

INVESTIGACION

Una familia de índices para la caracterización global de la dieta

PERSPECTIVAS EN NUTRICIÓN HUMANA
ISSN 0124-4108 Número 11. Julio de 2004
Universidad de Antioquia. Medellín. Colombia págs. 11-23

Pedro A. Monterrey

Dr. en Ciencias Matemáticas (Ph. D).
Consultor Estadístico.
E-mail: pedromonterrey@yahoo.com.mx

Gisela M. Pita

Dra. en Medicina.
Máster en Salud Ambiental y en Nutrición.
Instituto de Nutrición e Higiene de los Alimentos, Cuba.

Resumen

El procedimiento utilizado habitualmente para evaluar los datos de las encuestas dietéticas cuantitativas, es el análisis individual del comportamiento de la ingesta de cada uno de los nutrientes. En general el análisis conjunto o multivariado de los datos se realiza para controlar el efecto de la ingestión calórica total y tropieza con la dificultad de que las ingestiones de los nutrientes

están relacionadas y no se reporta en la literatura cómo proceder, para la aplicación de esas técnicas, teniendo en cuenta las correlaciones de los datos de la dieta. Se han construido índices para caracterizar de manera conjunta los datos de la dieta, éstos índices se han elaborado con el interés de caracterizar la calidad y la variedad de la dieta del individuo. En el presente trabajo se

PALABRAS CLAVE:

Encuestas dietéticas, métodos multivariados, distancias estadísticas, índices para la evaluación cuantitativa de la dieta de grupos.

introduce una familia de distancias estadísticas que permitirán el uso de métodos multivariados para el análisis conjunto de los datos dietéticos teniendo en cuenta las peculiaridades de los mismos. A partir de ellas se construyeron, índices para obtener una caracterización cuantitati-

va global de la dieta de los individuos en el contexto de su grupo poblacional. Estos índices serán de utilidad en los casos en que el interés del análisis de datos se centra en las comparaciones o evoluciones de la dieta de grupos de individuos o de poblaciones.

An Index family for the global diet characterization

Summary

The most frequent technique for quantitative data analysis in dietetics is the univariate analysis, in which the nutrients effect is determined one by one. Generally speaking, multivariate analysis is limited to control the effect of total caloric intake. The use of multivariate statistical data analysis techniques is limited or influenced by dependences between different nutrients intakes. There are not clear evidences in literature about the best way for taking these correlations into account. The clearest way, in literature, for combining food or nutrients intake

is the construction of indexes to quantify diet quality or dietary diversity. In the present article we define a family of statistical distances, which will permit the use of statistical multivariate data analysis techniques in dietetics data analysis taking into account the characteristics of each one. Based on these distances, we constructed indexes to obtain a quantitative characterization of an individual diet relative to the group's diet. Such indexes will be useful in cases when data analysis involves group comparisons or diagnostics of groups or population's diet.

KEY WORDS:

Dietetics surveys, multivariate methods, statistical distances, index for group quantitative diet evaluation.

INTRODUCCIÓN

El procedimiento utilizado habitualmente para evaluar los datos de las encuestas dietéticas cuantitativas es el análisis individual del comportamiento de la ingesta de cada uno de los nutrientes. Este análisis permite detectar los nutrientes dónde se presentan más deficiencias y asociar éstas con la presencia de determinado evento adverso a la salud; pero tiene la limitación de evaluar individualmente, como piezas fraccionadas, componentes de un proceso que es parte de un todo y de esta forma las diferentes interacciones entre los nutrientes pudieran interferir en las conclusiones a que se arribe respecto a la dieta del individuo en su conjunto.

Un complemento de los análisis individuales es el uso de métodos multivariados, dónde se analicen conjuntamente varios nutrientes, para brindar una visión global de la dieta. En general, en la práctica diaria de la epidemiología nutricional, estas técnicas se limitan a procedimientos que controlan la ingesta de los nutrientes ajustándolas en correspondencia con la ingesta calórica total.

Es claro que el uso de métodos estadísticos multivariados se ve dificultado por el hecho de que las ingestas de muchos nutrientes están estrechamente interrelacionadas y dependen, desde muchos puntos de vista, de la ingesta calórica total (1). Estas peculiaridades de los datos

dietéticos limitan el uso de los procedimientos estadísticos multivariados al no existir análisis sobre la efectividad y deficiencias de las distancias estadísticas utilizadas comúnmente en ellos.

