APLICACIÓN DEL ANALISIS DISCRIMINANTE PARA EXPLORAR LA RELACIÓN ENTRE EL EXAMEN DE ICFES Y EL RENDIMIENTO EN ALGEBRA LINEAL DE LOS ESTUDIANTES DE INGENIERÍA DE LA UTP EN EL PERÍODO 2001-2003

PATRICIA CARVAJAL
Candidato a Magister en Investigación Operativa y Estadística
Universidad Tecnológica de Pereira-UTP
pacarva@utp.edu.co

ALVARO TREJOS
Candidato a Magister en Investigación Operativa y Estadística,
Profesor
Facultad de Ingeniería Industrial
Universidad Tecnológica de Pereira, UTP
alvarot@utp.edu.co

JOSÉ SOTO MEJÍA
Profesor Titular PhD.
Facultad de Ingeniería Industrial
Universidad Tecnológica de Pereira
jomejia@utp.edu.co

RESUMEN
El presente articulo explora la relación entre los puntajes de ICFES [1] en cada una de las áreas del conocimiento con el rendimiento en álgebra lineal por parte de los estudiantes de ingenierías de la Universidad Tecnológica de Pereira, UTP, a través de la técnica estadística multivariada conocida como análisis discriminante. Los datos históricos que sirven como base del análisis han sido proporcionados por la Oficina de Planeación de la UTP. Para el análisis de los datos se utilizó el software estadístico SPSS (Statistical Package for Social Sciences, versión 11.5)
El análisis estadístico multivariado desarrollado en éste artículo permite concluir que los puntajes obtenidos en las pruebas del ICFES, como único criterio de ingreso a la UTP, dejan serios cuestionamientos sobre su capacidad de prever el desempeño futuro del estudiante en los programas de ingeniería.

PALABRAS CLAVES: Análisis Discriminante, Criterio de Lamba Wilks, Correlación Canónica, Función Discriminante, ICFES

ABSTRACT
The present I articulate explores the relation between the grades of ICFES [1] in each one of the areas of the knowledge with the yield in linear algebra on the part of the students of engineerings of the Technological University of Pereira, UTP, through the multivaried statistical technique of discriminate analysis.

The historical data that serve as it bases of the analysis have been provided by the Office of Planning of the UTP. For the analysis of the data statistical software SPSS (Statistical for Package Social Sciences was used, version 11.5) the developed multivaried statistical analysis in this one article allows to conclude that the grades obtained in the tests of the ICFES, like only criterion of entrance to the UTP, they leave serious doubts on its capacity to anticipate the future academic performance of the student in the engineering programs

KEYWORDS: *Discriminate Analysis, Lambda Wilks, Canonical Correlation, Discriminate Function, ICFES.*

1. INTRODUCCIÓN

Uno de los requisitos para ingresar a un programa de ingenierías en la Universidad Tecnológica de Pereira (UTP) es el resultado del examen de ICFES[1], por áreas del conocimiento. El artículo 14 del nuevo reglamento estudiantil de la U.T.P (Acuerdo No 28 del 16 de diciembre del 2003)[2] establece una ponderación para cada una de las áreas del examen ICFES, dependiendo del programa académico al cual se aspira ingresar. Para el caso de las ingenierías los mayores pesos corresponden a física, matemática y lenguaje con una ponderación del 20% para cada una de ellas. De ésta manera, implícitamente se está asumiendo que aspirantes a ingresar a los programas de ingeniería, que hayan obtenido alto puntaje en el examen del ICFES, tendrán también buen rendimiento académico en las asignaturas propias del programa de ingeniería. Por tanto, se espera que el rendimiento académico del estudiante de ingeniería, en asignaturas tales como: álgebra lineal, física, matemáticas, etc. sea explicado por el desempeño del mismo en las pruebas de ICFES.

