

ESTUDIO DE ASOCIACIÓN ENTRE *DROSOPHILA* Y LEVADURAS USANDO EL ANÁLISIS DE DEPENDENCIA

Doris Gómez Ticeran¹, Blanca Martínez², Gregoria Ramón³,
Víctor García⁴, Carlos Peche⁵

RESUMEN.- En este artículo se presenta la metodología de Análisis de Dependencia (ANADEP) para tablas de contingencia propuesta y desarrollada por Cordeiro (1990), usando la medida de asociación M , en el estudio de estructura de asociación en tablas de contingencia bidimensional. Se usa dicho método en el estudio realizado en las localidades de Anta (Huaraz) y Choquechaca (Caraz) del departamento de Ancash en el año 2003. Usando ANADEP se concluye que existe asociación entre las especies de moscas del género *Drosophila* y los tipos de levadura en las plantas huesped-cactáceas.

PALABRAS CLAVE.- ANADEP, estadística M , estadística $Ji - Cuadrado$, género *Drosophila*, levadura.

ABSTRACT.- In this paper, we present the methodology of the Analysis of Dependence (ANADEP) for suggested contingency tables developed by Cordeiro (1990), using the association measures M , in the study of structure of association in bidimensional contingency tables. We use this method in the study which took place in Anta (Huaraz) and Choquechaca (Caraz) of the department of Ancash during 2003. Using ANADEP, we conclude that there exists association among species of *Drosophila* flies and types of yeast in the hosting cactaceas plants.

KEY WORDS.- ANADEP, M statistics, Chi - Square, *Drosophila* genus, yeast.

1. INTRODUCCIÓN

Dentro del orden Díptera, la familia Drosophilidae está compuesta de 62 géneros, 25 de los cuales se encuentran en la región neotropical que incluye a nuestro país. El género *Drosophila* es el que tiene la mayor cantidad de especies descritas, estimándose su número en más de 2000 (Vásquez, 2002). El Perú, por su ubicación geográfica y sus diversas zonas climáticas debido a la Cordillera de los Andes, presenta áreas excelentes para llevar a cabo una serie de estudios con este género, cuyas reconocidas cualidades, especialmente en trabajos de genética y evolución, datan de más de 80 años y cuyos logros han hecho que el biólogo evolucionista Martín Brookes considere a esta mosca "Héroe anónimo del siglo XX" en Fly: The Unsung Hero of 20th century Science (Wright, 2002). Sin embargo, los conocimientos acerca de sus relaciones tróficas específicas motivan abarcar estudios de las relaciones de dependencia con los microorganismos causantes de la fermentación de sus plantas hospederas, como las cactáceas.

¹ Universidad Nacional Mayor de San Marcos. Facultad de Ciencias Matemáticas. e-mail:dgomezt@unmsm.edu.pe

Desde el punto de vista estadístico, para estudiar dichas relaciones de dependencia, cuando se tiene datos cualitativos presentados en tablas de contingencia, son muchos los métodos estadísticos para explorar y confirmar estructuras de asociación entre individuos de una población. Entre las metodologías confirmatorias están los modelos log lineales ya ampliamente discutidos en Bishop, Fienberg y Holland (1975) entre otros, y los métodos de mínimos cuadrados ponderados discutidos y aplicados por Grizzle, Starmer e Koch (1969). 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 Benzecri (1976), Lebart e Fenelon (1973), Lebart, Morineu y Fenelón (1984), Grenacre (1984), 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, Lebart, Morineau y Fenelón (1984). No obstante su gran aplicación y simplicidad de interpretación geométrica, la medida de disimilaridad usada en este análisis 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 Cordeiro (1990) a proponer la teoría del 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 (1973), quienes demostraron que M satisface las propiedades requeridas para un coeficiente de asociación.

Cordeiro (1990) 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.

