PESQUIMAT , Revista de la F.C.M. de la
Universidad Nacional Mayor de San Marcos
Vol. XIII N°2, pp. 77-85, Lima - Perú, Diciembre 2010

Cuantificación de las Categorías de Drosófilas y Levaduras Mediante los Métodos de Promedios Recíprocos y Análisis de Varianza

Doriz Gómez[1], *Jorge Condado*[2], *Liliana Huamán*[2],
Walter Acuña[2], *Martha Gonzales*[2] & *Gregoria Ramón*[2]

Resumen: Un problema latente en tablas de contingencia bidimensional es la cuantificación de los niveles de las variables categóricas involucradas. En el presente trabajo se eligen los métodos descriptivos multivariantes no lineales, el Método del Promedio Recíproco y el Método de Análisis de Varianza para la cuantificación de los niveles de *drosófilas* y levaduras [4], variables categóricas de la tabla de contingencia bidimensional. Los resultados encontrados muestran que fue posible la cuantificación de las categorías *drosófilas* y levaduras mediante las metodologías señaladas, las mismas que producen cuantificaciones diferentes a las tradicionalmente usadas para las categorías de las filas y columnas de la tabla de contingencia.
Palabras clave: Drosófilas. Método del Promedio Recíproco. Método de Análisis de varianza.

Quantification of the Categories of Drosophilas and Yeast by the Methods of Reciprocal Averages and Variance Analysis

Abstract: A latent problem in two-dimensional contingency tables is to quantify the levels of categorical variables involved. In this paper we choose the nonlinear multivariate descriptive methods, Reciprocal averaging method and Variance Analysis method for quantifying levels of drosophilas and yeast, categorical variables of two-dimensional contingency table. The results show that it was possible to quantify the categories drosophilas and yeast by the methods mentioned, the same ones that produce different quantifications of the categories traditionally used for the rows and columns of the contingency table.
Key words: Drosophila. Reciprocal averaging method. Analysis of variance method.

1. Introducción

Cuando el comportamiento de los datos no es normal, la obtención de las relaciones lineales no recoge la componente no lineal contenida en los datos, generando que muchas veces no se capte dicha información. En tal sentido, cuantificar las relaciones no lineales entre variables se ha convertido en un tema que ha merecido la atención de muchos investigadores.

Para evitar tal pérdida de información en los datos, aparecieron los métodos descriptivos multivariantes no lineales denominados métodos MUNDA. Autores como Greenacre, Lebart y Nishisato entre otros, propusieron metodologías tales como el Análisis de Correspondencia, el Promedio Recíproco y el Análisis de Varianza.

Para estudiar relaciones de dependencia, cuando se tiene datos cualitativos presentados en tablas de contingencia, son muchos los métodos estadísticos para explorar y confirmar estructuras

[1]UNMSM, Facultad de Ciencias Matemáticas, e-mail: dorisgomezt@gmail.com
[2]UNMSM, Facultad de Ciencias Matemáticas

de asociación entre individuos de una población. Entre las metodologías confirmatorias están los modelos log lineales ya ampliamente discutidos en [2] entre otros, y los métodos de mínimos cuadrados ponderados discutidos y aplicados por [8]. Estos dos métodos se basan justamente en modelos explicativos de relaciones funcionales entre las probabilidades de las celdas de la tabla de contingencia. Como metodología exploratoria importante se tiene el análisis de correspondencia, discutida y muy aplicada en [9] y [6] entre otros. El análisis de correspondencia es extremadamente útil en el estudio de estructuras de asociación y se ha utilizado con mucha frecuencia en Francia, en el estudio de grandes muestras relacionando profesiones y causas de muerte, profesiones e ítems de consumo, profesiones y lugares de esparcimiento en las vacaciones [9]. No obstante su gran aplicación y simplicidad de interpretación geométrica, la medida de disimilaridad usada con mayor frecuencia es la ji cuadrado, basada en las frecuencias observadas (o) y las frecuencias esperadas (e), que tiene serios inconvenientes porque es unidireccional, esto es $\sum(o-e)^2/e \neq \sum(e-o)^2/o$. Este hecho hace que la matriz usada en el estudio de las asociaciones sea semidefinida positiva, lo que implica una pérdida de dimensión que puede ser determinante para entender aspectos importantes de la tabla de contingencia en la que se resume la información para variables cualitativas. Esta desventaja del análisis de correspondencia motivó a [3] a proponer la teoría de Análisis de Dependencia (ANADEP) para tablas de contingencia bidimensional y multidimensional, usando la medida de asociación M, entre filas y columnas de la tabla de contingencia, dada por Khan e Ali en 1973, quienes demostraron que M satisface las propiedades requeridas para un coeficiente de asociación, creando matrices de dependencia que emulan matrices de correlaciones para variables cuantitativas. [3] construyó, a partir de la tabla de contingencia-parámetros como las *dependencias*, *codependencia* y *coeficientes de dependencia* que emulan para el caso de datos categóricos, las varianzas, covarianzas, coeficientes de correlación, correspondientes a datos numéricos, dando origen a la teoría ANADEP. También a partir de los *coeficientes de dependencia* postuló la teoría del Análisis Factorial de Dependencia (ANADEPF).