En la práctica, no todos los estudiantes que ingresan a los programas de ingeniería tienen buen rendimiento académico en las asignaturas enunciadas. A manera de ejemplo, hay una parte de ellos que pierde la asignatura Álgebra Lineal. Podría ser de interés poder predecir, de acuerdo con los resultados del examen de ICFES, si un aspirante a ingresar a un programa de ingeniería tendrá

[1] ICFES. Instituto Colombiano de Fomento para la Educación Superior, encargado de efectuar los Exámenes de Estado para el ingreso a la Educación superior
[2] Acuerdo 18 de diciembre de 2003 emanado del Consejo Superior de la UTP que reglamenta los derechos y deberes de sus estudiantes.

Fecha de recepción:
Fecha de aceptación:

éxito académico o no al cursar la asignatura Álgebra Lineal.

En términos simples, el análisis discriminante es una técnica estadística multivariada de especial interés en la investigación puesto que permite establecer la pertenencia probabilística de un individuo a grupos previamente establecidos a través de funciones de discriminación,

En éste artículo se presentan los resultados del intento de generar a través del Análisis Discriminante una función de discriminación que permita predecir si un estudiante que ingresa a las ingenierías de la U.T.P, tendrá éxito o no en la asignatura Álgebra Lineal. La validez de la función discriminante dependerá de su efectividad a la hora de clasificar a un estudiante de forma correcta en uno de los dos grupos (los que ganarán o los que perderán Álgebra Lineal). Para lo cual habrá que esperar hasta que el estudiante haya efectivamente cursado la asignatura.
La sección 2, El Análisis Discriminante, presenta un resumen de los elementos teóricos básicos del análisis discriminante, basado en [4]. La sección 3, Obtención de la Función Discriminante, presenta los datos y el procedimiento seguido al intentar obtener una función discriminante usando el paquete estadístico SPPS (versión 11.5)
Finalmente, la sección 4, Resultados y Conclusiones, evalúa y analiza los resultados obtenidos.

2. EL ANALISIS DISCRIMINANTE

El análisis discriminante parte de una tabla de datos de n individuos a los que se les ha medido p variables cuantitativas independientes, que actúan como perfil de características de cada uno de ellos. Una variable cualitativa adicional, dependiente o clasificatoria, con dos ó más categorías, define el grupo al que cada individuo pertenece. Es pues una tabla nx(p+1) en que cada caso figura con un perfil y una asignación a grupo.

A partir la tabla nx(p+1) se obtendrá un modelo matemático discriminante contra el cual será contrastado el perfil del nuevo individuo cuyo grupo se desconoce para, en función de un resultado numérico, ser asignado al grupo más probable.

Cuando la clasificación de individuos se pueda asignar solamente a dos grupos a partir de k variables clasificadoras, solo se requiere una función discriminante.
La función discriminante de Fisher D, que aparece en la ecuación (1), se obtiene como función lineal de k variables explicativas como:

$$D = u_1 X_1 + u_2 X_2 + ... + u_k X_k \qquad (1)$$

Las puntuaciones discriminantes son los valores que se obtienen al dar valores a X_1, X_2, …, X_k en la ecuación (1). El problema matemático consiste en obtener los coeficientes de ponderación u_j, que mejor permitan diferenciar los individuos de un grupo del otro. Si consideramos que existen n observaciones, se puede expresar la función discriminante como en la ecuación (2)

$$D_i = u_1 X_{1i} + u_2 X_{2i} + ... + u_k X_{ki} \quad i = 1,2,...,n \qquad (2)$$

D_i es la puntuación discriminante correspondiente a la observación i-ésima. Expresando las variables explicativas en desviaciones respecto a la media, la relación anterior se puede expresar en forma matricial como sigue en la ecuación (3):

$$\begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} ... X_{k1} \\ X_{12} & X_{22} ... X_{k2} \\ & \\ & \\ X_{1n} & X_{2n} ... X_{kn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_k \end{bmatrix} \qquad (3)$$

En notación compacta se puede escribir como en la ecuación matricial (4) .

