producción propia es aún menor. Esto origina que la característica de este cluster (las 81 filas que lo componen) sea el de presentar valores que están por debajo de la norma nutricional. Este cluster tiene como característica que el 100% de las familias encuestadas se encuentran en situación de inseguridad alimentaria.

El ajuste de Bonferroni nos muestra la significancia de cada variable en la formación del cluster, con base en el estadístico T de student. En este caso, la variable más significativa en la formación de este cluster es el Nivel de suficiencia alimentaria, seguido del Requerimiento neto, que como era de esperarse, asume valores por debajo de la media, y correspondiendo con el grupo que identificamos por debajo de la norma o en situación de inseguridad alimentaria.



**Figura 1.** Ajustes de Bonferroni para variables CLUSTER 1 y 2 (ESAH 1)

La Disponibilidad calórica por producción propia no se muestra, en absoluto, significativa en la formación del cluster.

Para el cluster 2, a partir del ajuste de Bonferroni, se encuentra que la variable más significativa en su formación es el Nivel de suficiencia alimentaria, el cual asume valores por encima de la media. Seguidamente, está Requerimiento neto, que es igualmente significativo, pero que en este cluster asume valores por debajo de la media. La variable Disponibilidad calórica diaria es en este cluster mucho más significativa que en el cluster 1.

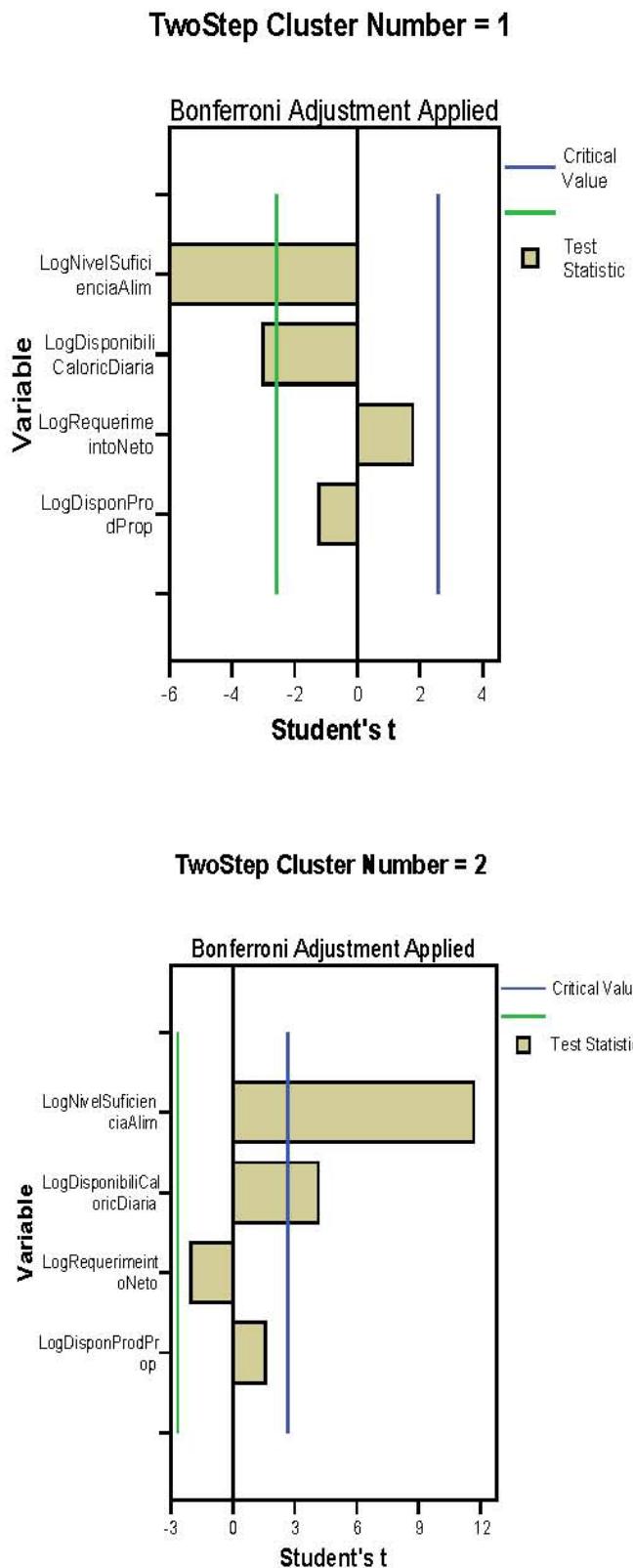
#### **Agrupamiento en dos etapas: ESAH 2.**

Con los resultados obtenidos, se conformaron dos grupos o clusters, tomando como criterio la variable Nivel de suficiencia alimentaria. Estos son: Cluster 1, “Nivel de suficiencia alimentaria baja” y Cluster 2, “Nivel de suficiencia alimentaria alta”.