Mediante el Análisis de Dependencia [3], se consiguió establecer relaciones de asociación entre las especies de moscas: *Drosofila Novemaristata*, *D. Inca*, *D.Huaylasi*, *D. Melanogaster* y las levaduras-tres cepas: *Saccharomyces Cerevisiac*, *Candida utilis* y *Rhodotorula sp.* [6].

El objetivo del presente estudio es usar los métodos descriptivos multivariantes no lineales, método del Promedio Recíproco y el método de Análisis de Varianza en tablas de contingencia de doble entrada, a fin de encontrar la mejor escala que permita cuantificar las categorías de *drosófilas* y levaduras [4].

2. Métodos Estadísticos

2.1. Método del Promedio Recíproco (MPR)

En el contexto de una tabla de contingencia bidimensional, sean:
$F = (f_{ij} : \quad i = 1, \ldots I; \quad j = 1, \ldots, J)$, la matriz de frecuencias absolutas conjuntas, donde $f_{ij} \leq 0$, es el número de individuos escogidos al azar, que pertenecen a los niveles $i = 1, \ldots, I$ y $j = 1, \ldots, J$ de las características A y B;
$f_{i\bullet}$ y $f_{\bullet j}$ los totales por filas y por columnas,
X_j e Y_i los pesos asociados a las columnas y a las filas, respectivamente.

El método del Promedio Recíproco se inicia imputando pesos a los niveles de las columnas, X_j, $j = 1, \ldots, J$, o a los niveles de las filas, Y_i, de la tabla de contingencia (según elección) y a pesar de que estos pesos son arbitrarios, la sugerencia es que se incluya al cero.

Con los pesos, X_j, de los niveles de las columnas, se calculan los pesos o ponderaciones de las filas según:

$$Y_i(Filas) = \frac{\sum_{j=1}^{J} f_{ij}(X_j)}{f_{1\bullet}}, \quad i = 1, \ldots, I \tag{1}$$

Luego se obtiene los valores ajustados por la media, es decir:

$$Y_i = Y_i - M \tag{2}$$

donde la media ponderada por los pesos Y_i, es: $M = \frac{\sum_{i=1}^{I} f_{i\bullet}(Y_i)}{n}$

Para conseguir las ponderaciones finales de las filas, se divide Y_i por G_Y, donde G_Y =máximo$|Y_i|$, o sea,

$$Y_i = \frac{Y_i}{G_Y} \tag{3}$$

Usando las ponderaciones obtenidas para las filas, se re-calcula las ponderaciones para las columnas según:

$$X_j(Columnas) = \frac{\sum_{i=1}^{I} f_{ij}(Y_i)}{f_{\bullet j}}$$

Se procede a encontrar los valores ajustados por la media, es decir,

$$X_j = X_j - N \tag{5}$$

donde $N = \frac{\sum_{j=1}^{J} f_{\bullet j}(Y_j)}{n}$.

Para obtener las ponderaciones finales de las columnas, se divide X_j por G_X, donde G_X =máximo$|X_j|$, o sea,

$$X_j = \frac{X_j}{G_X} \tag{6}$$

Se repiten los pasos aneriores hasta conseguir la convergencia para los pesoso ponderaciones asignados a los niveles de las columnas y filas respectivamente, llegando a las ponderaciones óptima y al igual que a los valores a los que convergen G_X y G_Y. $\rho = \sqrt{G_X G_Y}$, es la máxima correlación entre filas y columnas de la tabla de contingencia [10].