D=Xu (4)

La variabilidad de la función discriminante (suma de cuadrados de las variables discriminantes respecto a su media) se expresa en la ecuación (5) como:

d'd=u'X'Xu (5)

La matriz **X'X** es una matriz simétrica expresada en desviaciones respecto a la media, por lo que puede considerarse como la matriz T de suma de cuadrados total de las variables explicativas de la matriz X. Según la teoría del análisis multivariante de la varianza, **X'X** se puede descomponer en la suma de la matriz entre grupos **F** y la matriz intragrupos **V** (o residual), como se muestra en la ecuación (6).

X'X=T=F+V (6)

Por tanto (5) puede expresarse como en (7)

d'd= u'X'Xu= u'Tu= u'Fu= u'Wu (7)

Los ejes discriminantes vendrán dados por los vectores propios asociados a los valores propios de la matriz $W^{-1}F$ ordenados de mayor a menor. Las puntuaciones discriminantes se corresponden con los valores obtenidos al proyectar cada punto del espacio k-dimensional de las variables originales sobre el eje discriminante.

Los centros de gravedad o centroides (vector de medias) son los estadísticos básicos que resumen la información sobre los grupos. Los centroides de los grupos I y II se muestran en la ecuación (8)

$$\overline{X}_I = \begin{bmatrix} \overline{X}_{1,I} \\ \overline{X}_{2,I} \\ \overline{X}_{k,I} \end{bmatrix} \qquad \overline{X}_{II} = \begin{bmatrix} \overline{X}_{1,II} \\ \overline{X}_{2,II} \\ \overline{X}_{k,II} \end{bmatrix} \tag{8}$$

Así, para los grupos I y II se obtienen los puntajes promedios de la función discriminante en cada grupo dados por la ecuación (9) :

$$\overline{D}_I = u_1\overline{X}_{1,I} + u_2\overline{X}_{2,I} + \ldots + u_k\overline{X}_{k;I} \tag{9}$$
$$\overline{D}_{II} = u_1\overline{X}_{1,II} + u_2\overline{X}_{2,II} + \ldots + u_k\overline{X}_{k;II}$$

El punto de corte discriminante C se calcula mediante la ecuación (10)

$$C = \frac{\overline{D}_I + \overline{D}_{II}}{2} \tag{10}$$

El criterio para clasificar el individuo i es el siguiente:

Si $D_i<C$, se clasifica al individuo i en el grupo I
Si $D_{i>}C$, se clasifica al individuo i en el grupo II

En general, cuando se aplica el análisis discriminante se le resta el valor de C a la función discriminante, como se expresa en la ecuación (11).

$$D - C = u_1X_1 + u_2X_2 + \ldots + u_kX_k - C \tag{11}$$

En este caso, se clasifica a un individuo en el grupo I si D-C>0, y en grupo II de otra manera.

Existen otros criterios de clasificación, entre los que se destacan el análisis de regresión y la distancia de Mahalanobis[3].

3. OBTENCION DE LA FUNCIÓN DISCRIMINANTE

Al momento de ingresar a la UTP a un estudiante de ingenierías, se le conocen algunas variables numéricas, de carácter académico como son: resultados del ICFES total y resultados para las áreas de biología, matemáticas, filosofía, historia, química, lenguaje, geografía. De acuerdo con el modelo de la sección 2, este conjunto de nueve variables corresponden a las variables independientes del modelo. Al cabo del 3er o 4º semestre de la carrera se conocerá la nota de álgebra lineal, la cual permitirá ubicar al estudiante en uno de dos grupos: ganadores o perdedores de la asignatura Álgebra Lineal. La variable de ubicación en los grupos es la variable dependiente o clasificatoria mencionada en la sección 2 y es de tipo cualitativa dicotómica por que tiene dos categorías mutuamente excluyentes. Con base en la información descrita anteriormente sería importante conocer lo siguiente:
¿Se puede usar el conjunto de las nueve variables independientes y la variable dependiente de carácter cualitativa, para construir una función matemática y estadística que permita clasificar los nuevos estudiantes en uno de los dos grupos establecidos por la variable cualitativa?
¿Que forma tiene la función – cuáles son sus parámetros – puede utilizarse esta función para clasificar nuevos estudiantes?
¿Cuáles son las posibilidades de no cometer error al aplicar la regla o función de decisión?
y para terminar la aplicación del análisis discriminante a manera de ejemplo se analizarán los resultados restantes del procesamiento:

3.1. Programa de SPSS para ejecutar el análisis discriminante

Con la base de datos mencionada, que contiene 15 variables de tipo académico para todos los estudiantes de ingeniería en el período 2000-2003. Se ejecutaron los comandos del programa SPSS necesarios para ejecutar [3] el procedimiento discriminante, así:

ANÁLISIS – CLASIFICAR – DISCRIMINANTE
VARIABLE DE AGRUPACION REPITIO ALGEBRA LINEAL (0-1)
INDEPENDIENTES: Icfes Total, Icfes matemática, Icfes Biología, Icfes filosofía, Icfes física, Icfes Historia, Icfes Química, Icfes Lenguaje, Icfes geografía

USAR Método de inclusión por pasos
CLASIFICAR

PROBABILIDADES PREVIAS: calcular según tamaño de grupo
MOSTRAR: resultados para los primeros 20 casos
TABLA DE RESUMEN CONTINUAR
SELECCIONAR VARIABLE DE SELECCIÓN FILTRO=1, ACEPTAR

3.2 Estadísticas Descriptivas Para Las Variables Independientes

Después de ejecutar en el software estadístico SPSS-versión 11.5 los comandos pertinentes al análisis discriminante las estadísticas descriptivas para las variables independientes correspondientes a los 68

[3] Distancia de Mahalanobis es una generalización de la distancia euclidiana que tiene en cuenta la matriz de covarianza de las variables

estudiantes que perdieron álgebra lineal y a los 461 estudiantes que la ganaron se presentan a en la Tabla 1. En la parte superior de la Tabla 1 se presentan los resultados para el grupo de estudiantes que no repitió Álgebra Lineal (461). La primera columna muestra el puntaje total del ICFES y el puntaje para cada una de las 8 asignaturas. En la segunda columna la media de cada una de las asignaturas que serán usadas como variables predictoras. En la mitad de la tabla aparece la misma información para los 68 estudiantes que repitieron álgebra lineal. En la parte inferior de la tabla están los totales sin dividir entre los que pierden o ganan álgebra lineal.

Estadísticos de grupo

REPETICION DE ALGEBRA LINEAL		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
NO REPITIO	PUNTAJE ICFES TOTAL	51,2407	3,83310	461	461,000
	ICF_Biologia	51,6963	5,48387	461	461,000
	ICF_Matematic	47,7918	6,21604	461	461,000
	ICF_Filosofia	49,5249	6,12390	461	461,000
	ICF_Fisica	51,5510	6,17359	461	461,000
	ICF_Historia	50,2278	5,33957	461	461,000
	ICF_Quimica	52,6312	5,78704	461	461,000
	ICF_Lenguaje	53,8525	5,55055	461	461,000
	ICF_Geografia	50,6941	5,38675	461	461,000
SI REPITIO	PUNTAJE ICFES TOTAL	49,6119	2,86202	68	68,000
	ICF_Biologia	50,9853	4,51364	68	68,000
	ICF_Matematic	45,5588	4,83576	68	68,000
	ICF_Filosofia	47,7794	5,48769	68	68,000
	ICF_Fisica	49,8088	5,89272	68	68,000
	ICF_Historia	50,0294	4,53501	68	68,000
	ICF_Quimica	51,9559	4,90639	68	68,000
	ICF_Lenguaje	53,2206	5,09564	68	68,000
	ICF_Geografia	50,2206	4,46828	68	68,000
Total	PUNTAJE ICFES TOTAL	51,0314	3,76000	529	529,000
	ICF_Biologia	51,6049	5,37045	529	529,000
	ICF_Matematic	47,5047	6,09835	529	529,000
	ICF_Filosofia	49,3006	6,06924	529	529,000
	ICF_Fisica	51,3270	6,16048	529	529,000
	ICF_Historia	50,2023	5,23959	529	529,000
	ICF_Quimica	52,5444	5,68178	529	529,000
	ICF_Lenguaje	53,7713	5,49368	529	529,000
	ICF_Geografia	50,6333	5,27624	529	529,000