El cluster 2 será llamado “Alto nivel de suficiencia alimentaria”, ya que la media de la variable Nivel de suficiencia alimentaria refleja un valor por encima de la norma. En este cluster vemos que la Disponibilidad calórica diaria es mayor que el Requerimiento neto diario, lo cual indica que estos hogares cubren sus requerimientos nutricionales.

La aplicación del ajuste de Bonferroni para las variables del cluster 1, nos indica la significancia de éstas en la formación del cluster, con base en el estadístico t de Student para variables continuas. En tal sentido, en el cluster 1 la variable más significativa fue el

Nivel de suficiencia alimentaria y, en menor grado, la Disponibilidad calórica diaria. Las variables restantes, esto es, el Requerimiento neto y la Disponibilidad por producción propia no se mostraron significativas en la conformación del cluster. En este caso, ambas variables, Nivel de suficiencia alimentaria y Disponibilidad calórica diaria, asumen valores por debajo de la media de cada variable, lo que se corresponde con el comportamiento del cluster; es decir, con hogares con mas requerimientos calóricos que disponibilidad, pues esta última variable asume valores negativos en el cluster.



**Figura 2.** Ajuste de Bonferroni para variables Clusters 1 y 2 (ESAH 2)

En el cluster 2, igualmente, se muestran significativos el Nivel de suficiencia alimentaria y la Disponibilidad calórica diaria, sólo que en este cluster estas variables asumen valores por encima de la media y valores positivos. En este caso, la disponibilidad calórica de estos hogares es mayor que el requerimiento. En promedio, para ambas encuestas (ESAH 1 y ESAH 2), las familias con alto nivel de suficiencia alimentaria sobrepasan, en 1.5 veces, los requerimientos calóricos netos diarios, puesto que disponen de 12098,02 cal/familia, cuando el requerimiento es de 8295,86 calorías/familia.

Las familias con bajo nivel de suficiencia alimentaria cubren solo el 68 % de sus requerimientos calóricos netos diarios, ya que disponen sólo de 7177 cal/familia, cuando el requerimiento es de 12.226 cal/ familia.

La técnica de agrupamiento en dos etapas arrojó resultados igualmente interesantes. Para ambas versiones de la encuesta, y con base en las variables nutricionales, los grupos obtenidos siempre fueron dos: uno con nivel de suficiencia alimentaria “baja” y otro con nivel de suficiencia alimentaria “alta”.

En ambas encuestas, el grupo Bajo nivel de suficiencia alimentaria mostró como característica que sus niveles de Requerimiento neto diario son mayores que los de Disponibilidad calórica diaria. Por contraparte, el grupo Alto nivel de suficiencia alimentaria muestra niveles de Disponibilidad calórica ligeramente mayores que los requeridos.

## Conclusiones

Para el caso de ambas encuestas (ESAH 1 y ESAH 2), en el cluster de Bajo nivel de suficiencia alimentaria, el promedio de requerimiento neto diario es de 12.260,59 calorías/familia, mientras que la Disponibilidad calórica diaria es de 7.177,286 calorías /familia. Esto representa un requerimiento calórico promedio aproximado de 2322,08 calorías/persona/día, y una disponibilidad calórica promedio aproximada de 1359,33 calorías/ persona/día.

Por otra parte, para ambas encuestas, en el cluster Alto nivel de suficiencia alimentaria, el promedio de Requerimiento Neto Diario es de 8.295,86 calorías/familia, siendo la disponibilidad calórica diaria promedio de 12.098,019 calorías/familia. Es decir, un requerimiento neto diario de 1.797,58 calorías/persona/día y una disponibilidad calórica promedio de 2.621,45 calorías/ persona/día.

El cálculo de valores promedio por persona/día se hizo en función de un promedio aritmético de las comunidades que conforman el cluster dividido, para el número promedio de personas por familia de esas comunidades; este promedio es sólo referencial y no refleja las características físicas de cada miembro de la familia. Para el cálculo de los requerimientos calóricos, éste sólo es usado a fines de tener valores referenciales de consumo y requerimiento calórico por persona; los promedios usados para la unidad de estudio “familia” si se hallan ajustados.