En el contexto descrito, se presenta los resultados teóricos más importantes del Análisis de Dependencia (ANADEP). Luego, esta teoría es aplicada a los datos encontrados en las colectas de moscas *drosófilas* en las localidades de Anta (Provincia de Huaraz) y Choquechaca (Provincia de Caraz), a fin de estudiar relaciones de asociación entre las especies de moscas y sus preferencias por nutrientes que contienen levaduras, presentes en las pencas infectadas de los cactus. El análisis se llevó a cabo en el Instituto de Investigación de la Facultad de Ciencias Matemáticas, creando matrices de dependencia que emulan matrices de correlaciones para variables cuantitativas.

2. ANÁLISIS DE DEPENDENCIA (ANADEP) POBLACIONAL

Sean A y B dos características de clasificación de individuos de una población (supuestamente infinita), con A_1, A_2, \dots, A_I y B_1, B_2, \dots, B_J los niveles de A y B respectivamente. Se define:

$$P = (p_{ij}; i = 1, \dots, I; j = 1, \dots, J) \quad (2.1)$$

la matriz de probabilidades, donde

$$p_{ij} = P(A_i \cap B_j)$$

es la probabilidad de que un individuo escogido al azar de ésta población pertenezca a los niveles i y j de las características A y B respectivamente.

Como $p_{ij} \geq 0$, para todo (i, j) y haciendo:

$$p_{i\cdot} = \sum_{j=1}^J p_{ij} \quad \text{para } 1 \leq i \leq I \quad (2.1)$$

$$p_{\cdot j} = \sum_{i=1}^I p_{ij} \quad \text{para } 1 \leq j \leq J$$

se cumple que:

$$\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1 \quad \text{entonces, los vectores fila:}$$

$$p'_{r,i} = \frac{(p_{i1}, p_{i2}, \dots, p_{iJ})}{p_{i\cdot}} \quad 1 \leq i \leq I \quad (2.3)$$

y los vectores columna:

$$p'_{c,j} = \frac{(p_{1j}, p_{2j}, \dots, p_{Ij})}{p_{\cdot j}} \quad 1 \leq j \leq J \quad (2.4)$$

son distribuciones de probabilidad.

También son distribuciones de probabilidad, el vector marginal fila

$$p'_{r,\pm} = (p_{\cdot 1}, \dots, p_{\cdot J}) \quad (2.5)$$

y el vector marginal columna

$$p'_{c,\pm} = (p_{1\cdot}, \dots, p_{I\cdot}) \quad (2.6)$$

generalmente denominados *perfiles probabilísticos* o solamente *perfiles*.

Con la notación descrita, que son las mismas de ANADEP, colocadas en la forma matricial se tiene la matriz de probabilidades y los vectores de perfiles.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & \cdots & p_{1J} \\ p_{21} & p_{22} & \cdots & \cdots & p_{2J} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{I1} & p_{I2} & \cdots & \cdots & p_{IJ} \end{bmatrix} \quad (2.7)$$

$$p_{c,\pm} = \begin{bmatrix} p_{1\bullet} \\ p_{2\bullet} \\ \cdots \\ p_{I\bullet} \end{bmatrix}, \quad p_{r,\pm} = \begin{bmatrix} p_{\bullet 1} \\ p_{\bullet 2} \\ \cdots \\ p_{\bullet J} \end{bmatrix} \quad (2.8)$$

$$p_{c,j} = \begin{bmatrix} p_{1j} / p_{\bullet j} \\ p_{2j} / p_{\bullet j} \\ \cdots \\ p_{Ij} / p_{\bullet j} \end{bmatrix}, \quad p_{r,i} = \begin{bmatrix} p_{i1} / p_{i\bullet} \\ p_{i2} / p_{i\bullet} \\ \cdots \\ p_{iJ} / p_{i\bullet} \end{bmatrix}; \quad (2.9)$$