En la estadística multivariada las distancias estadísticas permiten introducir una cuantificación en las diferencias entre individuos a partir del análisis del comportamiento simultáneo de varias variables aleatorias. En el campo de la nutrición, esto sería útil para cuantificar las diferencias (o semejanzas) entre dos individuos teniendo en cuenta la ingesta total de diferentes nutrientes, lo que permitirían hacer un análisis global de la dieta involucrando simultáneamente varios nutrientes de interés. De esta forma podrían obtenerse valoraciones globales de la dieta y se podría hacer uso de los métodos estadísticos para la agrupación, lo que potenciaría el análisis de datos en la epidemiología nutricional. Lamentablemente no existen antecedentes conocidos del uso de distancias en las publicaciones revisadas donde se analizan datos de la dieta; su uso se reduce a consideraciones univariadas para detectar los indicadores que más discriminen entre poblaciones en el marco de la construcción de pruebas de tamizaje. En ese contexto se identifican la distancia estandarizada y su adaptación, al caso de varianzas diferentes, cono-

El objetivo del presente trabajo es introducir una familia de distancias estadísticas que se adecuen a las características de los datos de la dieta y construir, a partir de ellas, índices para caracterizar la dieta del individuo...

cida como el índice de detectabilidad (2) (3), que no son más que distancias construidas a partir de la estandarización de un puntaje y que, en su estructura, son semejantes a la distancia de Fréchet (4).

En el análisis de datos de la dieta la combinación de datos se realiza para construir índices que cuantifiquen la calidad de la dieta. Esos índices fueron agrupados por Kant, en 1996, en diferentes categorías: índices para caracterizar nutrientes individuales, índices para agrupar nutrientes, puntajes para caracterizar el consumo de productos de origen animal y vegetal y puntajes para caracterizar patrones de alimentación (5). En este contexto se destacan los índices para evaluar la calidad de la dieta (6) y los índices para cuantificar la diversidad de la dieta (7). Todos estos índices se construyen sobre la base de evaluaciones cuantitativas de la dieta de los individuos y su función es cuantificar aspectos de la dieta que permitan obtener caracterizaciones de los mismos.

El objetivo del presente trabajo es introducir una familia de distancias estadísticas que se adecuen a las características de los datos de la dieta y construir, a partir de ellas, índices para caracterizar la dieta del individuo en el contexto de un grupo. Estos índices, según la clasificación de Kant, son índices para agrupar nutrientes, pero se diferencian de los mencionados anterior-

mente en que el interés que se persigue con ellos no es caracterizar algún atributo de la dieta del individuo sino representar la dieta del individuo en el contexto del grupo. Las distancias que se introducen constituirán un buen instrumento para extender el uso de los métodos de la estadística multivariada al campo de la nutrición y los índices permitirán una visión global de la dieta del individuo, relativa al grupo, lo que será un instrumento importante en los estudios epidemiológicos donde se deba comparar la dieta de varios grupos: estudios de casos y controles, cohortes o en evaluaciones de impacto de acciones.

Los índices construidos fueron aplicados, a manera de ejemplo, en el análisis de la variación estacional de la dieta de embarazadas en Cuba.

MATERIALES Y MÉTODOS

Para la combinación de datos de la ingesta de diferentes nutrientes se observó la necesidad de utilizar datos estandarizados, que es el criterio de construcción del puntaje Z. De esta forma se dispone de valores comparables, al ser independientes de la escala de medición, y por ello numéricamente combinables. En correspondencia la familia de distancias que se propone tiene la siguiente estructura:

$$D_{jk} = \sum_i c_i \frac{|n_{ij} - n_{ik}|}{a_i}$$

Donde:

D_{jk} : denota la distancia (o diferencias en la ingestión) entre el individuo j y el k.

\sum_i : representa la suma de la expresión que se escribe a continuación del símbolo, expresión que depende de cada uno de los nutrientes evaluados los que se simbolizan individualmente con la letra i.

n_{ij} : representa la ingesta del nutriente i por parte del individuo j.

n_{ik} : representa la ingestión del nutriente i por parte del individuo k.

