

# UTILIZACIÓN DEL ANÁLISIS DE CLUSTER CON VARIABLES MIXTAS EN LA SELECCIÓN DE GENOTIPOS DE MAÍZ (*Zea mays*)

Dunia Chávez Esponda\*, Ileana Miranda Cabrera\*\*, Mario Varela Nualles\*\*\*, Lianne Fernández\*\*\*\*

\*Universidad Agraria de La Habana (UNAH), Cuba.

\*\*Centro Nacional de Sanidad Agropecuaria (CENSA), Cuba.

\*\*\*Instituto Nacional de Ciencias Agrícolas (INCA), Cuba.

\*\*\*\*Instituto de Investigaciones Fundamentales en Agricultura Tropical “Alejandro de Humboldt” (INIFAT), Cuba

## ABSTRACT

The cluster analysis or numeric taxonomy is the method that is used to establish individuals' groups, unifying inside oneself group to those elements that have characteristic similar. In the classification methods it plays an important list the index of similarity or it distances used then of it depends it in great measure that the final results have the possible biggest dependability. This index or coefficient is selected according to the variables measures. When there is presence of mixed (quantitative, qualitative and binary) variables he/she intends to use the coefficient of Gower, however in the agricultural sphere their use is very scarce. In this work the Analysis of Cluster was used, calculating the distance womb among individuals starting from the coefficient Gower's similarity, with the objective of to study the agronomic behavior of the experimental varieties of corn (*Zea mays*) and to select genotypes with more resistance potentialities to plagues and that they have bigger growth to use it like base in the genetic cruzamiento. This method allowed to contain and to characterize the studied varieties, that which is of great importance for the investigators to decide if to discard a variety or to follow it cultivating and more important still, to use the genotypes like transgenic plants, aspect that is having excellent results in our country.

**KEY WORDS:** cluster analysis, similarity of Gower, genotypes of corn.

MSC 62P10

## RESUMEN

El análisis de cluster o taxonomía numérica es el método que se utiliza para establecer grupos de individuos, unificando dentro de un mismo grupo a aquellos elementos que tengan características similares. En los métodos de clasificación juega un rol importante el índice de similitud o distancia utilizada pues de ello depende en gran medida que los resultados finales tengan la mayor confiabilidad posible. Este índice o coeficiente se selecciona de acuerdo a las variables medidas. Cuando hay presencia de variables mixtas (cuantitativas, cualitativas y binarias) se propone usar el coeficiente de Gower, sin embargo en la esfera agropecuaria su uso es muy escaso. En este trabajo se empleó el Análisis de Cluster, calculando la matriz de distancia entre individuos a partir del coeficiente de similitud de Gower, con el objetivo de estudiar el comportamiento agronómico de las variedades experimentales de maíz (*Zea mays*) y seleccionar genotipos con más potencialidades de resistencia a plagas y que tengan mayor crecimiento para utilizarlo como base en el cruzamiento genético. Este método permitió agrupar y caracterizar las variedades estudiadas, lo cual es de gran importancia para los investigadores decidir si descartar una variedad o seguirla cultivando y más importante aún, utilizar los genotipos como plantas transgénicas, aspecto que está teniendo excelentes resultados en nuestro país.

## 1. INTRODUCCIÓN

La especie *Zea mays*, de origen americano, se dispersó desde su centro de origen en Mesoamérica a todo el Continente Americano (MANRIQUE CHÁVEZ, 1997) y a Europa (BOSH et al., 1997). Este cereal es considerado como el tercer cultivo en importancia en el mundo después del trigo y arroz (EYZAGUIRRE Y LINARES, 2004). Se utiliza en la alimentación humana y animal aunque también tiene uso industrial. En México, es el cultivo de mayor importancia económica y social, debido a que forma parte de la dieta de la mayoría de los mexicanos, principalmente, los de escasos recursos que viven en áreas marginales. Por otra parte, en la sierra peruana, los maíces de altura constituyen un recurso importante en la alimentación diaria de las familias rurales (MANRIQUE CHÁVEZ, 1997).