Se construye la matriz delta, Δ , en base a las expresiones dadas en (2.7) y (2.8),

$$\Delta = \frac{1}{\sqrt{2}} \left(\sqrt{P} - \sqrt{p_{c,\pm}} \sqrt{p_{r,\pm}} \right) \text{ de donde } I_x J, \quad (2.10)$$

$$\Delta = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{p_{11}} - \sqrt{p_{1\bullet} p_{\bullet 1}} & \sqrt{p_{12}} - \sqrt{p_{1\bullet} p_{\bullet 2}} & \cdots & \sqrt{p_{1J}} - \sqrt{p_{1\bullet} p_{\bullet J}} \\ \sqrt{p_{21}} - \sqrt{p_{2\bullet} p_{\bullet 1}} & \sqrt{p_{22}} - \sqrt{p_{2\bullet} p_{\bullet 2}} & \cdots & \sqrt{p_{2J}} - \sqrt{p_{2\bullet} p_{\bullet J}} \\ \vdots & \vdots & \cdots & \vdots \\ \sqrt{p_{I1}} - \sqrt{p_{I\bullet} p_{\bullet 1}} & \sqrt{p_{I2}} - \sqrt{p_{I\bullet} p_{\bullet 2}} & \cdots & \sqrt{p_{IJ}} - \sqrt{p_{I\bullet} p_{\bullet J}} \end{bmatrix};$$

y se forma la matriz, D , de dependencias y codependencias entre las columnas,

$$D_c = \Delta' \Delta = \begin{bmatrix} d_{c,11} & d_{c,12} & \cdots & d_{c,1J} \\ d_{c,21} & d_{c,22} & \cdots & d_{c,2J} \\ \vdots & \vdots & \vdots & \vdots \\ d_{c,J1} & d_{c,J2} & \cdots & d_{c,JJ} \end{bmatrix} \quad (2.11)$$

donde:

$$d_{c,jj} = \frac{1}{2} \sum_{i=1}^I \left(\sqrt{p_{ij}} - \sqrt{p_{i \bullet} p_{\bullet j}} \right)^2 \quad (2.12)$$

$$d_{c,jj^*} = \frac{1}{2} \sum_{i=1}^I \left(\sqrt{p_{ij}} - \sqrt{p_{i \bullet} p_{\bullet j}} \right) \left(\sqrt{p_{ij^*}} - \sqrt{p_{i \bullet} p_{\bullet j^*}} \right) \quad (2.13)$$

y la matriz, Ψ_c , de coeficientes de codependencias entre las columnas se expresa como

$$\Psi_C = \begin{bmatrix} 1 & \delta_{c,12} & \dots & \delta_{c,1J} \\ \delta_{c,21} & 1 & \dots & \delta_{c,2J} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{c,J1} & \delta_{c,J2} & \dots & 1 \end{bmatrix} \quad (2.14)$$

que contiene los coeficientes de codependencias entre las columnas j y j^* , $j \neq j^*$, donde:

$$\delta_{c,jj^*} = \frac{d_{c,jj^*}}{\sqrt{d_{c,jj} d_{c,j^*j^*}}} \quad \text{y} \quad \delta_{r,jj} = 1. \quad (2.15)$$

Equivalentemente, la matriz de dependencias y codependencias para las filas es

$$D_r = \Delta \Delta' = \begin{bmatrix} d_{r,11} & d_{r,12} & \dots & d_{r,1I} \\ d_{r,21} & d_{r,22} & \dots & d_{r,2I} \\ \vdots & \vdots & \vdots & \vdots \\ d_{r,I1} & d_{r,I2} & \dots & d_{r,II} \end{bmatrix} \quad (2.16)$$

que contiene las dependencias y codependencias entre las filas i y i^* , $i \neq i^*$, donde

$$d_{r,ii} = \frac{1}{2} \sum_{j=1}^J \left(\sqrt{p_{ij}} - \sqrt{p_{i \bullet} p_{\bullet j}} \right)^2 \quad (2.17)$$

$$d_{r,ii^*} = \frac{1}{2} \sum_{j=1}^J \left(\sqrt{p_{ij}} - \sqrt{p_{i \bullet} p_{\bullet j}} \right) \left(\sqrt{p_{ij^*}} - \sqrt{p_{i \bullet} p_{\bullet j^*}} \right) \quad (2.18)$$

y la matriz, Ψ_r , de coeficientes de codependencias entre las filas

$$\Psi_r = \begin{bmatrix} 1 & \delta_{r,12} & \dots & \delta_{r,1I} \\ \delta_{r,21} & 1 & \dots & \delta_{r,2I} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{r,I1} & \delta_{r,I2} & \dots & 1 \end{bmatrix}$$

que contiene los coeficientes de codependencias entre las columnas i y i^* , $i \neq i^*$, donde:

$$\delta_{c,jj^*} = \frac{d_{c,jj^*}}{\sqrt{d_{c,jj} d_{c,j^*j^*}}} \quad \text{y} \quad \delta_{r,ii} = 1. \quad (2.19)$$