**Tabla 4.** Resumen de agrupamiento en dos etapas por criterios.

Criterio		Alto nivel de suficiencia alimentaria	Bajo nivel de suficiencia alimentaria
Promedio de personas por familia	ESAH 1	5,22	
	ESAH 2	5,29	
Requerimiento neto diario por familia	ESAH 1	8295,86	12260,59 cal/fam
	ESAH 2	8295,86	12260,59 cal/fam
Disponibilidad calórica diaria por familia	ESAH 1	12098,02	7177,28 cal/fam
	ESAH 2	12098,02	7177,28 cal/fam
Requerimiento por persona	ESAH 1	1797,58	2322,08 cal/perso
	ESAH 2	1797,58	
Disponibilidad calórica diaria por persona	ESAH 1	2621,45	1359,33 cal/perso
	ESAH 2	2621,45	

Según la tabla de requerimientos de energía y nutrientes de la población venezolana (mayo 2000) del INN-CONICIT, el promedio de ingesta calórica promedio para una persona (con excepción de embarazadas, niños, atletas y ancianos) es de 2300 calorías /persona/día. En el grupo de bajo nivel de suficiencia alimentaria, esta ingesta calórica es de 1359,33 calorías/persona/día, mientras que el grupo de alto nivel de suficiencia alimentaria presenta una ingesta calórica en promedio es 2621,45 calorías/persona/día.

De los anterior, se concluye que, aún cuando el ingreso familiar crezca, la mejoría de las variables nutricionales no lo hace en la misma proporción, pudiéndose incluso presentar el empeoramiento de estas variables. Esto se halla relacionado con problemas en el acceso oportuno y eficiente a los alimentos por parte del responsable del hogar, más que con problemas “económicos” que imposibiliten la compra de éstos.

### Agradecimiento

La asesoría nutricional y uso de datos fue posible gracias al Subcomponente Nutrición del Programa PESA-MAT-FAO, en las personas de Siloyde Rivas y René Sleiman.

Por el apoyo estadístico y académico, al Profesor Giampaolo Orlandoni y al Instituto de Estadística de Facultad de Ciencias Económicas de la ULA (Mérida).

### Referencias Bibliográficas

- Abreu Olivo, E.; Ablan, E. (2004). ¿Qué ha cambiado en Venezuela desde 1970 en cuanto a la disponibilidad de alimentos para el consumo humano? *Agroalimentaria*, 19, 13-33.
- Banfield, J.; Raftery, A. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 49:803-821.
- Castillo, R.; Morales, A. (2004). Propuesta para interpretar los cambios en el consumo alimentario.

- Agroalimentaria. 19: 61-74.
- Coleman, D.; Dong, X., Hardin, J., Rocke, D. y Woodruff, D. (1999). Some computational issues in clusters analysis with no a priori metric. Computational Statistics & Data análisis. 31: 1-11.
- Dehollain, P. (1995). Conceptos y Factores Condicionantes de la seguridad alimentaria en hogares. Agroalimentaria. 1.
- FAO. (2004). Manual de Objetivos y Metodología de la Encuesta de Seguridad Alimentaria (ESAH).
- Flores, L., Villa, A. e Iglesias, Luis. (1996). Estratificación multivariante. Criterios de evaluación. Estadística Española. 38 (141): 19-35.
- Hair, A.; Tatham y Black. (1999). *Análisis Multivariante*. Madrid: Prentice Hall.
- Hyman, G., Larrea, C.; Farrow, A. (2005). Methods, results and policy implications of poverty and food security mapping assessments. Food Policy. 30: 453-460.
- Johnson, R.; Wichern, D. (1999). Applied Multivariate statistical Analysis. New Jersey: Prentice Hall.
- Mercado, C.; Lorenzana, P. (2000). Acceso y Disponibilidad Alimentaria Familiar. Validación de instrumentos para su medición. Caracas: Fundación Polar.
- Peña, Daniel. (2002). *Ánalisis de datos multivariantes*. Madrid: Editorial McGraw Hill.
- Pérez, C. (2005). *Métodos Estadísticos Avanzados con SPSS*. Madrid: Editorial Thomson.
- Proyecto UTF/VEN/008/VEN FAO-MAT. Programa Especial de Seguridad Alimentaria (2006). Resultados de la situación de seguridad alimentaria en algunas comunidades rurales apoyadas por el Programa Especial de Seguridad Alimentaria (PESA) 2004-2006.
- Rodriguez, M. (2006). Análisis socioeconómico de la Encuesta de Seguridad Alimentaria en el Hogar. PROYECTO UTF/VEN/008/VEN FAO-MAT. Caracas: Programa Especial de Seguridad Alimentaria.
- Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences*. Londres: University of Cincinnati, Lawrence Erlbaum Associates publishers.
- The SPSS TwoStep Cluster Component. A Scalable component enabling more efficient customer segmentation. SPSS White paper-Technical report, [en línea]. Recuperado el 8 de noviembre de 2011, de <http://www-01.ibm.com/software/analytics/spss/>