El cultivo de maíz en Cuba ha constituido un elemento básico en la alimentación de la población humana, del ganado y de las aves. Las condiciones climáticas para el cultivo del maíz en Cuba no son óptimas en comparación con otras regiones de América, sin embargo, la diversidad de usos de este cereal en el país justifica más su cultivo que los

rendimientos que se obtienen (FERNÁNDEZ et al., 2004). Por esta razón, el maíz es uno de los cultivos económicos en los que se ha investigado y se continúa avanzando en la mejora genética.

La variabilidad genética existente en Cuba constituye una riqueza potencial que puede ser de gran valor estratégico para el país (FERNÁNDEZ et al., 2006). HATHEWAY (1957) determinó la existencia en Cuba de siete razas de maíz: Criollo, Tusón, Canilla, Argentino, *White Pop*, *Yellow Pop*, y *White dent*, reconocidas por sus características distintivas en la planta, la mazorca y el grano. El autor señala, sobre la base de evidencias arqueológicas e históricas que las tres primeras deben haberse originado a partir de razas introducidas en Cuba antes de la conquista de América y las cuatro restantes deben haberse introducido desde otras regiones de América después de esa fecha; la raza Argentino desde Argentina, *White dent* desde México, *White Pop* y *Yellow Pop* desde los Estados Unidos de América, pero luego de posteriores cruzamientos y el aislamiento geográfico de Cuba se formaron razas genuinamente cubanas. Una vez introducidas las variedades cubanas se han realizado estudios de caracterización con el objetivo de elegir aquellas que resulten más resistentes a enfermedades y al mismo tiempo mantengan caracteres agronómicos favorables con rendimientos adecuados, por lo que se ha trabajado en la identificación de variedades a partir de un conjunto de descriptores medidos.

El análisis de cluster es el método que se utiliza para establecer grupos de individuos, unificando dentro de un mismo grupo a aquellos elementos que tengan características similares (LINARES, 1990). En los métodos de clasificación juega un rol importante el índice de similitud o distancia utilizada pues de ello depende en gran medida que los resultados finales tengan la mayor confiabilidad posible (MIRANDA, 1997). Este índice o coeficiente se selecciona de acuerdo a las variables medidas.

Existen procesos biológicos en los que es necesario medir parámetros de naturaleza mixta. Esto produce una matriz de datos con variables cuantitativas, cualitativas y binarias para la cual no se ha encontrado una medida de similitud adecuada. Aunque algunos autores (MIRANDA Y TORRES, 1998; GALLINARI Y MARIN, 2001) han tratado de implementar distancia que permitan un tratamiento para unificar la taxonomía convencional con la molecular, hasta hoy la estadística no ha ofrecido una solución algorítmica para obtener los resultados más confiables y los biólogos y taxónomos desconocen las ventajas y desventajas de una u otra métrica, realizan agrupamientos seleccionando métricas y métodos sin tener conocimiento estadístico previos a la selección. En la literatura se reporta que en presencia de variables mixtas (cuantitativas, cualitativas y binarias) se propone usar el coeficiente de Gower (GOWER, 1971), sin embargo en la esfera agropecuaria su uso es nulo.

En este trabajo se empleó el Análisis de Cluster, calculando la matriz de distancia entre individuos a partir del coeficiente de similaridad de Gower (GOWER, 1971), con el objetivo de estudiar el comportamiento agronómico de las variedades experimentales de maíz (*Zea mays*) con vistas a seleccionar genotipos que posean más potencialidades de resistencia a plagas y que tengan mayor crecimiento para utilizarlo como base en el cruzamiento genético (GÓMEZ, 2006).

