

# APLICACIÓN DEL ANÁLISIS DISCRIMINANTE EN EL ESTUDIO DE LAS ENFERMEDADES DIARREICAS EN EL GANADO PORCINO

Dunia Chávez Esponda

Departamento Ciencias Básicas. Universidad Agraria de La Habana (UNAH). Carretera de Tapaste y Autopista Nacional, San José de las Lajas, La Habana. Cuba.

## RESUMEN.

Se presenta un ejemplo de aplicación de la técnica de Análisis Discriminante para estudiar las enfermedades diarreicas en el ganado porcino. La función discriminante óptima fue utilizada para predecir si un cerdo se enfermaría o no teniendo en cuenta las condiciones climáticas y del lugar donde se encontraban estos animales. Se logró además determinar los caracteres más significativos ante esta situación, estos fueron: la humedad del local, la temperatura ambiental y la humedad relativa. Esta función fue obtenida utilizando las diferentes funciones discriminantes de los sistemas estadísticos estudiados, teniendo en cuenta que el error de clasificación fuera mínimo.

## ABSTRACT

An example of the use of Discriminant Analysis techniques is developed for studying diarrheic diseases in pork cattle. The optimal discriminant function was used for predicting if a pork is going to be sick or not taking into account the climatic conditions and the place where the animals are situated. In addition were determined the most significant characteristics in that situation, they were humidity of the local, temperature of the environment and relative humidity. That function was obtained using different discriminant functions of the studied statistical systems taking into account the classification error be minimum.

**KEY WORDS:** optimal discriminant function, misclassification, Fisher discriminant

MSC 62P10

## 1. INTRODUCCIÓN.

El Departamento de Prevención de la Facultad de Medicina Veterinaria de la Universidad Agraria de La Habana, entre otras funciones, se encarga de prevenir determinadas enfermedades en diferentes especies de animales, entre ellas está la prevención de la diarrea en los cerdos.

En el mundo y en Cuba no se conocen estudios para prevenir las enfermedades diarreicas desde el punto de vista netamente estadístico a través de un modelo. Estas malezas han sido estudiadas desde el punto de vista veterinario, a través de las manifestaciones clínicas del animal. En la bibliografía consultada se presentan vacunas para la enfermedad, guías técnicas de prevención pero ningún trabajo desde este punto de vista; utilizando técnicas de la Estadística Multivariada, CARVAJAL (2002).

Durante algunos años, este departamento ha venido presentando serios problemas con la detección temprana de una posible enfermedad diarreica en el ganado porcino. A consecuencia de esto, los animales se enfermaban, se deshidrataban y morían sin poder prevenir a tiempo la pérdida de animales tan importantes en nuestro país.

El objetivo del presente trabajo es encontrar un modelo que permita predecir si el cerdo se enfermará o no de diarrea, teniendo en cuenta algunos indicadores corporales, condiciones climáticas y condiciones del lugar donde se encontraban estos animales. Además, se pretende concluir cuáles de estos indicadores o variables son más significativas en este tipo de estudio. De esta manera se podrá apreciar como las técnicas de Estadística Multivariada tienen gran aplicación en muchos problemas reales de índole agropecuaria.

Como es conocido, en la actualidad, el análisis estadístico es el complemento básico de casi todas las investigaciones. Muchos investigadores lo ven como la confirmación de un resultado y en la práctica, la estadística no solo confirma, sino que con cierto grado de incertidumbre

predeterminado por una probabilidad de error, obtiene resultados sobre la base del experimento, ANDERSON (1968).

En múltiples situaciones de la vida no puede hacerse uso de los métodos matemáticos determinísticos para conocer la verdad, ya que el conocimiento del fenómeno es incompleto y están presentes factores aleatorios que no permiten explicarlo. De ahí que en el desarrollo de un gran número de ciencias particulares como la Biología, la Agricultura, etc, los métodos estadísticos jueguen un papel fundamental y por tanto se haya extendido en la época actual un nuevo estilo de trabajo donde el estadístico y el especialista de la ciencia particular trabajen mancomunadamente, BOUZA- SISTACHS (2002).