$|\cdot|$: representa el módulo o valor absoluto de un número, es decir, su valor sin signo. El valor absoluto de 5 ($|5|$) es el mismo que el de -5 ($|-5|$), es decir 5.

c_i y a_i son coeficientes que al tomar diferentes valores determinan las diferentes distancias de la familia que se define.

Cuando $c_i = 1$ y a_i es igual a la desviación estándar del nutriente i (s_i), la distancia D recibe el nombre de *distancia de Fréchet*, muy conocida entre aquellos que están familiarizados con los métodos estadísticos. En general los datos de la ingesta de nutrientes son asimétricos, eso hace que sus coeficientes de variación, y por ello sus varianzas, sean elevados. Como al dividir por el coeficiente a_i en cada sumando de la definición de la distancia se hace un cambio de escala (recordar la construcción y significado del puntaje Z), es aconsejable, en algunos casos, utilizar en el cociente una medida de dispersión menos sensible que la desviación estándar frente a la presencia de valores extremos de las variables. En correspondencia se propone considerar, como valor para a_i , el recorrido intercuartílico

(RI) que es otra medida de dispersión, con el atributo de ser menos sensible que la desviación estándar frente a la presencia de valores muy extremos: la distancia así construida se denominará *distancia de Fréchet modificada*.

La distancia de Fréchet, o la de Fréchet modificada, tienen el inconveniente de que, en su construcción, todos los nutrientes tienen la misma importancia o el mismo aporte. Eso hace que, en la suma, los nutrientes que están correlacionados tengan un efecto, sobre las diferencias entre individuos, mayor que lo deseado y magnifiquen diferencias que no son tales o tan marcadas en la dieta general del individuo. Una solución para compensar ese efecto es multiplicar el sumando correspondiente a cada nutriente, en la definición de D, por una constante

(c_i) que tenga en cuenta la dependencia entre los valores de la ingesta de cada uno de los nutrientes considerados.

El coeficiente de determinación (R^2) caracteriza el porcentaje de la variación de una variable que es explicada por otras y por ello $1-R^2$ es una medida de la información de una variable que *no es explicada* por las otras. Por esa razón se propone considerar $c_i = 1-R^2$, siendo R^2 el coeficiente de determinación construido de la regresión entre el nutriente i y el resto de los nutrientes. En ese caso la distancia de Fréchet la nueva distancia construida ($a_i = s_i, c_i = 1 - R^2$) se denominará *distancia R de Fréchet* y en el caso de la distancia de Fréchet modificada se denominará *distancia R de Fréchet modificada* ($a_i = R^2, c_i = 1 - R^2$).

A partir de las distancias se construyeron índices para caracterizar globalmente la dieta del individuo en el contexto del grupo. La idea es simple: se construye un individuo "ficticio" con los puntajes mínimos en cada uno de los nutrientes analizados, y el valor del índice para un individuo cualquiera es la distancia entre él y el individuo "ficticio" que representaría la ingesta "más baja" en el grupo. Aplicando este proceder a la distancia de Fréchet y a la distancia R de Fréchet se construyeron, respectivamente, dos índices: el índice de Fréchet (IF) y el índice R (IR) para describir el comportamiento global de la dieta de un individuo en el contexto del grupo.

Los índices introducidos fueron aplicados al análisis de las variaciones estacionales de la dieta de las embarazadas de un estudio transversal, en el que se seleccionaron 203 casos de nacimientos fisiológicos ocurridos desde febrero 2000 hasta enero 2001 en el Hospital Ginecobstétrico "América Arias", ubicado en la Ciudad de la Habana Cuba, para estudiar la relación madre/hijo de las vitaminas antioxidantes. En Cuba la disponibilidad de algunos alimentos es fuertemente estacional y, en correspondencia, estos casos fueron distribuidos en los cuatro trimestres en los que se divide el año según los periodos de estacionalidad en la disponibilidad de alimentos (1^{er} trimestre de febrero-abril, 2^{do} trimestre de mayo-julio, 3^{er} trimestre de agosto-octubre, 4^{to} trimestre de noviembre-enero). Los datos de la dieta fueron obtenidos mediante una encuesta de frecuencia semicuantitativa de consumo de alimentos referida a los tres meses anteriores al parto. Los detalles de este estudio, las características del muestreo, las diferentes determinaciones realizadas y sus resultados principales se encuentran en el trabajo de Pita y colaboradores (8).

