

Análisis discriminante canónico con técnicas gráficas multivariadas aplicado a un diseño con dos factores

Canonical discriminant analysis with graphical hypothesis-error and structure of a multivariate design with two factors

Javier Lorbes Medina¹, Yelitza Garcia-Orellana¹, Manuel Milla Pino² y Lisbeth Diaz¹

¹Departamento de Ingeniería. Decanato de Agronomía. Universidad Centroccidental "Lisandro Alvarado". Barquisimeto, Venezuela. javierlorbes@ucla.edu.ve; yelitzagarcia@ucla.edu.ve; ldiaz@ucla.edu.ve;

²Instituto universitario de Tecnología de Yaracuy, San Felipe, Eepartamento de Ingeniería. Decanato de Agronomía. Universidad Centroccidental "Lisandro Alvarado". Barquisimeto, Venezuela. memilla22@yahoo.com.mx. §Autor de correspondencia.

Fecha de recepción: 09/02/2015 Fecha de aceptación del artículo: 10/02/2015

Resumen

Se describen dos técnicas gráficas multivariantes en el que se aplica e ilustra la metodología para proveer una vista en baja dimensión de resultados obtenidos de un diseño con dos factores multivariado basado en el análisis discriminante canónico con las gráficas Hipótesis-Error (HE), que provee una comparación visual directa de las matrices de covarianza para la hipótesis y error; y de estructura discriminante canónica, que muestra una vista alternativa para todas las variables en un espacio bidimensional que maximiza las diferencias entre los tratamientos y proveen un resumen visual compacto de las características resaltantes de los resultados, mostrando todas las observaciones, medias de tratamientos y sus relaciones con las respuestas. En una aplicación, se demuestra el alcance y potencial que ofrece el análisis discriminante canónico con estas gráficas como alternativa para el análisis de datos e interpretación de resultados provenientes de diseños experimentales complejos.

Palabras clave

Análisis discriminante canónico, gráfica HE canónica, variables canónicas.

Abstract

In this paper two are described graphics techniques for multivariate data in the context of canonical discriminant analysis applied to a multivariate design with two factors and graphic representations of results generated application. In particular, the methodology described and illustrated to provide a low dimensional view data obtained from a design with two factors based multivariate canonical discriminant analysis with graphic data in reduced for multiple responses such as graphical range Hypothesis-Error (HE), which provides a direct visual comparison of the covariance matrices for the hypothesis and error, and the graph of canonical discriminant structure, which shows an alternative view for all variables in two-dimensional space that maximizes the differences between the groups, which provide a compact visual summary of the salient features of the results as they show all observations, group means, and their relationships to varying responses. In one application, the scope and potential of the canonical discriminant analysis with these graphs as an alternative for complex data analysis demonstrated experimental designs.

Keywords

Canonical discriminant analysis, canonical discriminant graphic, graphics HE, graph canonical structure, canonical variables.

1. Introducción

La necesidad de obtener conclusiones válidas desde el punto de vista estadístico, y por ende biológico, en relación con los resultados extraídos de muchos experimentos, particularmente del campo agronómico, hace que se planteen estudios que generen metodologías de análisis acordes a los intereses particulares de cada investigador, muchas de las cuales existen pero son poco conocidas o estudiadas o bien conocidas pero no utilizadas, sea por lo novedosa o por la complejidad que encierran. Muchas de estas metodologías subyacen en el ámbito de la estadística multivariante donde el problema de reducción de la dimensionalidad de los datos es fundamental para cada aplicación experimental.

En muchos problemas experimentales, principalmente agronómicos y biológicos, las observaciones realizadas no son independientes entre sí, aunque pueden ser divididas en grupos conocidos a priori, y en esta situación, la atención está generalmente dirigida a la discriminación entre los grupos. Sin embargo, existen las llamadas variables canónicas, que son una combinación lineal de las variables originales, pero orientada a maximizar las diferencias entre los grupos, en lugar de la variabilidad total. En otras palabras, dado un conjunto de observaciones experimentales multivariadas, se busca una nueva perspectiva que, con reducido número de dimensiones, muestre mejor las diferencias entre los grupos, dejando de lado las diferencias entre los individuos y este procedimiento implica el análisis de datos multivariante (Onofri y Ciricofolo, 2004).