Cabe resaltar que las matrices anteriores contienen las dependencias y codependencias en variables cualitativas (emulan varianzas y covarianzas para variables numéricas) y los coeficientes de codependencias (emulan coeficientes de correlaciones para variables numéricas) (Cordeiro, 1990). Estos parámetros tienen interpretaciones y algunas propiedades muy semejantes a las covarianzas, varianzas, desviación estándar y el coeficiente de correlación para variables numéricas. Para analizar una semejanza entre las codependencias y covarianzas es necesaria la siguiente definición.

Definición 1.

La columna j , o el nivel B_j , donde $1 \leq j \leq J$, de la categoría **B** de la tabla de contingencia, es independiente de las filas $A_i, i = 1, 2, \dots, I$, o de la categoría **A**, si

$$p_{ij} = P(A_i \cap B_j) = P(A_i)P(B_j) = p_{i\bullet}p_{\bullet j} \quad (2.20)$$

para todo $i = 1, \dots, I$; es decir: $p_{c,j} = p_{c,\pm}$.

Esta definición en tablas de contingencia, indica independencia de las categorías **A** y **B** para los niveles de una en relación a la otra. Si todos los niveles de una categoría fueran independientes de la otra, entonces las categorías **A** y **B** son independientes, lo que es evidente por (2.20).

Con ésta conceptualización, un resultado semejante al existente para covarianzas entre variables numéricas se muestra a continuación:

Si la columna j fuese independiente de las filas, entonces por (2.13), $d_{c,jj^*} = 0$, para todo $j \neq j^*$, y, $\delta_{c,j}^2$ también será cero. Esto quiere decir que el perfil columna $p_{c,j}$ no se diferencia del perfil marginal $p_{c,\pm}$, esto es, el nivel B_j de la categoría **B** no es diferente, según la categoría **A**, del vector marginal. Cuando esto sucede, el nivel B_j puede ser ignorado del estudio de asociación entre las categorías **A** y **B**. Es decir,

$$\begin{aligned} d_{c,jj^*} &= \frac{1}{2} \sum_{i=1}^I \left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}} \right) \left(\sqrt{p_{ij^*}} - \sqrt{p_{i\bullet}p_{\bullet j^*}} \right) \\ &= 0 \end{aligned}$$

y consecuentemente $\delta_{c,jj^*} = 0$ para todo j^* .

Existen situaciones donde $d_{c,jj^*} = 0$ y las columnas j y j^* están asociadas, de la misma manera que cuando se obtiene una covarianza igual a cero para variables cuantitativas no garantiza independencia estadística. Un valor positivo de d_{c,jj^*} , significa que las probabilidades de los perfiles columna j y j^* tienden a estar del mismo lado cuando son comparadas con las probabilidades del perfil marginal, mientras que un valor negativo indica tendencia contraria. Esto es, si se condiciona la columna j , la probabilidad de un nivel i de la categoría **A** es mayor que su probabilidad marginal, y condicionada a la columna j^* , el valor de la probabilidad condicional de A_i tiende a ser menor que la de la marginal, y viceversa.

Si la codependencia d_{c,jj^*} no es cero, su valor positivo o negativo va a depender también de los pesos marginales de las columnas j y j^* ; de igual manera, si la columna j fuera independiente de

las filas, su dependencia $d_{c,ij}$ será proporcional a su peso marginal, esto es, columnas con pesos marginales mayores tienden a ser más dependientes (e importantes en el estudio de asociación) de las filas.

Khan y Ali (1973) propusieron la estadística

$$M = \left[\frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J (\sqrt{p_{ij}} - \sqrt{p_{i\cdot} p_{\cdot j}})^2 \right]^{1/2} \quad (2.21)$$

como medida de asociación entre filas y columnas de una tabla de contingencia y demostraron que la estadística M satisface muchas de las propiedades requeridas para un coeficiente de asociación.