La constante multiplicativa para ajustar la unidad de Y_i es C_r, donde

$$C_r = \sqrt{\frac{n}{\sum_{i=1}^{I} f_{i\bullet} Y_I^2}} \tag{7}$$

y la constante multiplicativa para ajustar la unidad de X_j, está dada por C_c, donde

$$C_c = \sqrt{\frac{n}{\sum_{j=1}^{J} f_{\bullet j} X_j^2}}$$

Las ponderaciones finales se obtienen multiplicando los pesos obtenidos en el último paso de las iteracciones multiplicado por para Y_i, y por C_c para X_j.
Los nuevos pesos se denominan "pesos o ponderaciones normalizadas" y resultan siendo:

$$Y_i = C_r Y_i \qquad X_j = C_c X_j. \tag{9}$$

Las ponderaciones normalizadas multiplicadas por el valor singular son las coordenadas principales o pesos proyectados:

$$Y_i = \rho Y_i \qquad X_j = \rho X_j. \tag{10}$$

2.2. Método de Análisis de varianza

Según el método de Análisis de Varianza a un criterio de clasificación, se otorgan las ponderaciones o escalas de manera que sea máxima las diferencias *entre las unidades* de investigación y que las diferencias *dentro de ellas* sea mínima. Así por ejemplo, si lo que se tienen son opiniones de estudiantes respecto a sus profesores, las opiniones de los estudiantes respecto de *un profesor (dentro)* deben ser lo mas parecidas, mientras que sus opiniones respecto a *los profesores (entre)* deben ser diversas. Este método se basa en el Principio de Consistencia Interna de Guttman (in [10]).
Sean:

$\vec{Y} = \begin{bmatrix} Y_1 \\ \dots \\ Y_I \end{bmatrix}$, $\vec{X} = \begin{bmatrix} X_1 \\ \dots \\ X_I \end{bmatrix}$ los vectores con los pesos para las filas y columnas,

F la matriz de datos,

$f_R = \begin{bmatrix} f_{1\bullet} \\ f_{2\bullet} \\ \dots \\ f_{I\bullet} \end{bmatrix}$, $f_C = \begin{bmatrix} f_{\bullet 1} \\ f_{\bullet 2} \\ \dots \\ f_{\bullet J} \end{bmatrix}$ los vectores con totales por filas y columnas de la matriz F,

D_C la matriz diagonal con los totales por columnas de la matriz F,
D_R la matriz diagonal con los totales por filas de la matriz F.

Los términos del análisis de varianza, suma de cuadrados total (SS_T), suma de cuadrados entre grupos o entre tratamientos (SS_B) y la suma de cuadrados dentro de los grupos o dentro de los tratamientos (SS_W), se expresan como:

$$SS_T = \vec{X}' \left[D_C - \frac{f_c f_c'}{f_t} \right] \vec{X} \quad SS_B = \vec{X}' \left[F' D_R^{-1} F - \frac{f_c f_c'}{f_t} \right] \vec{X} \tag{11}$$

$SS_W = \vec{X}' \left[D_C - F' D_R^{-1} F \right] \vec{X}$ sujeto a las restricciones, $f_C' \vec{X} = 0$, $SS_T = f_t$.

Se procede a encontrar los pesos de las categorías de manera que hace máximo la suma de cuadrados entre tratamientos, SS_B, sujeto a la condición $SS_T = f_t$y la suma de pesos igual a cero,por lo que en la solución se usan multiplicadores de Lagrange.

$$Q(\vec{X}, \lambda_1, \lambda_2) = SS_B - \lambda_1 (SS_T - f_t) - \lambda_2 (\vec{X}' f_c f_C' \vec{X} - 0)$$

$$\partial Q(\vec{X},\lambda_1,\lambda_2)/\partial\vec{X}=0 \qquad \partial Q(\vec{X},\lambda_1,\lambda_2)/\partial\lambda_1=0 \qquad \partial Q(\vec{X},\lambda_1,\lambda_2)/\partial\lambda_2=0$$
$$[F'D_R^{-1}F-\lambda D_C]\vec{X}=0\rightarrow F'D_R^{-1}F\vec{X}=\lambda D_C\vec{X}$$

Esta ecuación se denomina ecuación del autovalor-generalizada, de donde debe obtenerse el autovalor (λ) y autovector $\vec{X}$ de la matriz D_C.

Se demuestra que $\lambda=\dfrac{\vec{X}'F'D^{-1}RF\vec{X}}{\vec{X}'D_C\vec{X}}=\dfrac{SS_B}{SS_T}=\eta^2$, donde η^2 es la *razón de correlación.*

Aquí usaremos indistintamente los términos: *razón de correlación* y *autovalor* indistintamente.