La limitación fundamental de las técnicas multivariantes es la complejidad en la presentación de los resultados y la interpretación de los mismos, sobre todo cuando se tienen muchas variables, debido a las interrelaciones entre ellas y a la variabilidad de características asociadas a los modelos, lo cual ha conducido a muchos investigadores a realizar análisis univariantes para cada una de las variables por separado, con los consecuentes errores u omisiones en la interpretación de sus resultados (Amaro et al. 2004).

Una técnica estadística multivariada bien desarrollada que puede ser usada como ayuda o complemento en la interpretación de los resultados de un diseño experimental con enfoque multivariado es el Análisis Discriminante Canónico (ADC), el cual permite la identificación de diferencias entre grupos (clases o tratamientos) a partir de medidas de características de los individuos pertenecientes a esos grupos y facilitando la comprensión de las relaciones entre las variables evaluadas dentro de esos grupos (Afifi y Clark, 1996). El ADC estima funciones lineales (funciones o variables canónicas) a partir de las variables cuantificadas realizando la separación de los grupos de individuos al maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos (Cruz-Castillo et al. 1994).

Esta técnica es ampliamente usada para analizar datos multivariados con estructuras de grupos (como los provenientes de diseños experimentales), y es matemáticamente equivalente a un análisis de varianza multivariado de una vía, que a menudo persigue objetivos similares y que en ocasiones se aplica conjuntamente con ésta (Amaro, 2001). Mientras que en análisis de varianza multivariante la meta es probar la igualdad del vector de medias a través de los niveles de las clases o grupos, en un ADC se encuentran combinaciones lineales de las variables cuantitativas que proveen máxima separación entre las clases o grupos.

El origen del método se remonta a Fisher (1936), o en forma más directa a Mahalanobis (1936), y ha recibido por los autores diferentes nombres, tales como: Análisis de Coordenadas Discriminantes (Seber, 1984), Análisis de Variables Canónicas (Gittins, 1985; Krzanowski, 1988; Gower, 1989); Análisis Factorial Discriminante (Lebart et al. 1995) y Análisis Canónico de Poblaciones (Cuadras, 1996; 2007) entre otros muchos.

Las metas del ADC son: primero, encontrar los ejes de discriminación más grandes entre los grupos identificados a priori, es decir, encontrar combinaciones lineales de las variables de partida con máximo poder discriminante entre los grupos;

segundo, probar si las medias de esos grupos a lo largo de esos ejes son significativamente diferentes, es decir, hallar una ordenación de los grupos de individuos cada uno representado por el vector de las medias en todas las variables, y tercero, estudiar la dimensionalidad de los datos. Cuando hay dos grupos a ser separados, la técnica es conocida como Análisis de la Función Discriminante (AFD); con más de dos grupos, la misma se conoce como Análisis de Variables Canónicas (AVC) o Análisis Discriminante Canónico.

El ADC tiene un considerable interés en el campo agronómico, por al menos dos razones. En primer lugar, gran parte de la experimentación agronómica tiene un enfoque multivariado, en el sentido de que, independientemente del objetivo y el tipo de diseño, los experimentos se repiten en varios bloques, en varios años y/o varias localidades, teniendo en consideración cada vez un número de elevado de características. Por otra parte, los datos son naturalmente divididos en grupos y los individuos son diferentes ya que pertenecen a diferentes tratamientos experimentales y tiene mucho más sentido discutir las semejanzas más o menos pronunciadas, en lugar de aceptar o rechazar la hipótesis nula a veces muy artificial (Onofri y Ciricifolo, 2004).

En años recientes, han surgido una variedad de nuevos métodos gráficos para datos multivariados en el contexto del modelo lineal multivariado, destinados a entender cómo la variación es reflejada en pruebas multivariadas y mostrar como los efectos de los factores son expresados a través de múltiples respuestas. Entre tales métodos gráficos se encuentran las gráficas HE canónicas (Friendly, 2006, 2007; Egesel et al. 2011; Friendly et al. 2013; Friendly y Sigal, 2014) que proveen una comparación visual directa de las matrices de covarianza de la hipótesis (H) y error (E), y su uso en aproximaciones o proyecciones de bajo rango de datos multivariados para desplegar una vista en baja dimensión de los resultados obtenidos al utilizar un análisis discriminante canónico. La gráfica de estructura discriminante canónica, además de la gráfica HE son nuevos métodos integrados al ADC y es diferente de otros métodos gráficos (por

ejemplo Escalamiento Multidimensional, Análisis de Componentes Principales, Biplots, etc.) en que permite evaluar variables respuestas y factores juntos.