En toda investigación agropecuaria están interactuando factores que pueden ir desde las características del animal, las condiciones del lugar hasta los factores ambientales o climáticos, los cuales ejercen un efecto conjunto en los resultados finales de un experimento. VARELA (1996).

En estos casos, no es adecuado llevar a cabo una serie de análisis univariado para cada una de las variables, ya que en ellos se ignoran estructuras de correlación llegando incluso a conclusiones falsas del resultado final del experimento, BOUZA- SISTACHS (2002).

En los procesos actuales se involucran un gran número de variables con comportamientos disímiles. De ahí que surge el análisis multivariado como el método esencial para resolver problemas en que estén involucrados varios indicadores. CUADRAS (1992) ; LINARES, ACOSTA y SISTACH (1986).

Dentro del Análisis Multivariado existen muchas técnicas de gran aplicación en las investigaciones agropecuarias. Debido a las características y objetivos del problema planteado, es posible buscar una solución a través de la técnica estadística de Análisis Discriminante, ANDERSON (1968).

Las primeras ideas de esta técnica surgen en la cuarta década del siglo XX, debido a algunas investigaciones biológicas y antropométricas. Mahalanobis en 1930 y Fisher en 1936 fueron sus principales fundadores. Esta técnica siempre se ha utilizado para clasificar un individuo en una de las poblaciones dadas a priori, a partir de la información brindada por muestras aleatorias de cada una de ellas. BOUZA- SISTACHS (2002). ; HERNÁNDEZ.(1997).

Existe otra técnica estadística para clasificar como es el Análisis de Cluster, pero esta última es una técnica exploratoria que sólo nos permite describir los datos. Sin embargo, resulta ventajoso utilizar el Análisis Discriminante, pues éste nos permite describir, seleccionar las variables que más influyen en el problema, construir una función (modelar) a partir de estas variables y, además, predecir en qué grupo se clasifica un nuevo individuo, el cual ha sido evaluado en dicha función. RODRÍGUEZ ( 2004).

Este método aunque es netamente estadístico, tiene como ventaja fundamental, la de predecir y saber si el cerdo se enfermará o no en un futuro, o sea, la prevención temprana de la enfermedad. Esto resulta importante, pues de esta manera se pueden evitar pérdidas en el sector animal.

## **2. MATERIALES Y MÉTODOS.**

El estudio partió de un experimento en el que se tenía un grupo de 84 cerdos a los cuales se le midieron algunos indicadores corporales y otros como son: variables climáticas y condiciones del lugar donde se encontraban. Se confeccionó una base de datos de 84 filas (cerdos) y 6 columnas (variables) en FOXPRO 5.0 con las siguientes variables, todas continuas:

X<sub>1</sub> = PESO (peso)  
X<sub>2</sub> = ESTATURA (estatura)  
X<sub>3</sub> = TEMP (temperatura ambiental)  
X<sub>4</sub> = HR (humedad relativa)  
X<sub>5</sub> = HUMD (humedad del local)

Se creó además la variable  $X_6$  denominada DIARR. La misma es binaria; toma el valor 1 si el cerdo tuvo diarrea y 0 en caso contrario.

El objetivo de los investigadores que plantearon el problema fue encontrar una función que les permitiera clasificar o decidir a qué clase pertenece un cerdo, del cual poseen sus valores de las variables analizadas.

Luego, para clasificar los cerdos en dependencia de estas variables se realizaron análisis discriminantes, y se consideraron las diferentes funciones discriminantes presentes en los sistemas estadísticos actuales. CHÁVEZ, BURGOS y HERNÁNDEZ (2000)

Los paquetes estudiados fueron: CSS, SAS, STATGRAPHICS, SPSS, STATISTICA, SYSTAT, STATITCF Y AD, CSS (1991); SAS,. (1982) ; SUÁREZ (1996). .

Las funciones discriminantes son lineales respecto a los parámetros y fueron las siguientes: la de Fisher, la canónica, la logística, la cuadrática respecto a las variables, la factorial y la calculada a través de la regresión. Una vez obtenidas estas funciones lineales, se seleccionó la función discriminante óptima, teniendo en cuenta que la estimación del error de clasificación sea mínima, pues éste es el parámetro de bondad del Análisis Discriminante. Este error se obtuvo utilizando los métodos de Resustitución y Validación Cruzada. BELLO (1998); . CHÁVEZ, D; BURGOS, T; HERNÁNDEZ, N. (2000). ; KNOKE (1996). ; TARDIF and HARDY. (1996).