Para la evaluación del comportamiento de los diferentes nutrientes y los índices construidos, en cada uno de los cuatro trimestres en que se consideró dividido en año, se utilizó el análisis de varianza de clasificación simple y la prueba de comparación múltiple de Bonferroni.

Para la significación de las diferencias se utilizó un nivel de significación del 5%. En los análisis los datos fueron transformados logarítmicamente para homogeneizar las varianzas.

RESULTADOS

El comportamiento de los nutrientes a considerar en el análisis actual se describe en la Tabla 1.

La Figura 1 presenta el comportamiento de algunos de ellos en los diferentes semestres y los valores p obtenidos como resultado de la

aplicación de la prueba de análisis de varianza para evaluar la significación de las diferencias entre trimestres. Todos los nutrientes, a excepción de las proteínas, los lípidos y la vitamina A, tienen un comportamiento diferenciado entre semestres. La mayoría reprodujo el comportamiento de la energía. Los trimestres 1, 3 y 4 tienen un comportamiento bastante semejante y en el segundo el consumo fue significativamente más bajo. Sin embargo los carotenos, la tiamina y el ácido fólico también presentaron valores bajos, y en ocasiones más bajos, en el trimestre 3.

TABLA 1

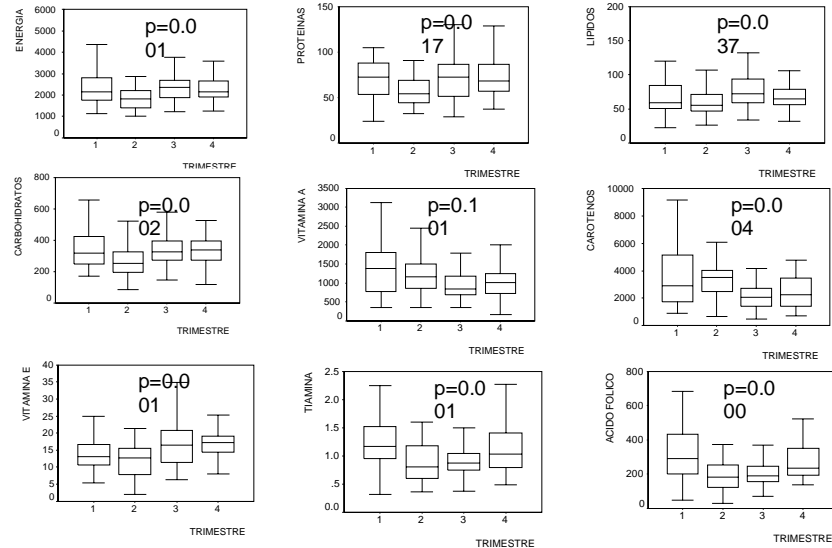
Descripción del comportamiento de los nutrientes evaluados (n = 203)

| Nutriente | Mínimo | Máximo | Media | D.S* |
|-----------------------|--------|---------|--------|--------|
| Energía (kcal) | 1009.0 | 5334.7 | 2243.5 | 296.8 |
| Proteínas Totales (g) | 24.0 | 186.59 | 69.3 | 22.9 |
| Lípidos (g) | 22.6 | 212.5 | 71.0 | 30.3 |
| Carbohidratos (g) | 83.7 | 747.5 | 326.0 | 114.6 |
| Vitamina A | 169.3 | 5553.3 | 1204.7 | 838.6 |
| Caroteno | 485.5 | 14990.6 | 2939.6 | 2241.1 |
| Vitamina E | 1.92 | 69.3 | 16.3 | 9.4 |
| Tiamina | 0.32 | 2.72 | 1.0 | 0.4 |
| Niacina | 4.2 | 48.2 | 12.6 | 5.1 |
| Rivoflavina | 0.4 | 4.08 | 1.7 | 0.8 |
| Piridoxina | 0.5 | 3.8 | 1.8 | 0.6 |
| Ácido Fólico | 30.8 | 716.9 | 250.5 | 126.2 |
| Vitamina C (mg) | 11.4 | 872.7 | 192.8 | 160.9 |
| Cobalamina (µg) | 0.8 | 24.8 | 6.1 | 4.2 |

*D. S: Desviación estándar

FIGURA 1

Comportamiento de algunos nutrientes en cada uno de los trimestres evaluados



Este análisis muestra diferencias entre trimestres, pero ¿es posible decir que en el trimestre 2 el consumo fue significativamente menor que en el resto, incluido el 3? Para responder a esta pregunta y ampliar el análisis anterior se utilizaron los índices de Fréchet y R.