El objetivo principal de este trabajo fue analizar un diseño con dos factores multivariante utilizando la técnica de análisis discriminante canónico con las gráficas Hipótesis-Error y de estructura discriminante canónica para una mejor presentación y fácil interpretación de resultados experimentales generados con este tipo de diseño.

2. Materiales y métodos

El material experimental usado en este estudio se obtuvo de un experimento poscosecha con frutos de pimentón (*Capsicum annuum* L.) de la variedad Júpiter, seleccionando para ello frutos almacenados a 10 °C. Los frutos se evaluaron con y sin envoltura plástica, a los 7, 14, 21 y 28 días de almacenamiento, para un arreglo factorial 2 X 4; es decir, se evaluaron los factores: Plástico (P): con envoltura plástica (C) y sin envoltura plástica (S) y Tiempo (T) de almacenamiento (7, 14, 21 y 28 días). Las 8 combinaciones de tratamientos (TRAT) codificadas interactivamente son: C-07, C-14, C-21, C-28, S-07, S-14, S-21, S-28. El diseño experimental utilizado fue completamente al azar con 3 repeticiones de cinco frutos cada una. Se midieron las siguientes variables respuestas: Pérdida de Peso (PP) en gramos, Firmeza (FIR) medida en newton, Sólidos Solubles Totales (SST), expresados en °Brix, Acidez Titulable (AT), expresado como g% de ácido cítrico, Relación Sólidos Solubles Totales/Acidez Titulable (SSAT) y pH.

El análisis estadístico fue hecho basado en el modelo lineal multivariante siguiente:

$$y_{ijkh} = \mu_h + P_{ih} + T_{jh} + (PXT)_{ijh} + \varepsilon_{ijkh}$$

donde $i=1,2$ $j=1,2,3,4$; $k=1,2,3$; $h=1,2, 3,6$; y_{ijkh} es el vector multivariado de la observación k sobre la variable h para la envoltura i y el tiempo j ; μ_h es el vector multivariado de medias general sobre la variable h ; P_{ih} es el

vector multivariado de efectos de la envoltura plástica i sobre la variable h , T_{jh} es el vector multivariado de efectos del tiempo j sobre la variable h , $(PXI)_{ijh}$ es el vector multivariado de efectos de interacción entre la envoltura i y el tiempo j sobre la variable h y ε_{ijkh} es el vector multivariado para errores aleatorios asociado con el vector de observaciones y_{ijkh} .

El análisis discriminante canónico con las técnicas gráficas Hipótesis-Error (HE) y de Estructura Discriminante Canónica fue utilizada para analizar un experimento poscosecha con dos factores y seis variables respuestas. Está basado sobre un modelo lineal multivariado, en las que para la gráfica HE, la elipse de la hipótesis (H) es obtenida de las sumas de cuadrados y productos para la hipótesis (efectos principales e interacción), mientras que la elipse del error (E) es obtenida de las sumas de cuadrados y productos para el error. Por otro lado, la gráfica de estructura discriminante canónica, que muestra una vista alternativa para todas las variables en un espacio bidimensional que maximiza las diferencias entre las combinaciones de tratamiento, provee un resumen visual compacto de las características resaltantes de los datos ya que muestra todas las observaciones, las medias de los tratamientos sobre las dos variables canónicas más grandes, círculos de confianza para estas medias, y vectores variables mostrando las correlaciones de las variables respuestas con las funciones canónicas.

Para generar las gráficas HE y de estructura canónica con el ADC se utilizaron los procedimientos PROC GLM y PROC CANDISC del programa estadístico SAS (SAS, 1999), conjuntamente con los programas macros CANPLOT, HEPLIT y HECAN para las representaciones gráficas (Friendly, 2006; 2007).

En las gráficas HE y de estructura canónica, el orden y separación de las medias de los tratamientos o niveles sobre cada variable canónica indicará cómo esa combinación lineal de las variables respuestas discrimina entre ellos, y los vectores variables sobre la gráfica indicarán la correlación de cada una de las variables observadas con las dimensiones

canónicas. Con ejes igualados, la longitud relativa de cada vector variable será proporcional a su contribución a la discriminación entre los tratamientos, como también, los ángulos entre los vectores indicarán aproximadamente las correlaciones entre las diferencias medias de los grupos, basado sobre la matriz H estandarizada proyectada en el espacio de las dimensiones canónicas.