Debe observarse que se cumple: $M^2 = \sum_{j=1}^J d_{c,ij} = \sum_{i=1}^I d_{r,ii}$. Es decir, sumando las dependencias de

todas las columnas o de todas las filas de la tabla de contingencia, se obtiene el coeficiente de asociación de Khan y Ali.

Definición 2 : La dependencia total de la tabla de contingencia es igual a

$$d^2 = M^2,$$

donde d^2 es una medida de asociación global de la tabla de contingencia.

Si cada columna fuese independiente de las filas, por (2.17) y (2.20) la dependencia total de la tabla de contingencia será cero y las categorías **A** y **B** serán independientes y en consecuencia, cualquier fila será independiente de las columnas.

3. DEPENDENCIAS Y CODEPENDENCIAS MUESTRALES

Sea $T = \{f_{ij}; 1 \leq i \leq I, 1 \leq j \leq J\}$, una tabla de frecuencias absolutas de clasificación de

$n = \sum_{i=1}^I \sum_{j=1}^J f_{ij}$ individuos muestreados (tamaño de muestra) de una población estratificada según

las categorías **A** y **B**, $A = (A_1, \dots, A_I)$ y $B = (B_1, \dots, B_J)$, donde f_{ij} es el número de individuos

muestreados de los niveles A_i de **A** y B_j de **B**. Haciendo $h_{ij} = \frac{f_{ij}}{n}$, $h_{i\cdot} = \sum_{j=1}^J h_{ij}$, $h_{\cdot j} = \sum_{i=1}^I h_{ij}$, los

estimadores máximo verosímiles de los parámetros dados de (2.12), (3.1) y (4.1) se obtienen substituyendo en dichas expresiones: p_{ij} por h_{ij} , $p_{i\cdot}$ por $h_{i\cdot}$, $p_{\cdot j}$ por $h_{\cdot j}$.

Bajo el supuesto de independencia entre filas y columnas de la tabla de contingencia y cuando el tamaño de muestra es grande, la dependencia total satisface la relación (Khan e Ali, 1973):

$$8n d^2 \approx \chi^2 \quad (3.1)$$

donde χ^2 es la estadística ji-cuadrado de Pearson y n es el tamaño de la muestra.

4. PRUEBA DE HIPÓTESIS PARA INDEPENDENCIA EN TABLAS DE CONTINGENCIA

La dependencia total muestral, \mathbf{d}^2 , bajo la hipótesis de independencia, satisface la relación asintótica (3.1).

Por lo tanto, el primer paso en el estudio de asociación en tablas de contingencia es el test de independencia entre las dos categorías de la tabla. Entonces, bajo el supuesto de independencia y el modelo probabilístico multinomial,

$$8n \mathbf{d}^2 \sim \chi_{(I-1)(J-1)}^2,$$

es decir, $8n \mathbf{d}^2$ tiene distribución ji cuadrado de Pearson con $(I-1)(J-1)$ grados de libertad.

También $\mathbf{d}^2 = \sum_{j=1}^J d_{c,jj} = \sum_{i=1}^I d_{r,ii}$, relación que permite descomponer la dependencia total en las

dependencias por filas o las dependencias por columnas.

Luego, es de interés averiguar si para cada una de estas representaciones de la dependencia total

\mathbf{d}^2 , hay alguna parte cuya contribución es nula, o no, en la correspondiente descomposición de \mathbf{d}^2 .

Por lo tanto, se debería contrastar

$$H_{c,j} : d_{c,jj} = 0 \quad (4.1)$$

contra su negación, para algún j , $1 \leq j \leq J$. La aplicación de éstas pruebas de hipótesis será motivo de otra investigación.