## 2. DESARROLLO

El Análisis de Cluster posee un algoritmo matemático en el que se establecen muchas interrogantes que aun no han tenido una respuesta incorporada a las investigaciones taxonómicas actuales (MIRANDA 1997). De echo, este método parte del cálculo de una matriz de semejanza entre los individuos (EVERITT, 1980) que debe ser obtenida a partir de una métrica  $d(X_i, X_j)$  que refiere la similitud, disimilaridad o distancia entre el individuo  $i$  y el  $j$ , dicha métrica se selecciona de acuerdo a las escalas de medición de las variables, así como de las estructuras de correlación en ellas existentes (LINARES 1990), pero aun se desconoce cual de las métricas ofrece la agrupación jerárquica más estable, más cercana al agrupamiento real entre individuos.

Si todas las variables (parámetros evaluados) son continuas existe un conjunto de distancias para calcular la similitud, siendo la más conocida la distancia Euclideana. Sin embargo esta métrica no es recomendable cuando las variables están dadas en escalas diferentes (VARELA, 1996). Aunque a veces se sugiere estandarizar las variables antes de continuar con el análisis, de manera general, no se conoce un criterio para la selección de la medida de similitud preciso y los patrones de agrupamiento obtenido empleando una u otra métrica pueden variar aun cuando se utilice el mismo método de aglomeración (MIRANDA, 1997).

Si todas las variables son binarias hay un conjunto de similitudes que precisa el grado de semejanza entre dos individuos. Una de las más usadas es la métrica de Jaccard que precisa un porcentaje calculado como  $\#$  de variables en que dos individuos coinciden /  $\#$  de variables analizadas. Tampoco, se ha precisado un criterio de selección, pero en

este caso las fórmulas de las métricas ofrecen una interpretación más rápida haciendo algo más sencilla la selección. (SOTO et al, 2006).

Algunos investigadores han trabajado con medidas de similitud para atributos binarios, especialmente en problemas de la biología molecular; 18 de estas métricas aparecen en el manual del paquete estadístico RadDistance y en la opción para el cálculo de distancias que ofrece el SPSS. (SPSS, 1993).

En caso de tener variables discretas algunos autores (VARELA, 1996; EVERITT, 1980) recomiendan construir variables binarias, de manera que si una variable tiene k niveles tendría k nuevas variables de presencia – ausencia.

El problema mayor es cuando las variables son algunas cuantitativas y otras cualitativas (mixtas), en tal caso algunos autores sugieren llevar todas las variables a la misma escala (binaria) (GORDON, 1990), otros sugieren combinar las similitudes empleando medida de ponderación (WOJCIECHOWSKI, 1987; MIRANDA, 1998) y otros plantean realizar un análisis con las variables de igual naturaleza y emplear las otras para interpretar los resultados (SOTO et al, 2006).

De manera general, se desconoce cual de los procedimientos ofrece mejores resultados (SOTO et al, 2006), no hay una técnica de selección de la similitud y no se conocen los convenientes o inconvenientes de la utilización de uno u otro criterio ni las características de las variables para las cuales se adecuan.

El coeficiente de similitud de Gower (GOWER, 1971) es una medida de similitud que permite la utilización simultánea de variables cuantitativas, cualitativas y dicotómicas. Aplicando este coeficiente de similitud se puede determinar el grado de semejanza o similitud entre individuos; a los cuales se le han medido características cualitativas, cuantitativas (continuas y discretas) y binarias. El mismo se define como sigue:

$$d_{ij}^2 = 1 - s_{ij}$$

$$s_{ij} = \frac{\sum_{h=1}^{p1} \left( 1 - \frac{|x_{ih} - x_{jh}|}{G_h} \right) + a + \alpha}{p1 + (p2 - d) + p3}$$

p1: número de variables cuantitativas,

a: número de coincidencias en 1

d: número de coincidencias en 0 de las p2 variables binarias

$\alpha$ : es el número de coincidencias para las p3 variables cualitativas.

$G_h$  es el rango de la h-ésima variable cuantitativa.

Rango= X máximo – X mínimo.