Así, la longitud relativa de cada vector variable será un indicativo de la proporción de varianza explicada por esa variable y el ángulo entre cada vector variable con cada función canónica será proporcional a su correlación (coeficiente de la estructura canónica total), por tanto, los coeficientes de estructura canónica total medirán la correlación lineal simple bivariada entre una variable respuesta original y la variable canónica, por lo que el coeficiente canónico reflejará la varianza que la variable observada comparte con la variable canónica, y puede ser interpretada en la valoración de la contribución relativa de cada variable a cada función canónica.

Las gráficas HE y de Estructura canónica son nuevos métodos integrados al ADC y que permite evaluar variables y factores juntos. También hace posible interpretar las relaciones entre las variables de acuerdo con el análisis de correlación, por representalos como vectores dibujados desde el origen. Si existe una correlación positiva entre dos variables, los vectores tendrán similar dirección, y mostrarán diversas direcciones en el caso de una correlación negativa. La longitud del vector relacionado indica qué tan bien la variación de la variable puede ser explicada en el espacio bidimensional. En otras palabras, cuanto más largo es un vector, mejor interpretación es posible sobre la variable que representa. Además, las formas de las elipses de la hipótesis (H) y del error (E) en las gráficas HE brinda información acerca de la distribución de los datos y las dimensiones canónicas. Si el círculo de la hipótesis es anular significa que la variación total es explicada por ambas dimensiones canónicas igualmente, mientras que si es un elipsoide plano entonces significa que la primera dimensión explica una proporción mayor que la segunda dimensión. El diámetro del círculo del error indica la tasa de

Tabla 1. Resumen del análisis discriminante canónico para la interacción Plástico X Tiempo

Variable canónica	Correlación canónica	Autovalor	Proporción autovalor	Proporción acumulada	Valor de Probabilidad
Plástico X Tiempo (Lambda de Wilks = 0,025; P = 0,0001)					
Can 1	0,96	11,25	0,848	0,848	0,0001
Can 2	0,81	1,87	0,141	0,989	0,0842
Can 3	0,37	0,16	0,011	1,000	0,7312

error de la hipótesis, y los niveles de los factores ubicados dentro de este círculo tienen una alta tasa de error.

El análisis se llevó a cabo de dos maneras. En un primer análisis, se utilizaron los residuales (desviaciones de las medias) de los vectores de variables respuesta ajustada, del modelo de efectos ajustados. Aquí, para el modelo de dos vías con interacción, la gráfica HE en rango reducido para el efecto de interacción (P X T), H_{PXT} es equivalente al análisis de $y|(P, T)$ es decir, $y_{ijk} - \bar{y}_{P(i)} - \bar{y}_{T(j)}$. El resultado, se ilustra a través de una gráfica HE canónica que resume toda la variación de las medias en la interacción en la elipse H y donde se grafican las medias (centroides) como puntos.

En un segundo análisis, se codificaron todas las combinaciones de los niveles de los dos factores interactivamente, así que la matriz H expresa las diferencias de todos los grupos; es decir, $H=H_p + H_T + H_{PXT}$, $E=E_w$ y el error dentro de celdas. El resultado, se ilustra a través de una gráfica de la estructura discriminante canónica de todos los grupos o combinaciones de tratamientos.

3. Resultados y discusión

En primer lugar, los valores obtenidos para cada una de las seis variables respuestas estudiadas fueron sometidos a análisis exploratorios de datos. Los resultados demostraron que ninguna variable presentó desvíos significativos con relación a las suposiciones de normalidad y homogeneidad de varianzas, indicando ausencia de limitaciones para la realización del análisis de la varianza de

los datos. Se comprobó también que el vector de errores tiene distribución multinormal, verificado por las pruebas de Mardia (Mardia, 1974) y Henze-Zirkler (Henze y Zirkler, 1990) usando el macro MULTNORM del SAS versión 8 (SAS, 1999), el cual reporta además las pruebas de Shapiro-Wilk (Shapiro y Wilk, 1965), para normalidad univariada y las pruebas multivariadas de Asimetría y Curtosis de Mardia y el estadístico T de Henze-Zirkler.

El análisis multivariante de los datos aplicando el ADC revela, en primera instancia, un efecto altamente significativo para la interacción Plástico Y Tiempo ($P < 0,01$) por el estadístico de prueba multivariante Lambda de Wilks. Los resultados se resumen en la Tabla 1.