### 3. RESULTADOS Y DISCUSIÓN.

El análisis discriminante se desarrolló en cada uno de los sistemas mencionados anteriormente, para obtener un conjunto de funciones lineales discriminantes con su correspondiente error. La estimación del error de clasificación se usó para seleccionar la función discriminante óptima, la cual serviría a los investigadores para clasificar a los cerdos en futuros enfermos o no. CHÁVEZ, BURGOS y HERNÁNDEZ (2000)

En la Tabla 1, se muestran las soluciones parciales obtenidas en cada uno de los sistemas que se relacionan en la primera columna. Como se puede apreciar, la solución óptima sería la última por poseer un menor error de clasificación. CHÁVEZ, BURGOS y HERNÁNDEZ (2000)

Además, a través de esta técnica, también se pudo concluir que la humedad del local era el carácter más significativo en este estudio, pues aparece en todas las funciones de la Tabla 1 y, porque al llevar a cabo los métodos de selección de variables: paso a paso, ascendente o descendente en muchos sistemas siempre la variables humedad del local estaba presente en el modelo.

Es posible apreciar también, que las variables temperatura ambiental y humedad relativa son significativas en este estudio, lo que en menor cuantía que la humedad del local.

Finalmente, se puede concluir que de las variables estudiadas, las que más influyen para prevenir las enfermedades diarreicas en los cerdos son: humedad del local, temperatura ambiental y humedad relativa con el uso de la siguiente función:

$$F(X) = -7.0765X_1 + 0.1335X_3 + 2.7190X_4 + 0.8777X_5 - 9.7874$$

Una vez obtenida la función, la regla a seguir para prevenir la diarrea sería evaluando las mediciones tomadas en dicha función y siguiendo el siguiente criterio:

Si  $F(X) > 0$ , entonces el cerdo se enfermará de diarrea.

Si  $F(X) \leq 0$ , entonces el cerdo no se enfermará de diarrea.

**Tabla1. Resultados del análisis discriminante. / Discriminant analysis results.**

Sist. Estad.	Tipo de Función	Función Discriminante	Error Resust.	Error Cross. Val.
CSS [8]	Canónica ascend.	$F(X)=0.1522 X_3+2.9001 X_4+ 0.7283 X_5 - 10.2553$	0.3131	--
SAS [17]	Canónica cuadrát.	$F(X)=0.2134X_1^2-1222.657X_2^2-0.1512X_3^2+ 50.131X_4^2+38.0324X_5^2+25.3216X_1X_2- 0.0524X_1X_3+21.1254X_1X_4+20.5311X_1X_5 +10.2004X_2X_3-800.6115X_2X_4- 664.6265X_2X_5+3.1045X_3X_4+2.7235X_3X_5 +101.2421X_4X_5-32.2221X_1+1330.1829X_2- 5.9298X_3-159.4511X_4-165.4141X_5 -301.9713$	0.2987	--
STAT GRAPHIS	Canónica	$F(X)=5.9110X_1-7.6125X_2+0.1324X_3+ 2.8998X_4+ 0.7023X_5 - 11.1227$	0.3221	--
SPSS	Canónica paso a paso	$F(X)=0.6988X_5 - 10.1630$	0.3312	--
STATIS TICA	Fisher ascend.	$F(X)=0.1223X_3+2.8672X_4+0.7011X_5 -11.2110$	0.3116	--
SYSTAT	Regresión	$F(X)=0.129X_3+1.325X_4+0.124X_5+1.198$	0.4980	--
STATITC	Factorial	$F(X)=0.6990X_1-0.2655X_2+0.2088X_3+ 2.870X_4+ 0.9024X_5$	0.3302	--
AD [18]	Selecc de vars.	$F(X)=-7.0765X_1+0.1335X_3+2.7190X_4+ 0.8777X_5 -9.7874$	0.2695	0.2101