Cuando todos los caracteres son binarios, el índice de similitud de Gower coincide con el índice de Jaccard. Cuando todos los caracteres son cualitativos (con más de dos estados), el índice es equivalente al coeficiente de coincidencias simple, definido como la relación del número total de coincidencias y el número total de caracteres. Cuando todos los caracteres son cuantitativos, el índice de Gower se asemeja a la medida absoluta de las distancias. (GOWER, 1971).

Autores como (LONDOÑO et al, 2007) señalan algunas de las propiedades del coeficiente de similitud de Gower (GOWER, 1971) que resultan especialmente ventajosas para la taxonomía de las especies. Las mismas son las siguientes:

- Las características morfológicas y moleculares que se miden en las especies involucran variables de diferente naturaleza (cuantitativa, cualitativa, binaria); el coeficiente de similitud de Gower (GOWER, 1971) es apto para calcular similitudes cuando se tienen tales mezclas de variables. (LONDOÑO et al, 2007).
- Usando el coeficiente de similitud de Gower (GOWER, 1971), es posible trabajar con bases de datos en las que faltan observaciones de algunas variables, sin prescindir de todo el vector que representa a la unidad muestral ni usar ningún método de imputación. Esta propiedad resulta muy útil en estudios taxonómicos pues a menudo aparecen observaciones faltantes. (LONDOÑO et al, 2007)

- Mediante el uso de este coeficiente es posible ponderar las variables de manera diferencial, dependiendo del papel que se quiera que cada una juegue en la ordenación. En este sentido, es posible asignar las ponderaciones dando mayor peso a las variables que en estudios precedentes han mostrado alta capacidad discriminante. (LONDOÑO et al, 2007).

En el problema práctico, el experimento se realizó en 55 campos de maíz situados 19 en el occidente y 36 en el oriente del país. En cada campo se muestrearon mensualmente 30 mazorcas dispuestas en un diseño de bloques parcialmente aleatorizado con 20 parcelas de 20m de longitud y una distancia de siembra de 0,90 x 0,35 m. Cada parcela constituyó una variedad. Se observaron los valores promedios de los descriptores agronómicos y morfológicos de los cuales hay 9 cuantitativos, 2 cualitativos y 1 binario:

LM: Longitud de la mazorca (cm).  
 DM: Diámetro de la mazorca (cm).  
 NHG: Número de hileras de granos.  
 NGH: Número de granos por hilera.  
 DT: Diámetro de la tusa (cm).  
 LG: Longitud del grano (cm).  
 AG: Ancho del grano (cm).  
 GrG: Grosor del grano (cm).  
 P100S: Peso de 100 granos.

FG: Forma del grano.  
 CT: Color de la tusa.

DHG: Disposición de hileras de grano.

**Tabla 1a.** Características (promedio) de las variedades de maíz analizadas en las regiones occidental y oriental de Cuba.

Variedad	LM	DM	NHG	NGH	DT	LG	AG	GrG	P100S
Argentino	18,98	4,545	12,64	42,14	2,355	1,205	0,89	0,38	33,165
Canilla	15,51	3,95	12,26	32,87	2,005	1,165	0,84	0,39	27,83
Canilla/ Trad	18,98	4,07	13,76	44,76	1,79	1,24	0,72	0,38	24
Criollo	16,09	4,542	13,61	32,356	2,855	1,0707	0,865	0,3942	29,072
Cuña	17,61	4,137	13,33	39,924	2,021	1,214	0,804	0,372	28,151
En. Mex.	14,79	3,79	12,29	29,57	2,6	0,79	0,83	0,46	15,5
Gíbara	19,77	4,317	12,16	39,7	2,39	1,1567	0,9267	0,39	33,386
Grande	19,22	4,53	13,28	41,36	2,36	1,13	0,86	0,39	33
Grano Ancho	17,8	4,56	12,96	35,32	2,5	1,22	0,97	0,38	34,33
Mexicano	17,98	4,617	14,32	39,443	2,65	1,13	0,85	0,4033	29,943
Morado	17,16	4,322	12,87	37,264	2,386	1,164	0,874	0,378	29,034
Pinto	15,83	4,44	12,96	35,96	2,42	1,15	0,845	0,37	32,915
Pollo	14,93	3,495	13,25	34,415	1,55	1,105	0,705	0,335	20,83
Rojo	15,53	4,35	12,9	35	2,38	1,16	0,84	0,41	30
SN	17,43	3,9	12,64	37,77	2,05	1,1	0,7	0,35	25,93
Tusa Fina	18,16	4,22	13,2	39,6	2,03	1,28	0,82	0,34	31,5
Tusa gruesa	17,1	4,54	13,28	35,74	2,455	1,275	0,885	0,385	31,585
Tusa Morada	17,12	3,98	12,48	39,8	2,02	1,07	0,87	0,4	24,5
Tuson	20,26	4,78	13,2	45,4	2,65	1,23	0,93	0,39	30,67
Yanelys	15,84	4,69	11,44	34,16	2,56	1,23	1,02	0,39	37,5