Se introduce una nueva notación, el vector, $\vec{W}=D_C^{-1/2}\vec{X}$, y se reescribe la *razón de correlación* en términos del nuevo vector:

$$\eta^2=\frac{\vec{W}'D_C^{-1/2}F'D_R^{-1}FD_C^{-1}\vec{W}}{W'\vec{W}}=\frac{\vec{W}'B'B\vec{W}}{\vec{W}'\vec{W}}=\eta^2,\ \text{donde}\ B=D_R^{-1/2}F'D_C^{-1/2}.$$

La forma estándar de la ecuación-autovalor es:

$$[B'B-\lambda I]\vec{W}=0\rightarrow\ \text{se obtiene}\ \vec{W}\ \text{y luego}\ \vec{X}=D_C^{-1/2}\vec{W}.$$

La solución trivial mencionada antes corresponde a $\lambda_1=1$ y $\vec{W}_0=\vec{1}$, respecto a la matriz B o a la matriz de datos F. Para eliminar ésta solución trivial, se continúa con el procedimiento estándar para calcular la matriz residual, digamos, la matriz C.

Se calcula la matriz residual:

$$C=B'B-\lambda_0\frac{\vec{W}_0\vec{W}_0'}{\vec{W}_0''\vec{W}_0}=B'B=\frac{D_C^{1/2}\vec{1}|vec1'D_C^{1/2}}{f_t}$$

y la primera componente es la solución de la siguiente ecuación: $(C-\eta^2I)\vec{W}=0$, con su máximo autovalor, digamos η_1^2 y su autovector asociado $\vec{W}_1$.

Considerando el máximo autovector asociado al máximo autovalor η_1^2, esto es $\vec{W}_1$, obtenemos el primer vector de pesos óptimos, $\vec{X}_1$,

$$\vec{X}_1=D_C^{-1/2}\vec{W}_1 \qquad W_1'W_1=\vec{X}_1'D_CX_1=f_t \tag{12}$$

Los escores óptimos para el sujeto i en la componente 1 están dada por: $Y_{i1}=\dfrac{1}{\eta_1}\dfrac{\sum_{j=1} f_{ij}X_{j1}}{f_{i\bullet}}$, donde el vector de escores de la primera componente está dados por $\vec{Y}_1=\dfrac{1}{\eta_1}D_R^{-1}F\vec{X}$.

De manera similar se pueden expresar SS_T, SS_B, SS_W en términos de los pesos para las filas, esto es, $\vec{Y}$, y maximizar la razón de correlación $\dfrac{SS_B}{SS_T}$ sujeto a la condición que la suma de los pesos de las respuestas es cero y que la suma de cuadrados de las respuestas ponderadas es igual a f_t. El resultado es la ecuación-autovalor:

$$[B'B-\lambda I]\vec{V}\rightarrow\ \text{se obtiene}\ \vec{V}\ \text{y luego}\ \vec{Y}=D_R^{-1/2}\vec{V} \tag{13}$$

Análisis y Discusión

Se llevaron a cabo colectactas de moscas *drosófilas* en las localidades de Anta (Provincia de Huaraz) y Choquechaca (Provincia de Caraz), dos áreas de grandes pisos de cactáceas; con la finalidad de estudiar la especificidad en la preferencia de los nutrientes que contienen las levaduras, fuente de proteínas de las drosófilas. Cabe señalar que *Drosophila* y Levaduras se incluyen entre los casos mas fascinantes de asociación a nivel mundial, prestándose a estudios de coevolución ([1] y [2]).

Se pretende cuantificar las categorías de las filas y de las columna de la tabla de contingencia que contiene los resultados de las colectas de moscas *drosófilas* en las localidades de Anta, Provincia de Huaraz y Choquechaca, Provincia de Caraz (Gómez, 2006), mediante las metodologías presentadas.