A partir de los valores descriptivos del comportamiento de cada nutriente presentados en la Tabla 1, se calculó el IF de los valores del consumo de todas las embarazadas. A manera de ejemplo se presenta el cálculo de IF para un individuo hipotético cuyo consumo es: energía

1850 Kcal, proteínas totales 90 g, lípidos 100 g, carbohidratos 500 g,.....vitamina C 300 mg, cobalamina 3 µg)*

El resultado de efectuar las operaciones indicadas es 35.3. Este valor no tiene un significado por sí mismo, es un número relativo que indica la ubicación del individuo en el grupo. Mientras mayor sea el índice mayor el consumo del individuo. La Figura 2 describe el comportamiento de los valores de IF en cada uno de los semestres; la aplicación del análisis de varianza de clasificación simple mostró diferen-

$$* \text{ IF} = \frac{1850 - 1009}{396.8} + \frac{90 - 24.0}{22.9} + \frac{100 - 22.6}{30.3} + \frac{500 - 83.7}{114.6} + \dots + \frac{300 - 11.4}{160.9} + \frac{3 - 0.8}{4.2}$$

cias significativas entre semestres ($p = 0.009$). Al observar la Figura 2 se aprecia que en el segundo trimestre se obtuvieron valores del índice de consumo de Fréchet mucho más bajo que en los restantes. La aplicación de la prueba de comparación múltiple de Bonferroni permitió aceptar diferencias significativas entre los trimestres 1 y 2 ($p = 0.015$) y los trimestres 2 y 4 ($p = 0.024$). El trimestre 3, aunque tuvo valores ligeramente inferiores a los de los trimestres 1 y 4, estos no llegaron a determinar diferencias significativas ($p = 1.000$ y $p = 0.617$, respectivamente). La afirmación de que tuvo valores intermedios se sustenta en que tampoco presentó diferencias con el trimestre 2 ($p = 0.077$). Es innegable que los valores de consumo del trimestre 2 fueron sustancialmente inferiores al

de los otros trimestres. De esta forma se aisló un patrón de comportamiento anual de la dieta de las embarazadas: en el segundo trimestre la ingesta dietética es inferior al resto del año.

¿Hasta qué punto las interacciones entre los nutrientes pudieron influir en las conclusiones anteriores? El uso de IR aclaró esta duda. La Tabla 2 presenta los coeficientes de determinación de las regresiones entre las diferentes variables.

Por el alto grado de interdependencia entre la ingesta de los nutrientes y la de energía, los coeficientes de determinación entre los nutrientes fueron calculados sin la energía para obtener los niveles reales de dependencia entre ellos y dilucidar así su real aporte a la distancia. La última columna de la Tabla 2 pre-