La prueba para la interacción P X T basado en el estadístico Lambda de Wilks tiene un valor de 0,025 con $P < 0,01$; lo que indica que el contraste multivariante para explicar la relación entre los efectos de envoltura plástica y el tiempo de almacenamiento es altamente significativo considerando las seis variables simultáneamente. La misma necesita tres dimensiones para ser representada, si bien las dos primeras recogen el 98,9% de la información. De las tres posibles dimensiones para la interacción, solamente la primera es significativa ($P < 0,01$); es decir, la dimensionalidad de la interacción es unidimensional. Esto se ve reflejado también en el hecho que la proporción del autovalor (o proporción de variabilidad explicada) por la primera variable canónica (Can 1) es de 0,848; lo que indica que la primera función canónica representa un 84,8% de la variación total en la interacción.

Tabla 2. Coeficientes de estructura canónica total de las dos primeras funciones canónicas y variación explicada por cada vector variable para la interacción Plástico X Tiempo.

Variable	Estructura Canónica		Variación explicada (R ²)
	Can 1	Can 2	
PP	0,95	-0,10	0,84
FIR	0,88	0,12	0,72
SST	-0,21	0,06	0,14
AT	0,49	0,28	0,28
SSAT	-0,27	-0,22	0,19
pH	-0,04	0,37	0,13

En la Tabla 2, se reportan los coeficientes de la estructura canónica total (también llamada estructura de correlación o pesos discriminantes canónicos) para la interacción Plástico X Tiempo. La matriz de estructura canónica total contiene las correlaciones entre las variables y las funciones discriminantes canónicas, donde se examinan las correlaciones simples entre las variables originales y las funciones canónicas. Estas correlaciones revelan las relaciones lineales bivariadas entre cada una de las variables y los ejes discriminantes canónicos. Así, los pesos canónicos reflejan la varianza que la variable observada comparte con la función canónica, y puede ser interpretada en la evaluación (o valoración) de la contribución relativa de cada variable a cada función canónica. También se muestra la proporción de variabilidad explicada por cada vector variable (R²).

Examinando las correlaciones de las variables con las funciones canónicas (coeficientes de estructura canónica), se puede ver que Can 1 está dominada por un valor grande para PP (0,95), seguido por FIR (0,88) y en menor medida por AT (0,49), siendo todas positivas. La segunda Can 2 está dominada por pH (0,37) seguido por AT (0,28) y SSAT (-0,22), aunque con valores o pesos no tan grandes como el primer eje canónico. Así, es evidente que la interacción (P X T) se debe principalmente a las variables PP, FIR y AT, y por tanto, son responsables de una porción importante de la discriminación entre las 8 combinaciones de tratamiento. La variable PP es la que presenta la mayor proporción de variabilidad explicada (R²) con un valor de 0,84, seguida de

FIR con 0,72. Todo esto se puede ver ilustrada en la Figura 1, donde se resume toda la variación de las medias en la interacción (elipse H) en la que se grafican las medias de cada combinación de tratamiento como puntos (centroides).

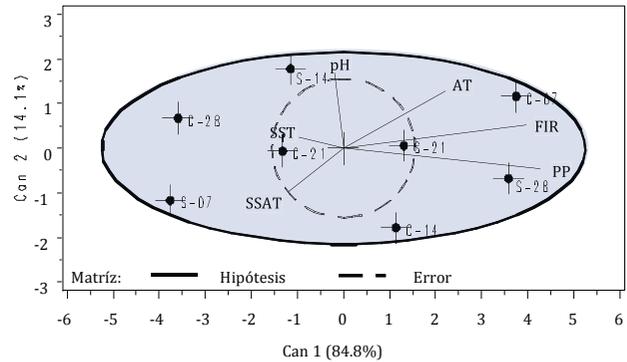


Figura 1. Gráfica Hipótesis-Error para la interacción (P X T).

En esta figura, se ilustra la Gráfica HE discriminante canónica para el efecto de interacción Plástico X Tiempo, en la que se hace una representación gráfica de las matrices H (Hipótesis) y E (Error) de un modelo lineal multivariado para las dos dimensiones canónicas que mejor discriminan entre las combinaciones de tratamientos. Aquí, las medias de las combinaciones de tratamientos son representadas por una elipse H y la variación del error es mostrada explícitamente como una elipse E (la cual es graficada como un círculo en el espacio canónico). El orden y separación de las medias de las combinaciones de tratamientos sobre los dos ejes canónicos indican que la primera combinación lineal (Can 1) discrimina los 8 tratamientos, mientras que la segunda (Can 2), prácticamente no discrimina ninguna, de lo cual se deduce la no significancia de la segunda variable canónica; es decir, básicamente toda la variación entre las combinaciones de tratamientos es explicada por la primera función.