Tabla 1. Estadísticos de grupo

Los 529 estudiantes sobre los que se realiza el análisis discriminante constituyen una muestra aleatoria del total de 741 estudiantes registrados en el período 2000-2003 y que cursaron álgebra lineal. La decisión de tomar una muestra aleatoria de 529 estudiantes es la de construir a partir de ella la función discriminante y posteriormente poder probar la calidad de la función obtenida con el resto de estudiantes no seleccionados (212).

3.3 selección de las variables discriminantes

En el análisis discriminante debe seleccionarse de aquel conjunto de variables independientes (PUNTAJE ICFES TOTAL e ICFES POR ÁREAS) aquel subconjunto que más discrimine a los grupos establecidos por la variable dependiente o clasificatoria (perdió o ganó álgebra lineal).

Los criterios que se consideran para la selección de las variables son el de Lambda Wilks (λ) y el estadístico F.
Lambda de Wilks para un conjunto de P variables independientes (P=9 en este trabajo), mide las desviaciones dentro de cada grupo respecto a las desviaciones totales sin distinguir los grupos y está dado por la ecuación (12).

$$l = \frac{S.C\,\mathrm{int}\,ragrupos}{S.C.totales} \qquad (12)$$

Donde en el numerador se encuentra la suma de cuadrados intra grupos, la cual mide el cuadrado de las desviaciones de cada uno de los datos con respecto a la media del grupo, llamado en la sección 2 centroide. En el denominador se encuentra la suma de cuadrados totales, la cual mide el cuadrado de las desviaciones de cada uno de los datos con respecto a la media de todos los datos (media total sin discriminar grupos).
La suma de cuadrados totales es el resultado de la suma de cuadrados inter-grupos (entre los grupos) y intra-grupos (dentro de los grupos),

Si el valor de λ es próximo a 1, los grupos estarán mezclados y el conjunto de variables independientes no será adecuado para construir la función discriminante, dado que el mayor porcentaje de variabilidad estaría siendo representado por la variabilidad dentro de los grupos.

El criterio Lambda de Wilks identifica las mejores variables discriminantes pero es el estadístico F el que determina que variables se deben tener en cuenta en el modelo. El criterio funciona de la siguiente manera, si el valor de F es mayor a 3.84 la variable podrá ser tenida en cuenta, de lo contrario será eliminada. El estadístico F como función del Lambda de Wilks se calcula de acuerdo con la ecuación (13):

$$F = \left(\frac{n-g-p}{g-1}\right)\left(\frac{1-l_{p+1}/l_p}{l_{p+1}/l_p}\right) \qquad (13)$$

donde:
n: es el número de estudiantes; en nuestro caso, n= 529
g: es el número de grupos, g=2 para este caso
p: el número de variables independientes; p=9
λp: Lambda de Wilks calculado antes de la inclusión de variable seleccionada

l_{p+1}: Lambda de Wilks calculado después de la inclusión de variable seleccionada

La Tabla 2, muestra los pasos para la selección de las variables a ser incluidas en la función discriminante siguiendo el criterio del Llambda de Wilks y el estadístico F.