Tabla N°1. Clasificación de moscas *Drodrosophilas* y la taxonomía de sus Alimentos. Localidades de Anta y Choquechaca. Ancash, 2003

LEVADURAS	Especies de *drosophila*				
	Huaylasi (H) ($X_1 = 0$)	Inca (I) ($X_2 = 1$)	Melonogaster (M) ($X_3 = 2$)	Novemaristata (N) ($X_4 = 3$)	Total
Saccharomyces Cerevisiae (SC) (Y_1)	17	51	162	63	293
Candida utilis (CV) (Y_2)	25	41	57	67	190
Rhodotorula sp (RO) (Y_3)	12	6	24	38	80
Total	54	98	243	168	563

2.3. Método del Promedio Recíproco (MPR)

Se Inicia asignando pesos arbitrarios a las categorías de las columnas, así, $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 4$ y obteniendo los pesos ponderados de las filas, media, pesos ponderados ajustados por la media y finalmente los pesos finales ponderados de las filas. En este primer paso, los valores son $Y_1(SC) = 1,9249$, $Y_2(CV) = 1,8737$, $Y_3(RO) = 2,1$,

$$M = 1,9325, \quad Y_1 = Y_1 - M = 1 - 0,0076, \quad Y_2 = Y_2 - M = 0,0588, \quad Y_3 = Y_3 - M = 0,1675$$

$$G_Y = \text{máximo}|Y_i| = 0,167, \quad Y_1 = \frac{Y_1}{G_Y} = -00454,$$

$$Y_2 = \frac{Y_2}{G_Y} = \frac{-0,0588}{0,1675} = -0,3510, \quad Y_3 = \frac{Y_3}{G_Y} = 1,0.$$

Usando los nuevos valores como ponderaciones de las filas, se calcula los nuevos promedios para las columnas, resultando:

$$X_1(H) = 0,0454, \quad X_2(I) = -0,1092, \quad X_3(M) = -0,0138, \quad X_4(N) = 0,0692$$

$$N = \frac{54(0,0454) + 98X_2 + 243X_3 + 168X_4}{29} = 0,000039$$

$$X_1 = X_1 - N = 0,0454 \qquad X_2 = X_2 - N = -0,1092$$

$$X_3 = X_3 - N = -0,0138 \qquad X_4 = X_4 - N = 0,0692$$

$$G_X = \text{máximo}|X_i| = 0,1092 \qquad X_1 = \frac{X_1}{G_X} = 0,4158$$

$$X_2 = \frac{X_2}{G_X} = -1,0 \qquad X_3 = \frac{X_3}{G_X} = -0,1264 \qquad X_4 = \frac{X_4}{G_X} = 0,6337$$

El proceso continua hasta conseguir la convergencia. A continuación se resumen las ponderaciones "escores" en las diversas iteraciones, llegando a la solución óptima en la cuarta iteración. Observe también que G_X y G_Y convergen a los valores 0,2407 y 0,2613 respectivamente.

Columnas	1	2	3	4	5
X_1	0	0,4158	1,0000	1,0000	1,0000
X_2	1	$-1,0000$	$-0,4718$	0,1842	0,1802
X_3	2	$-0,1264$	$-0,6348$	$-0,6807$	$-0,6808$
X_4	3	0,6337	0,8611	0,7696	0,7691
G_X	0,1092	0,1613	0,2487	0,2467	0,2407
Filas	1	2	3	4	5
Y_1	$-0,0454$	$-0,3359$	$-0,5550$	$-0,7062$	$-0,7055$
Y_2	$-0,3510$	0,0964	0,4346	0,6683	0,6682
Y_3	1,0000	1,0000	1,0000	1,0000	1,0000
G_Y	0,1675	0,2501	0,3368	0,2614	0,2613

Nishisato (1988c) mostró que $\rho = \sqrt{G_X G_Y} = 0,2508$, es la máxima correlación entre filas y columnas.
Según [7] la constante multiplicativa para ajustar la unidad de Y_1, Y_2, Y_3 es $C_r = 1,3462$ y según [8], la constante multiplicativa para ajustar la unidad de X_1, X_2 y X_3 está dada por $C_C = 1,4462$.

Las ponderaciones finales se obtienen multiplicando los pesos obtenidos en el último paso de las interacciones multiplicando por C_r para Y_1, Y_2, Y_3 y por C_C para X_1, X_2 y X_3. Estos nuevos pesos se denominan "pesos o ponderaciones normalizadas" $X_i C_C$, $Y_i C_r$, para los niveles de la especie *drosophila* (columnas): Huaylasi, Inca, Melanogaster y Novemaristata y para los niveles de las levaduras (filas): Saccharomyces, Candida utilis y Rhodotorula sp, respectivamente.