Se confeccionó una base de datos en Excel que contenía 20 filas (variedades) y 12 columnas (descriptores). Los valores promedio de los descriptores cuantitativos y los valores de los diferentes estados de los descriptores para cada carácter cualitativo utilizado en la caracterización de las variedades se muestran en las tablas 1a y 1b. Como puede

observarse para los caracteres cuantitativos (Tabla 1a) existe una diversidad considerable entre una y otra variedad, más aún entre las frecuencias relativas de los estados de los descriptores cualitativos (Tabla 1b) utilizados.

En la Tabla 1b se observa que las características más frecuentes encontradas en las variedades estudiadas fueron: la forma del grano (contraído), el color de la tusa (blanco) y una disposición regular de hileras de grano.

**Tabla 1b.** Frecuencias relativas de las características cualitativas de las variedades de maíz analizadas en las regiones occidental y oriental de Cuba.

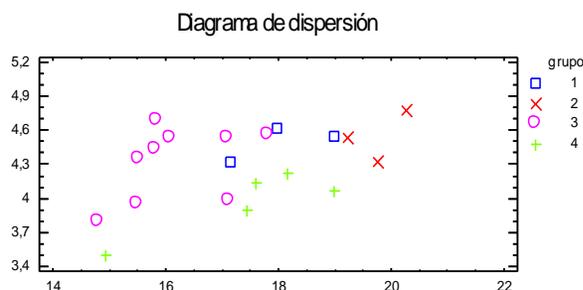
Descriptores	Por ciento (%)
<b>FG</b>	<i>Contraído 80 % Redondo 15% Dentado 5%</i>
<b>CT</b>	<i>Blanco 80% Morado 5% Jaspeado 15%</i>
<b>DHG</b>	<i>Regular 85% Irregular 15%</i>

Se calculó la matriz de similaridad entre los individuos a partir del coeficiente de Gower (S) (GOWER, 1971). Posteriormente se calculó la matriz de distancia como ( $D^2=1-S$ ).

Tomando como entrada la matriz D, se realizó un Análisis de Cluster empleando el método de aglomeración de Ward (WARD, 1963) utilizando el software estadístico STATISTICA Versión 6.0.

En el diagrama de dispersión (figura 1) se puede observar la formación de 4 grupos de variedades de maíz:

**Figura 1.** Diagrama de dispersión.



Al realizar el Dendograma (figura 2) se muestra cuales son las variedades que conforman los 4 grupos, teniendo en cuenta que poseen características similares de los descriptores estudiados:

Grupo 1: Variedades: Argentino, Morado y Mexicano.

Grupo 2: Variedades: Gíbara, Grande y Tuson.

Grupo 3: Variedades: Canilla, Rojo, Criollo, Grano Ancho, Tusa Guesa, Yanelys, Pinto, Tusa Morada y En. Mex.