#### 4. CONCLUSIONES.

1. Se logra encontrar una función discriminante óptima con la cual se puede prevenir si el cerdo se enfermará o no de enfermedades diarreicas. Esta es la siguiente:

$$F(X) = -7.0765X_1 + 0.1335X_3 + 2.7190X_4 + 0.8777X_5 - 9.7874$$

2. Como complemento de ello, se pudo concluir que la humedad del local es la característica más significativa en la clasificación de cerdos en enfermos o no de diarrea, así como, la temperatura ambiental y la humedad relativa también son significativos en este estudio.
3. Con el uso de la técnica de Análisis Discriminante y la forma en que se obtuvo la Función Discriminante Óptima, se ofrece a los investigadores de todas las ramas, una metodología para dar soluciones más completas a las aplicaciones prácticas de esta teoría.

Received March 2006  
Revised March 2007

#### REFERENCIAS.

ANDERSON, T. W. (1968). **An Introduction to multivariate statistical analysis**. John Wiley. New York. 121-142.

BELLO, A.L. (1998). On the performance of rank transforms discriminant method in error-rate estimation. **J. Statist. Comput. Simul.** 48: 153-165.

BOUZA, C; SISTACHS, V. (2002). **Estadística. Teoría Básica y ejercicios**. Editorial Félix Varela. La Habana.

CARVAJAL, A. M. (2002). Epidemiología y diagnóstico de la infección por el coronavirus de la diarrea epidémica porcina.

<[http://www.cibernetia.com/tesis\\_es/CIENCIAS\\_AGRARIAS/CIENCIAS\\_VETERINARIAS/1](http://www.cibernetia.com/tesis_es/CIENCIAS_AGRARIAS/CIENCIAS_VETERINARIAS/1)>.

CHÁVEZ, D; BURGOS, T; HERNÁNDEZ, N. (2000). Metodología propuesta para el cálculo de la Función Discriminante Óptima. **Revista Investigación Operacional**. 21, 218-223.

CID, D. 2002. Caracterización de estirpes de Escherichia Coli aisladas de diarreas neonatales de cerdos.  
<[http://www.cibernetia.com/tesis\\_es/CIENCIAS\\_AGRARIAS/CIENCIAS\\_VETERINARIAS/VIROLOGIA\\_VETERINARIA/2](http://www.cibernetia.com/tesis_es/CIENCIAS_AGRARIAS/CIENCIAS_VETERINARIAS/VIROLOGIA_VETERINARIA/2)>.

COOLEY, W. W. and P. R. LOHNES.(1971). **Multivariate Data Analysis**. John Wiley. New York. 563p.

CSS (1991) **STATISTICA**. StatSoft, N. York.

CUADRAS, C. M. (1996). **Métodos de Análisis Multivariante. Estadística y Análisis de Datos**. Promociones y Publicaciones Universitarias S. A. Barcelona.

HERNÁNDEZ, N. P.(1997). Regla de clasificación para dos poblaciones normales con matriz de covarianza muestral singular. **Revista Investigación Operacional**. 13, 19-25.

KNOKE, D. J. (1996). The robust estimation of classification. Error rates. **Comp and Maths. Appls**. 12 , 253-260.

LINARES, G; ACOSTA, L; SISTACH, V. (1986). **Estadística Multivariada**. Universidad de La Habana.

RODRÍGUEZ, M.( 2004). Análisis Discriminante.  
<<http://www.telefonica.net/web2/biomates/disc/disc.htm>>.

SAS,. (1982). **Introduction to SAS. User´s Guide: Statistics**. SAS Institute Inc, N. Jersey.

SUÁREZ, I. (1996). Sistema de Cómputo de Análisis Discriminante. **Trabajo de Diploma. Universidad de La Habana**.

TARDIF, B and HARDY. (1996). Assessing the relative contribution of variables in canonical discriminant análisis. **Biometrics** 44, 69-76.

VARELA, M. (1996). Aplicación de la estadística multivariada en las ciencias agropecuarias. **Programa y Resúmenes X Seminario Científico INCA. Cultivos Tropicales** 17, 21-26.