5. APLICACIÓN

Con la finalidad de estudiar la asociación entre las especies de moscas del género *Drosophila* y sus preferencias por los nutrientes que contienen levaduras (su fuente de proteínas), presentes en las pencas infectadas de los cactus, se realizaron dos salidas de colectas de especies de *Drosophila* a las localidades de Anta (Provincia de Huaraz) y Choquechaca (Provincia de Caraz), en agosto y septiembre del año 2003 respectivamente. Las muestras de levaduras se recogieron de los cactus en descomposición y se colocaron en tubos esterilizados que contenían medios de cultivo.

Los materiales biológicos colectados se trasladaron a los laboratorios de Genética Evolutiva y Microbiología Ambiental y Biotecnología de la Facultad de Ciencias Biológicas de la Universidad Nacional Mayor de San Marcos, para su determinación e identificación, utilizando para ello los protocolos correspondientes a levaduras y *Drosophilas*.

Las cepas de levaduras identificadas corresponden a *Saccharomyces cerevisiae*, *Candida utilis* y *Rhodotorula sp* y las moscas determinadas fueron *Drosophila novemmaristata*, *D. inca*, *D. huaylasi* y *D. melanogaster*. Luego se procedió a preparar el medio a base de cada cepa de levadura, agar y sacarosa y se separaron en 12 frascos (tres para cada cepa) colocándose en cada uno tres moscas hembras grávidas de cada especie. Este procedimiento se repitió dos veces, teniendo en

cuenta el ciclo biológico de cada especie de *Drosophila*, la temperatura promedio de $25^{\circ}C \pm 2^{\circ}C$, el alimento colocado en cada frasco (40 mililitros) y la humedad del bioterio.

Las moscas adultas emergidas de cada frasco fueron analizadas y contadas, dando un margen de una semana para la emergencia.

Con los datos que son resultado del experimento, se generó la siguiente tabla de frecuencias bidimensionales.

$$T = \{f_{ij}; 1 \leq i \leq 3, 1 \leq j \leq 4\}$$

Levaduras	Especies de <i>Drosophila</i>				Total
	huaylasi	Inca	melanogaster	novemmaristata	
<i>Saccharomyces cerevisiae</i>	17	51	162	63	293
<i>Candida utilis</i>	25	41	57	67	190
<i>Rhodotorula</i> sp	12	6	24	38	80
Total	54	98	243	168	563

en la que se consigna las frecuencias bidimensionales, con $I = 3$ niveles para la variable levaduras (fila), y $J = 4$ niveles para la variable especies de *Drosophila* (columna), siendo ambas de naturaleza cualitativa.

La medida de asociación entre filas y columnas de una tabla de contingencia, que expresa la dependencia total de la tabla de contingencia y obtenida con la fórmula (2.21) toma el valor:

$$d^2 = M^2 = \sum_{j=1}^4 d_{c,jj} = \sum_{i=1}^3 d_{r,ii} = 0,01239.$$

Para el nivel de significación de 0,05, el valor de la estadística de prueba $8n d^2 = 55,8$, resulta altamente significativo comparado con el valor teórico de la distribución ji-cuadrado de Pearson con 6 grados de libertad, $\chi^2_{(0,95,6)} = 12,59$. Por lo que se rechaza la hipótesis de independencia total entre las filas y columnas de la tabla de contingencia, es decir, al nivel de significación de 0,05 existe asociación entre las especies de *Drosophila* y levaduras.

Es importante resaltar, que para probar la hipótesis de independencia en la tabla de contingencia, usando la metodología tradicional se requiere hallar el valor de la estadística de prueba ji cuadrado. Dicho valor es $\chi_c^2 = 53,64$ y se ha obtenido usando el software estadístico SPSS (Gómez y col.,2005), como se ilustra en el siguiente reporte. El valor encontrado es muy similar al valor de la estadística de prueba, $8n d^2 = 55,8$.

6. CONCLUSIONES

A partir de los resultados estadísticos se plantean las siguientes conclusiones.

1. Siendo el valor de la dependencia total $\sum_{j=1}^4 d_{c,jj} = \sum_{i=1}^3 d_{r,ii} = 0,01239$ y con un nivel de significación

del 0,05, el valor de la estadística de prueba $8n d^2 = 55,8$, nos conduce al rechazo de la hipótesis de independencia entre las categorías de la variable levaduras y las especies de moscas del género *Drosophila*. Por lo tanto, se concluye que al nivel de significación de 0,05 existe asociación entre las especies de *Drosophila* y las diferentes categorías de levaduras.