Variables no incluidas en el análisis

Paso		Tolerancia	Tolerancia mín.	F para introducir	Lambda de Wilks
0	PUNTAJE ICFES TOTAL	1,000	1,000	11,338	,979
	ICF_Biologia	1,000	1,000	1,039	,998
	ICF_Matematic	1,000	1,000	8,051	,985
	ICF_Filosofia	1,000	1,000	4,938	,991
	ICF_Fisica	1,000	1,000	4,773	,991
	ICF_Historia	1,000	1,000	,085	1,000
	ICF_Quimica	1,000	1,000	,837	,998
	ICF_Lenguaje	1,000	1,000	,784	,999
	ICF_Geografia	1,000	1,000	,477	,999
1	ICF_Biologia	,547	,547	2,784	,974
	ICF_Matematic	,681	,681	1,255	,977
	ICF_Filosofia	,848	,848	,955	,977
	ICF_Fisica	,664	,664	,080	,979
	ICF_Historia	,724	,724	2,943	,973
	ICF_Quimica	,564	,564	2,961	,973
	ICF_Lenguaje	,629	,629	2,114	,975
	ICF_Geografia	,830	,830	,577	,978

Tabla 2. Variables no incluidas en el análisis

La Tabla 2 está dividida en dos pasos, en el paso cero, la variable con menor Lambda de Wilks (ver quinta columna) corresponde al Puntaje Total del ICFES (ver primera columna) con ë=0.979 y un valor del estadístico F=11.338 (ver cuarta columna).

Este resultado indica que el Puntaje ICFES total, explica muy poco la pertenencia de un estudiante en el grupo de los que ganan ó pierden álgebra lineal, ello se debe al hecho de tener un valor ë próximo a 1. Pero la variable debe ser tenida en cuenta en el modelo discriminante por ser estadísticamente significativa, F>3.84.

Como se observa en la Tabla 2, en el paso 1 no se tiene en cuenta para el análisis la variable (Puntaje ICFES total) ya que fue extraída en el paso 0. En el paso 1 la variable de menor Lambda de Wilks es el puntaje de ICFES en Historia y en Química (0.973 para ambas) pero no obstante no deben incluirse en la función discriminante por no cumplir el criterio de inclusión dado por el estadístico F definido anteriormente.

3.4. Determinación de la Función Discriminante

Se tiene n=529 estudiantes en la población y p variables independientes (p=9), cada estudiante puede ser considerado como un punto en un espacio p dimensional. Las coordenadas de cada punto se obtendrán a partir de las nueve variables independientes (puntajes ICFES).

A partir de la representación de los n=529 estudiantes, en R^9, se trata de extraer un nuevo espacio de menor dimensión [2] (Þ) tal que, al proyectar la nube de puntos sobre dicho espacio los puntos correspondientes a estudiantes pertenecientes a un mismo grupo estén próximos entre sí y los correspondientes a alumnos pertenecientes a distintos grupos estén alejados. Los ejes de este nuevo espacio serán las funciones discriminantes. El número de funciones discriminantes corresponde al mínimo entre el número de grupos menos uno y el número de variables independientes como expresado en la expresión (14):

$$\text{Min}(g-1,p) \qquad (14)$$

Donde g es el número de grupos y p es el número de variables.Dado que se tienen dos grupos sólo se tendrá en cuenta una función discriminante con una sola variable independiente (Puntaje ICFES total). La forma de la función es la mostrada en la ecuación (15).

$$Di = \hat{a}o + \hat{a}_1 X_i \qquad (15)$$

Donde

Di: es la puntuación discriminante para el estudiante i

Xi: es el puntaje de ICFES para el individuo i

âo, $\hat{a}_1$ son los coeficientes estimados a partir de los datos, de modo que los valores calculados por la función discriminante cumpla los siguientes características:

a) Maximice los autovalores (valores propios de las funciones canónicas discriminantes). Donde los autovaloeres están definidos como en la ecuación (16)

$$autovalores = \frac{S.C.entre\ grupos}{S.C.\text{int}\ ragrupos} \qquad (16)$$

De la ecuación (16), queda claro que un autovalor grande implica que la suma de cuadrados entre grupos es grande comparado con la suma de cuadrados intragrupos. Lo anterior significa que la función estaría discriminando los grupos.

b) Minimice Lambda de Wilks (ë) como se describe en la ecuación (17)

$$l = \frac{S.C.\text{int}\ ragrupos}{S.C.totales} \qquad (17)$$

c) Maximice la correlación canónica (Þ) como se describe en la ecuación (18)

$$h = \sqrt{\frac{S.C.entre\ grupos}{S.C.totales}} \qquad (18)$$

El estadístico de Lambda Wilks también puede ser utilizado para escoger las mejores funciones discriminantes.