FIGURA 2
Comportamiento del Índice de Fréchet en cada uno de los trimestres

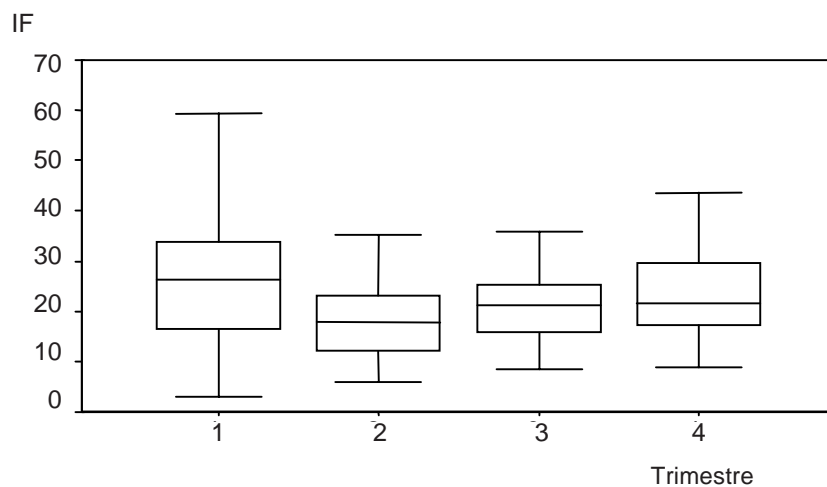


TABLA 2
Coefficientes de determinación

| Nutriente | Coefficientes de determinación | |
|-------------------|--------------------------------|------------------|
| | (R ²) | 1-R ² |
| Energía | 0.999 | 0.001 |
| Proteínas totales | 0.973 | 0.027 |
| Lípidos | 0.995 | 0.005 |
| Carbohidratos | 0.998 | 0.002 |
| Vitamina A | 0.963 | 0.037 |
| Caroteno | 0.872 | 0.128 |
| Vitamina E | 0.837 | 0.163 |
| Tiamina | 0.898 | 0.102 |
| Niacina | 0.911 | 0.089 |
| Rivoflavina | 0.957 | 0.043 |
| Piridoxina | 0.859 | 0.141 |
| Ácido Fólico | 0.894 | 0.106 |
| Vitamina C | 0.862 | 0.138 |
| Cobalamina | 0.964 | 0.036 |

senta los factores por los que se multiplica el valor estandarizado de cada nutriente en el cálculo de IR. Como se aprecia la energía es una de las variables que menos aporta al índice pues sus valores son explicados, en un 99.9%, por los valores de los otros nutrientes, dicho con otras palabras los valores observados de la energía son consecuencia de los valores de los otros nutrientes, lo que se tiene en cuenta a la hora de obtener el valor numérico del índice. Utilizando de nue-

vo los datos hipotéticos mencionados en el ejemplo de cálculo de IF, el IR en ese caso sería*

El resultado de los cálculos es 2.7. La interpretación y utilidad de este número es la misma que se indicó para IF.

La Figura 3 describe el comportamiento de la ingesta de las embarazadas, en cada uno de los trimestres del estudio, consolidado mediante los valores de IR. De nuevo se observó una diferencia sustan-

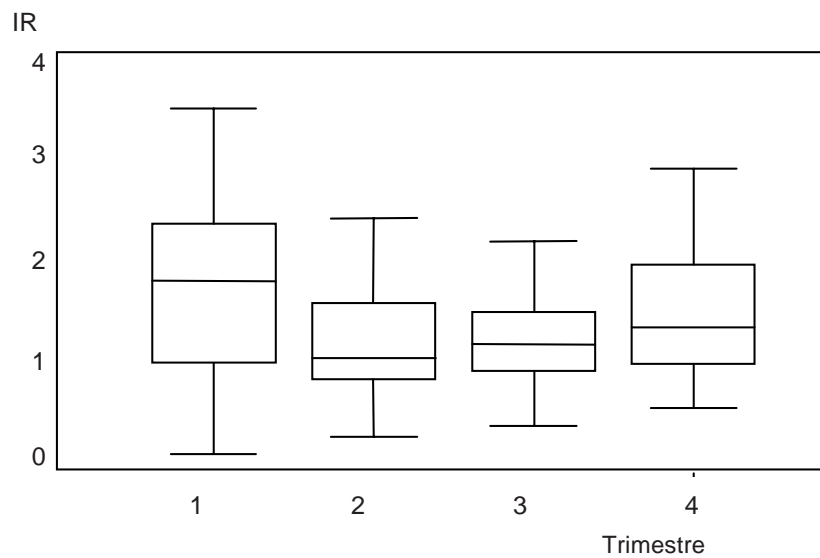
$$* \text{ IR} = (1 - 0.999) \frac{1850 - 1009}{396.8} + (1 - 0.973) \frac{90 - 24.0}{22.9} + \dots + (1 - 0.964) \frac{3 - 0.8}{4.2}$$

cial en los valores del segundo trimestre, pero la significación estadística de las diferencias detectadas mediante el análisis de varianza fue mayor que en el caso de IF ($p = 0.004$). La prueba de comparación múltiple de Bonferroni, sin embargo, reafirmó las conclusiones a

que se arribó con IF. Los valores del trimestre 2 se diferencian del 1 ($p = 0.009$) y del 4 ($p = 0.047$), no así del 3 ($p = 1.000$). El trimestre 3 no se diferenció significativamente en su comportamiento del 1 ($p = 0.066$) ni del 4 ($p = 0.335$).