2. El valor de la estadística de prueba usando la metodología del ANADEP, $8n d^2 = 55,8$, es similar al valor de la estadística de prueba obtenido usando la metodología tradicional, $\chi_c^2 = 53,64$.
3. En lo que respecta a las especies de moscas del género *Drosophila*, la dependencia de *D. huaylasi* y *D. inca* son las menores, mientras que la dependencia de *D. novemaristata* es el doble de la *D. inca*, y la de *D. melanogaster* es la mayor de todas.
4. En lo que respecta a las levaduras, *Saccharomyces cerevisiae* y *Rhodotorula* sp aportan igualmente a la dependencia, siendo un poco menor el aporte de *Candida utilis*.
5. Desde el punto de vista de las especies del género *Drosophila*, la levadura *Saccharomyces cerevisiae* es preferida por las *Drosophila* que no prefieren *Candida utilis* y la levadura *Rhodotorula* sp, asimismo se observa que *Candida utilis* y *Rhodotorula* sp son igualmente preferidas por las drosófilas.

REFERENCIAS BIBLIOGRÁFICAS

- [1] **Belo M. e Lavaca P.** *Associação entre drosophila e leveduras. Atração e produtividade.* Naturalia. N°7, 35- 45, (1982).
- [2] **Bishop Y., Fienberg S. and Holland P.** *Discrete multivariate analysis: Theory and Practice.* The MIT Press, Cambridge, Massachusetts, USA., (1975).
- [3] **Clark V. and May S.** *Computer-Aided Multivariate Analysis.* Chapman/Hall/CRC, (2004).
- [4] **Cordeiro Antonio.** *Análise de Dependencia.* Tese de Livre Docencia. Estado de Sao Paulo, Brasil: Universidade Estadual de Sao Paulo, (1990).
- [5] **Fontdevilà A., Wasserman M., Pla C., Pilares L., De Armengol R., Suyo M., Sánchez A., Vázquez J., Ruiz A. and García J. L.** *Description and evolutionary relationships of two species of the Drosophila mulleri cluster (Diptera: Drosophilidae).* Ann. Ent. Soc. Am. 83: 444-452, (1990).
- [6] **Gómez D.** *Análise de dependencia por trajetória.* Sao Paulo, Brasil. 121p. Disertação de Doutorado. ESALQ, Universidade de Sao Paulo, (1997).
- [7] **Gómez D., Vázquez J., Martínez B., Alcarráz M., Adriañola Y. y Condado J.** *Estudio de asociación entre drosophila y levaduras usando análisis de dependencia.* Res. XIV Reunión Científica ICBAR, UNMSM, Lima-Perú (2005).
- [8] **Grizzle J., Starmer C. and Koch G.** *Analysis of categorical data by linear models.* Biometrics, 25, 489-504, (1969).
- [9] **Grenacre M. J.** *Theory and applications of correspondence analysis.* London: Academic Press, (1984).
- [10] **Lebart L. et Felon J. P.** *Statistique et Informatique Appliquées.* Paris: Dunod., (1973).
- [11] **Lebart L., Morineau A. and Felon J.** *Multivariate descriptive statistical analysis.* New York : John Wiley & Sons, (1984).
- [12] **Khan A. and Ali S.** *A new coefficient of association.* Ann. Inst. Statist. Math., 25, 41-50, (1973).
- [13] **Pillai K. and Young D.** *On the exact distribution of Hotelling's generalized T_0^2 .* J. Multivar. Analysis, 1, 90-107, (1971).
- [14] **Vázquez J. et al.** *Description and evolutionary relationships of two species of the Drosophila Mulleri cluster.* Entomological Society of América, (1990).
- [15] **Vázquez J.** *Genética en poblaciones de Drosophila Fallén (Diptera-Hexapoda).* Libro de artículos y resúmenes del I Congreso Peruano de Genética Animal: 62-64. Soc. Per. de Genética, (2002).