Por otra parte, se puede ver que la mayor parte de la variación entre los tratamientos puede ser explicada por las diferencias entre las combinaciones (S-07) y (C-28) respecto de las combinaciones (C-07) y (S-28), siendo los vectores variables PP y FIR las

Tabla 3. Resumen del análisis discriminante canónico para tratamientos.

Variable canónica	Correlación canónica	Autovalor	Proporción autovalor	Proporción acumulada	Valor de Probabilidad
Tratamientos (Lambda de Wilks = 1,85E-6; P = 0,0001)					
Can 1	0,99	377,22	0,906	0,906	0,0001
Can 2	0,98	30,04	0,072	0,978	0,0001
Can 3	0,93	6,16	0,015	0,993	0,0001
Can 4	0,82	1,98	0,004	0,997	0,0034
Can 5	0,65	0,75	0,002	0,999	0,0570
Can 6	0,44	0,23	0,001	1,000	0,1864

que diferencian ampliamente estos tratamientos; es decir, el 84,8% de la variación de las medias de las 8 combinaciones de tratamientos es explicada por una sola dimensión canónica (Can 1) la cual corresponde a valores grandes para C-07 y S-28 y pequeñas para S-07 y C-28 sobre la variable PP fundamentalmente y en menor medida por FIR. Por otra parte, la variable PP es la que tiene la mayor correlación positiva con Can 1 y la mayor longitud relativa (R^2), lo que indica que este vector variable tiene la mayor proporción de variabilidad explicada (0,84), seguida por FIR (0,72). Resumiendo, las variables PP y FIR, son las que presentan mayor correlación con la primera función canónica y mayor proporción de variabilidad explicada (Ver Tabla 2).

La otra manera de analizar la interacción es considerando las combinaciones de tratamientos codificadas interactivamente. La información se muestra en las Tablas 3 y 4, e ilustrada en la Figura 2 siguientes:

El prueba global para tratamientos basado en el estadístico Lambda de Wilks tiene un valor de 1,85E-6 con $P < 0,01$; indicando que el contraste multivariante para diferenciar entre las 8 combinaciones de tratamiento es altamente significativo. Aunque las primeras cuatro variables canónicas resultan altamente significativas ($P < 0,01$), solo las dos primeras (Can 1 y Can 2) recogen el 97,8% de la información (primer plano factorial bidimensional).

Tabla 4. Coeficientes de estructura canónica total de las dos primeras funciones canónicas y variación explicada por cada vector variable para tratamientos.

Variable	Estructura canónica		Variación Explicada (R^2)
	Can 1	Can 2	
PP	-0,76	0,63	0,98
FIR	0,78	0,58	0,97
SST	-0,68	-0,25	0,82
AT	0,86	0,09	0,96
SSAT	-0,95	-0,16	0,96
pH	-0,65	-0,02	0,84

Examinando las correlaciones de las variables con las funciones canónicas (Tabla 4), se puede ver que Can 1 está dominada por valores grandes para todas las variables, aunque con mayor peso SSAT (-0,95) seguido por AT (0,86) y en menor medida por pH (-0,65). La segunda (Can 2) está dominada por PP (0,63) y FIR (0,58), respectivamente; además, todas las variables presentan altas proporciones de variabilidad explicada (R^2) indicando que todas las variables contribuyen con la discriminación de los tratamientos en un plano bidimensional.

En la Figura 2, se puede ver que en este primer plano factorial (Can 1 vs. Can 2), del 97,8% de la variabilidad total explicada por ambos ejes canónicos, el 90,6% de la variación entre las medias de los tratamientos (centroides) es explicada por la primera dimensión canónica (Can 1), alineada principalmente con las variables SSAT, AT, PP

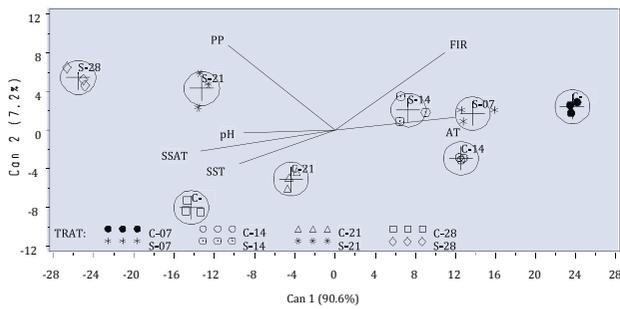


Figura 2. Gráfica de la estructura discriminante canónica mostrando las 8 combinaciones de tratamientos codificadas interactivamente.

y FIR; es decir, la primera función canónica está reflejando el tiempo de almacenamiento, estando los tiempos de 21 y 28 días asociados con altos valores (negativos) de las variables PP y SSAT y los tiempos de 7 y 14 días con altos valores (positivos) de las variables AT y FIR, donde se destaca que el primer eje canónico refleja un contraste entre los frutos con (C) y sin (S) envoltura plástica a los 21 y 28 días de almacenamiento asociado con altos valores (negativos) de las variables SSAT (-0,95) y PP (-0,76) respecto de los días 7 y 14 asociado con altos valores (positivos) de las variables AT (0,86) y FIR (0,78).