En la Tabla 3, se presentan la información para probar la significancia ("calidad") de la función discriminante obtenida:

Lambda de Wilks

Paso	Número de variables	Lambda	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	1	,979	1	1	527	11,338	1	527,0	,001

Tabla 3. Prueba estadística de la función canónica discriminante

Donde gl1=1, es el número de funciones discriminantes, gl2=1, es número de variables en la función discriminante, en éste caso corresponde a la variable- Puntaje Icfes total y gl3=527, corresponde es el tamaño de la muestra (529) menos el número de grupos, 2..

La Tabla 4, muestra la información que permite medir la calidad de la función discriminante

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	,022[a]	100,0	100,0	,145

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Tabla 4. Autovalores y Correlación Canónica

En la Tabla 4 aparecen los autovalores (0.022) y la correlación canónica (0.145) que se encuentran próximos a cero, estos son síntomas claros que la función discriminante no logra diferenciar la media de los dos grupos (estudiantes que perdieron o ganaron álgebra lineal), y ello indica que los datos están poco diferenciados en los dos grupos de acuerdo con las puntuaciones discriminantes.

Aun así en la Tabla 5, se presenta los coeficientes estimados para la función discriminante.

Coeficientes de las funciones canónicas discriminantes

	Función
	1
PUNTAJE ICFES TOTAL	,269
(Constante)	-13,704

Coeficientes no tipificados

Tabla 5. Coeficientes de las funciones canónicas discriminantes

La siguiente ecuación (19) permite calcular la puntuación discriminante para el estudiante i.

$D_i = -13.704 + 0.269X_i$, donde, (19)

D_i: puntuación discriminante para el estudiante i
X_i: puntaje de ICFES total del estudiante i

$\hat{A}o = -13.704$ y $\hat{a}_1 = 0.269$

4. CONCLUSIONES Y TRABAJOS FUTUROS

Entre las variables independientes ICFES total, Icfes en biología, matemáticas, filosofía, física, historia, química, lenguaje, geografía; la menos mala (según los criterios de Lambda de Wilks , autovalores, y la correlación canónica (Þ)) es el Puntaje ICFES total, la que aunque es estadísticamente significativa para la clasificación tiene poca significancia práctica de acuerdo con los resultados del criterio de Lambda Wilks.

El análisis estadístico multivariado desarrollado en éste artículo permite concluir que los puntajes obtenidos en las pruebas del ICFES, como único criterio de ingreso a la UTP, dejan serios cuestionamientos sobre su capacidad de prever el desempeño futuro del estudiante en los programas de ingeniería. Dado que el análisis se centró en el desempeño en Álgebra Lineal, queda para futuros estudios establecer la generalidad del cuestionamiento anteriormente mencionado.

En caso de corroborarse en futuros estudios la poca efectividad de los puntajes del ICFES para predecir el rendimiento académico del postulante a ingresar a la UTP, no debería ser utilizado como único criterio de selección.

De seguir con este criterio podría estarse negando talvez la posibilidad de ingreso de algunos postulantes con bajos puntajes en el ICFES y que podrían tener un buen rendimiento al no estar este último asociado con los puntajes del ICFES

5. BIBLIOGRAFÍA

[1] Dallas E. Johnson, Métodos multivariados aplicados al análisis de datos. Thompson 1998

[2] Richard A. Jonhson – Dean W. Wichern, Applied Multivariate Statistical Analysis. Cuarta edición. Prentice may

[3] Ferrán Aranaz Magdalena. SPSS para Windows Análisis estadístico Mc Graw Hill. 2001

[4] Perez, Cesar. Técnicas estadísticas con SPSS 10. Prentice Hall 2001. Capítulo 13