FIGURA 3

Comportamiento del Índice R en cada uno de los trimestres



DISCUSIÓN

Las distancias estadísticas introducidas fueron traducidas a índices para evaluar globalmente la dieta de un individuo. En el análisis de la dieta de las embarazadas y con el interés de decidir si había o no un efecto estacional que, de manera significativa, pudiera estar determinando que efectivamente la cuantía total de la ingesta dietética cambia según el momento del año se

tropezó, cuando se hizo el análisis individual de los nutrientes, con que en algunos casos las diferencias no eran significativas, aunque se observaba una marcada tendencia a la disminución en el trimestre 2 y no tan marcada en el 3. Al consolidar todos los datos en un índice quedó marcada la diferencia del trimestre 2, respecto a los otros, y una disminución no tan marcada en el 3. Al

Los índices definidos a partir de ellas no pretenden evaluar al individuo sino caracterizar el comportamiento de la dieta de un grupo por lo que serán de utilidad en los casos en que el interés del análisis de datos se centre en las comparaciones o evoluciones de la dieta de grupos de individuos o de poblaciones.

despojar este análisis del efecto de interacción entre nutrientes la conclusión del efecto del trimestre 2 quedó más resaltada.

El resultado obtenido refleja una característica de la alimentación de la población de la Ciudad de la Habana ya que se corresponde a la confluencia de varios factores:

1. Las vacaciones escolares que en Cuba abarcan los meses de junio-julio-agosto.
2. Las vacaciones de los trabajadores por el verano y, especialmente, de las madres trabajadoras porque la escuela recesa (finales de julio y agosto).
3. Disminución de la disponibilidad de alimentos por la etapa de tránsito entre las cosechas de primavera (abril-mayo) y la de fin de año.

Estos factores determinan que los trimestres 2 y 3, pero especialmente el 2, sean un momento de máxi-

ma demanda de alimentos en el hogar, pues escolares y trabajadores durante el resto del año almuerzan en sus centros de trabajo o estudio, y simultáneamente sean un momento de disminución en la disponibilidad de los alimentos a que se puede acceder por los grupos familiares.

Las distancias introducidas no tienen antecedentes en el análisis de datos de la dieta, y constituirán un punto inicial para propiciar un uso más profundo y extendido de diferentes técnicas del análisis de datos multivariados en el análisis de datos de la dieta, al introducir en su formulación las interacciones que existen entre estos datos. Los índices definidos a partir de ellas no pretenden evaluar al individuo sino caracterizar el comportamiento de la dieta de un grupo por lo que serán de utilidad en los casos en que el interés del análisis de datos se centre en las comparaciones o evoluciones de la dieta de grupos de individuos o de poblaciones.

Referencias

1. Willet W. Nutritional Epidemiology. New York: Oxford University; 1990.
 2. Stoltzfus RJ, Habicht JP, Rasmussen KM, Hakimi M. Evaluation of indicators for use in vitamin A intervention trials targeted at women. Int J Epidemiol 1993; 22 (6): 1111-1118.
- Habicht JP, Meyers LD, Brownie C. Indicators for identifying and counting the improperly nourished. Am J of Clin Nutr 1982; 35 (5): 1241-1254.
3. Linares Girela D. Análisis de datos. La Habana: Universidad de la Habana, Facultad de Matemática Cibernética; 1990.

4. Kant AK. Indexes of overall diet quality: a review. *J Am Diet Assoc.* 1996; 96 (8): 785-791.
 5. Haines PS, Siega-Riz AM, Popkin BM. The diet quality index revised: a measurement instrument for populations. *J Am Diet Assoc* 1999; 99: 679-704.
- Ruel MT. Operationalizing dietary diversity: a review of measurement issues and research priorities. *J Nutr.* 2003 Nov;133(11 Supl 2): 3911S-3926S.
6. Pita G, Pineda D, Martín I, Monterrey P, Serrano G, Macías C. Ingesta de macronutrientes y vitaminas de un grupo de embarazadas a través de un año de estudio. *Revista Cubana de Salud Pública* 2003; 29(3): 220-27.
 7. Gibson RS. *Principles of Nutritional Assessment.* New York: Oxford University; 1990.

FECHA DE RECEPCIÓN: 11 de diciembre 2003

FECHA DE ACEPTACIÓN: 14 de junio 2004