La segunda dimensión canónica Can 2 explica un 7,2% de la variación, determinada principalmente por las variables PP (0,63) y FIR (0,58); donde se puede deducir que Can 2 refleja el efecto de envoltura plástica, siendo más marcada esta diferencia en los días 21 y 28 sobre PP. Es decir, se refleja un contraste entre los frutos con cubierta plástica (C-21, C-28) respecto de los frutos sin cubierta plástica (S-21, S-28) a los 21 y 28 días, respectivamente, sobre la variable PP y entre los frutos con cubierta plástica a los 21 y 28 días (C-21, C-28) respecto de los frutos con y sin cubierta plástica a los 7 y 14 días (C-07, C-14, S-07, S-14) sobre la variable FIR. En otras palabras, la firmeza de los frutos está asociada con los primeros 14 días y la pérdida de peso con últimos 14 días de almacenamiento.

Se destaca también que todos los vectores variables presentan una alta proporción de variabilidad explicada (R^2), siendo dominada principalmente

por las variables PP y FIR con 0,98 y 0,97; respectivamente. Esto se ve reflejado también en el hecho que los vectores variable PP y FIR separan claramente los frutos con y sin envoltura plástica a los 21 y 28 días (coordenadas negativas) y los frutos con y sin envoltura plástica a los 7 y 14 días (coordenadas positivas) porque al proyectar sus círculos de confianza sobre el vector que representa a estas variables las proyecciones son disjuntas. Igual razonamiento se puede hacer para el efecto de envoltura plástica en cada tiempo de almacenamiento.

En síntesis, los resultados muestran que el análisis discriminante canónico con las técnicas gráficas Hipótesis-Error y de Estructura Discriminante Canónica usadas conjuntamente, permite identificar relaciones importantes entre las variables y los niveles de los factores evaluados de una manera compacta. Específicamente, se pudo diferenciar claramente las combinaciones de tratamientos en la interacción. Las ocho combinaciones de tratamiento aparecen gráficamente bien diferenciados. Se identificaron las variables más importantes en la discriminación de las combinaciones de tratamiento (interacción). Las diferencias entre los niveles de los factores están básicamente sobre el primer eje canónico.

Se pone de manifiesto que, con los datos del experimento presentado, se ha mostrado un procedimiento alternativo para analizar e interpretar resultados significativos obtenidos al utilizar diseños con dos factores, como análisis subsiguientes. En este sentido, el análisis discriminante canónico como técnica estadística multivariada de reducción de la dimensión y estrechamente relacionada al análisis de varianza multivariado, posibilita estudiar la estructura de varios grupos de individuos con relación a una serie de variables medidas dentro de esos grupos y permitir un mejor entendimiento de las relaciones entre las variables con la condición de que la separación entre los distintos grupos sea máxima con respecto a la variabilidad dentro de los grupos.

Las dos técnicas gráficas descritas e ilustradas en este trabajo y desplegadas para un diseño con dos

factores multivariante, proporcionan maneras de visualizar y entender adecuadamente los resultados de pruebas multivariantes. Específicamente, la gráfica HE canónica muestra directamente como las medias de los grupos difieren sobre las dos funciones canónicas conjuntamente, en relación a la variación dentro de grupos. La gráfica de estructura discriminante canónica muestra una vista alternativa de todas las variables, en un espacio bidimensional que maximiza las diferencias entre los grupos. La superposición de vectores variables, círculos de confianza para las medias de los tratamientos y anotaciones canónicas en rango reducido, ayudan a hacer estas gráficas más interpretables en relación a los datos originales y el resumen en baja dimensión extraída de tales gráficas.

4. Conclusiones

El análisis e interpretación de los efectos significativos obtenidos en diseños experimentales con dos factores y la identificación de las variables importantes para la discriminación de los niveles de los mismos, es posible por gráficas multivariantes con despliegue en espacios de baja dimensión, como las gráficas HE y de estructura discriminantes canónica que permiten un resumen visual compacto de los resultados y proveen una base para la visualización de todas las variables conjuntamente en una sola gráfica.