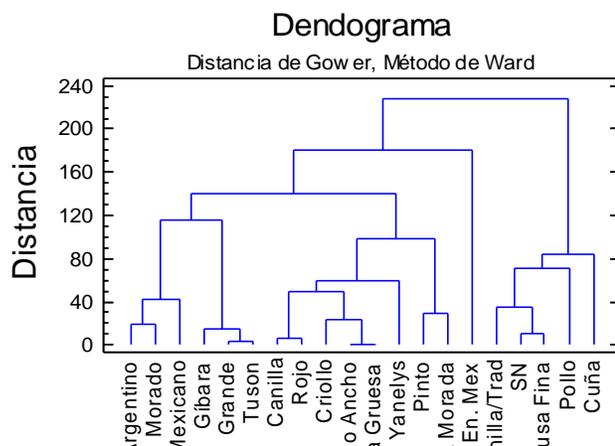
Grupo 4: Variedades: Canilla/Trad, SN, Tusa Fina, Pollo y Cuña.

En este agrupamiento resultaron más significativos los descriptores que enmarcan las longitudes de la mazorca y del grano, o sea, las variables: LM, DM, LG, AG, FG, además de P100S que identifica el peso de 100 granos que tiene la mazorca (Tabla 2).

A partir de estos resultados el investigador puede decidir la preferencia de cultivar las variedades Gíbara, Grande y Tuson por poseer características agronómicas en promedio superiores al resto de las variedades. Así mismo pueden ser variedades recomendadas para estudios de cruzamiento genético en la obtención de plantas transgénicas.

Por otra parte, el uso de la métrica de Gower con el método de Ward dio lugar a un agrupamiento jerárquico que eliminó los problemas de encadenamiento que obtuvo FERNANDEZ *et al.*, (2006) con estos mismos datos al emplear la técnica del vecino más cercano con la distancia Euclidiana.

**Figura 2.** Dendograma obtenido utilizando la distancia de Gower y el método de Ward.



Tal y como señaló GOWER (1967), el método del vecino más cercano no ofrece grupos claramente separados porque tiende a una agrupación longitudinal.

**Tabla 2.** Significación de los descriptores medidos.

	F	signif.
LM	22,51709	0,000005
DM	7,25979	0,002727
NHG	0,60005	0,624251
NGH	15,86578	0,600047
DT	1,36959	0,287830
LG	4,82817	0,014026
AG	3,29012	0,047849
GrG	0,32049	0,810447
P100S	17,08149	0,000030
FG	0,68548	0,057387
CT	1,72938	0,201216
DHG	1,04167	0,400945

Utilizar diferentes medidas de distancia y distintos criterios de enlace sigue siendo la manera más recomendada para obtener resultados consistentes y agrupamientos naturales.

No obstante, a partir de estos resultados, se puede recomendar que en presencia de variables mixtas se use en primera instancia la métrica de Gower con el método de Ward.

### 3. CONCLUSIONES.

1. Se obtuvieron cuatro grupos de variedades de maíz en los cuales se agrupan varios genotipos con características similares. A partir de ello, se pueden tomar las variedades Gíbara, Grande y Tusón como las recomendadas para estudios genéticos.
2. Los caracteres más significativos en este tipo de estudio resultaron ser las longitudes de la mazorca y del grano y el peso de 100 granos.

3. Se recomienda utilizar la distancia de Gower y el método de Ward para agrupamiento jerárquico a partir de caracteres mixtos.