Columnas Ponderaciones normalizadas		Filas Ponderaciones normalizadas	
X_1	$1(1,4462) = 1,4462$	Y_1	$-0,7055(1,3462) = -0,9497$
X_2	$0,1802(1,4462) = 0,2606$	Y_2	$0,6682(1,3462) = 0,8995$
X_3	$-0,6808(1,4462) = -0,9846$	Y_3	$1(1,3462) = 1,3462$
X_4	$0,7691(1,4462) = 1,1123$		

2.4. Métodos de Análisis de Varianza

Se identifican los elementos matriciales:

$$F = \begin{bmatrix} 17 & 51 & 162 & 63 \\ 25 & 41 & 57 & 67 \\ 12 & 6 & 24 & 38 \end{bmatrix} \qquad f_C = \begin{bmatrix} 54 \\ 98 \\ 243 \\ 168 \end{bmatrix} \qquad f_R = \begin{bmatrix} 293 \\ 190 \\ 80 \end{bmatrix} \qquad \vec{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$D_C = \begin{bmatrix} 54 & 0 & 0 & 0 \\ 0 & 98 & 0 & 0 \\ 0 & 0 & 243 & 0 \\ 0 & 0 & 0 & 168 \end{bmatrix} \qquad D_R = \begin{bmatrix} 293 & 0 & 0 \\ 0 & 190 & 0 \\ 0 & 0 & 80 \end{bmatrix} \qquad f_t = 563 = n;$$

donde las ponderaciones óptimas que se buscan son $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$ para las drosophilas, $\vec{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ para las levaduras.
Mediante el software Matlab se obtiene:

$$\frac{f_c f' : c}{f_t} = \begin{bmatrix} 5,1794 & 9,3996 & 23,3073 & 16,1137 \\ 9,3996 & 17,0586 & 42,2984 & 29,2433 \\ 23,3073 & 42,2984 & 104,8828 & 72,5115 \\ 16,1137 & 29,2433 & 72,5115 & 50,1314 \end{bmatrix}$$

$G_2 = transF * invDr * F =$??

3. Conclusiones

Fue posible la cuantificación de las categorías drosófilas y levaduras mediante las metodologías Método del Promedio Recíproco y Análisis de Varianza, pertenecientes a los denominados métodos descriptivos multivariantes no lineales, MUNDA. Sin embargo, dichas metodologías producen cuantificaciones diferentes para las categorías de las drosófilas y levaduras.

4. Agradecimiento

Al Consejo Superior de Investigaciones del Vicerrectorado de Investigación de laUNMSM, por el soporte financiero otorgado el año 2008 para la ejecución del proyecto de investigación cuyo resultado es la presente publicación.

REFERENCIAS BIBLIOGRÁFICAS

[1] Belo,M.; LAVACA,P.(1982). Asociaçao entre drosophila e leveduras. Atraçao e produtividade. Naturalia. N0 7, 35- 45.

[2] Bishop,Y; Fiemberg, S. y Holland,P. (1975). Discrete multivariate analysis: Theory and Practice. The MIT Press, Cambridge, Massachusetts, USA.

[3] Cordeiro,A. (1990). Análise de dependencia. Tese de Livre Docencia. Universidade Estadual de Sao Paulo. Estado de Sao Paulo. Brasil.

[4] Gómez D. y et. al. (2006). Estudio de asociación entre drosophila y levaduras usando análisis de dependencia. PESQUIMAT. Revista de Investigación de la Facultad de Ciencias Matemáticas. UNMSM. Vol IX , N0 1, pág. 63-72.

[5] Greenacre, M. & Blasius,J. (1994). Múltiple Correspondence Analysis and Related Methods. Boca Raton:Chapmanand Hall/CRC. New York.

[6] Grenacre M. (1984). Theory and applications of correspondence analysis. London: Academic Press.

[7] Greenacre , M. & Blasius, J.(2006). Múltiple Correspondence Analysis in the Social Sciences. Academic Press. London.

[8] Grizzle,J.; Starmer, C. e Koch,G. (1969). Analysis of categorical data by linear models. Biometrics, 25, 489-504.

[9] Lebart,L.; Morineu, A. ; Fenelón, J. (1984). Multivariate descriptive statistical analysis. New York : John Wiley & Sons.

[10] Nishisato, S. (2007). Multidimensional nonlinear descriptive análisis. Boca Raton: Chapman & Hall/CRC. New York.

[11] Seber, G.(1984). Multivariate observations. John Willey, New York.

[12] Vásquez, J. (2002). Genética en poblaciones de Drosophila Fallén (Diptera-Hexapoda). Libro de artículos y resúmenes del I Congreso Peruano de Genética Animal: 62-64. Sociedad Peruana de Genética.