El procedimiento puede ser visto como una extensión del análisis de varianza multivariado, cuyo énfasis está en la reducción de la dimensión, y que casi siempre es aplicado a diseños de una vía, es posible extenderlo a diseños de dos vías ofreciendo una buena alternativa para el análisis de datos en diseños con dos factores. Se pone de manifiesto con la aplicación experimental las bondades que estas técnicas gráficas multivariadas poseen en el análisis e interpretación de resultados significativos surgidos de modelos de diseños experimentales complejos.

Referencias

1. Afifi, A.A. and V. Clark. (1996). *Computer-Aided Multivariate Analysis. 3ed.* Chapman & Hall, London.
2. Amaro, R., J.L. Vicente Villardón, y M.P. Galindo Villardón. (2004). *MANOVA-Biplot para arreglos de tratamientos con dos factores basado en modelos lineales generales multivariantes.* Interciencia, 29(1), 26-32.
3. Amaro, R. (2001). *MANOVA-Biplot para diseños con varios factores basado en modelos lineales generales multivariantes.* Tesis Doctoral. Universidad de Salamanca. España.
4. Cruz Castillo, J.G., S. Ganeshanandam, B.R. Mackay, G.S. Lawes, C.R.O. Lawoko and D.J. Woolley (1994). *Applications of canonical discriminant analysis.* Hortscience, 29, 1115-1119.
5. Cuadras, C.M. (1996). *Métodos de Análisis Multivariante.* EUNIBAR, Barcelona, España.
6. Cuadras, C.M. (2007). *Nuevos Métodos de Análisis Multivariante.* CMC Editions, Barcelona, España.
7. Egesel, C.O., F. Kahnman and M.K. Güi (2011). *Discrimination of maize inbreds for kernel quality traits and fatty acid composition by a multivariate technique.* Acta Scientiarum Agronomy. Maringá, 33(4), 613-620.
8. Fisher, R.A. (1936). *The use of multiple measurements in taxonomic problems.* *Annals of Eugenics*, 7, 179-188.
9. Friendly, M. (2006). *Data Ellipses, HE Plots and Reduced-Rank Displays for Multivariate Linear Models: SAS Software and Examples.* *Journal of Statistical Software*, 17(6), 1-43.
10. Friendly, M. (2007). *HE Plots for Multivariate General Linear Models.* *Journal of Computational and Graphical Statistics*, 16, 1-23.
11. Friendly, M., G. Monette and J. Fox (2013). *'Elliptical insights: Understanding statistical methods through elliptical geometry'.* *Statistical Science* 28(1), 1-39.
12. Friendly, M. and .M. Sigal (2014). *Recent Advances in Visualizing Multivariate Linear Models.* *Revista Colombiana de Estadística*, 37(2), pp. 261-283.

13. Gittins, R. (1985). *Canonical Analysis: A Review with Applications* in Ecology, Berlin: Springer-Verlag.
14. Gower, J.C. (1989). *Generalized canonical analysis. In: Multivay Data Analysis.* (R. Coppi y Bolasco, eds.). Elsevier Science Publisher, New York, pp. 221-232.
15. Henze, N. and Zirkler, B. (1990). A Class of Invariant Consistent tests for Multivariate Normality. *Communications in Statistics, Part A - Theory and Methods*, 19(10), 3595-3617.
16. Krzanowski, W.J. (1988). *Principles of Multivariate Analysis: A User's Perspective.* Oxford: Clarendon Press.
17. Lebart, L., A. Morineau y M. Piron. (1995). *Statistique exploratoire multidimensionnelle.* Dunod, París.
18. Mardia, K. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B*, 36, 115-128.
19. Onofri A. e E. Ciricofolo. (2004). *Characterization of yield quality in durum wheat by canonical variate analysis.* Proceedings VIII ESA Congress "European Agriculture in a global context", Copenhagen, 11-15 July 2004, 541-542.
20. Statistical Analysis System (SAS) Institute (1999). *SAS/STAT User's Guide, Version 8,* Cary, NC. USA.
21. Seber, G.A.F. (1984). *Multivariate Observations.* John Wiley & Sons, Inc. New York.
22. Shapiro, S.S. and M.B. Wilk. (1965). *An Analysis of Variance Test for Normality* (complete samples). *Biometrika*, 52.