RECEIVED OCTOBER , 2008  
REVISED DECEMBER 2009

#### REFERENCIAS.

- [1] BOSH, L, CASAÑAS, F, SÁNCHEZ, E y NUEZ, F. (1997): Variability of Maize landraces from northwest Spain. **Plant Genetic Resources Newsletter**. 112: 90-92.
- [2] EVERITT, B. (1980) : **Cluster analysis**. 2th edition. McGraw-Hill Book Company, N. York.
- [3] EYZAGUIRRE, P y LINARES, O. (2004): **Home Gardens and Agrobiodiversity..** Smithsonian Institution Washington.
- [4] FERNÁNDEZ LIANNE, TORRES, M, SÁNCHEZ, M Y RABI, O. (2004) : **El cultivo del Maíz en Cuba**. XX Reunión Latinoamericana del Maíz. Lima. 56-61.
- [5] FERNÁNDEZ LIANNE, CASTIÑEIRAS LEONOR, CRISTÓBAL, R, GARCÍA MARITZA y FUNDORA ZOILA. (2006) : Estudio de la variabilidad *In situ* de maíces tradicionales cubanos en dos regiones rurales de Cuba. **Revista Agrotecnia de Cuba**, 30,40-46.
- [6] GALLINARI, ALESSANDRA., MARÍN, J. M. (2001) : **Métodos de clasificación aplicados en Biología Molecular. Medidas de Similitud**. Disponible <<http://pdg.cnb.uam.es/cursos/complutense2001/pages/classMeth/ts/d035.htm>>.. Consultado: 24 de octubre del 2007.
- [7] GÓMEZ, O. (2006) : **Comportamiento agronómico de tres genotipos de maíz (*Zea mays* L.) bajo diferentes sistemas de labranza en condiciones de sabana**. Jornada Científica del Maíz, 24, 77-83.
- [8] GORDON, A. D. (1990) : **Cluster classification**. Wiley New York.
- [9] GOWER, J. (1967) : A comparison of some methods of cluster analysis. **Biom. J.** 23: 623-637.
- [10] GOWER, J. (1971) : A general coefficient of similarity and some of its properties. **Biometrics** 27, 857-872. 1971.
- [11] HATHEWAY, W. H. (1957) : Races of maize in Cuba. Publication 453. National **Academy of Sciences-National Research Council**. Washington, D. C.
- [12] LINARES, G. (1990) : **Análisis de datos**. Universidad de la Habana. Facultad de Matemática – Cibernética, La Habana..
- [13] LONDOÑO, G. C., LAVALETT, LELIA., GALINDO, PURIFICACIÓN., y AFANADOR, LUCÍA. (2007) : Uso de métodos multivariantes para la agrupación de aislamientos de *Colletotrichum spp* con base en características morfológicas y culturales. **Rev. Fac. Nal. Agr. Medellín**. 60, : 3671-3690.
- [14] MANRIQUE CHÁVEZ, P. A. (1997) : El maíz en el Perú. **Series Tecnologías, Consejo Nacional de Ciencia y Tecnología. (CONCITEC)**, Lima..
- [15] MIRANDA, ILEANA. (1997) : **Análisis de cluster como estrategia multivariada de clasificación. Solución a un problema taxonómico**. Tesis para optar por el título de Master en Matemática Aplicada a las Ciencias Agropecuarias. CENSA- ISAAC..
- [16] MIRANDA, ILEANA y TORRES, VERENA. (1998) : Coeficientes de similaridad para variables mixtas I. Nueva propuesta. **Rev. Protección Veg.** 13, 127-131.

[17] SOTO, A. J., PONZONI, I. y VÁZQUEZ, G. E. (2006). Análisis numérico de diferentes criterios de similitud en algoritmos de clustering. **Mecánica Computacional**. XXV: 993-1011. Disponible <<http://www.cimec.org.ar/ojs/index.php/mc/article/view/551/525>>. Consultado: 17 de octubre del 2007.

[18] SPSS (1993): **SPSS FOR WINDOWS. Release 6.0** SPSS Inc. 1989-1993. Licensed to Univ. Autónoma Madrid. 819225. Madrid, España.

[19] STATSOFT (1998): **STATISTICA FOR WINDOWS. Versión 6.0**. StatSoft Inc. 1984-2001. Licensed 74104. Tulsa, USA.

[20] VARELA, M. (1996) : Aplicación de la estadística multivariada a las ciencias agrícolas. Programa y Resúmenes X Seminario Científico INCA. **Cultivos Tropicales** 17, 115- 116.

[21] WARD, J. (1963) : Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**. 58: 236-244.

[22] WOJCIECHOWSKI, T. J. (1987) : Nearest neighbour classification rule for mixtures of discrete and continuous random variables. **Biometrics** 29, : 953-959.