

DISCRIMINACIÓN POR RENDIMIENTO CURRICULAR: Aplicación al Plan 90 de la Facultad de Ciencias Económicas y de Administración.

Jorge Blanco, Alina Machado, Laura Nalbarte*

Instituto de Estadística, Facultad de Ciencias Económicas y de Administración,
Universidad de la República.

* Los tres investigadores pertenecen al área de Análisis Multivariado del Instituto de Estadística.

Resumen

En el marco de la evaluación de la enseñanza universitaria, el Instituto de Estadística de la Facultad de Ciencias Económicas se encuentra abocado al estudio del Plan 90 de la Facultad en sus distintos aspectos: rendimiento curricular de los estudiantes, duración, deserción, entre otras.

En esta investigación se aplican diversas técnicas de Análisis Multivariado, con el objetivo de encontrar una distribución de rendimiento curricular para los estudiantes de las generaciones 1992 a 1995, de la Facultad de Ciencias Económicas y de Administración.

Se presenta el modelo obtenido para aquellos que optaron por la carrera de Contador Público y obtuvieron el título intermedio. El modelo que permite discriminar los estudiantes según su rendimiento, surge de la aplicación de funciones discriminantes generalizadas, utilizando como variables aquellas que surgen del Análisis de Componentes Principales y los grupos que surgen del Análisis de Cluster.

Abstract

In the framework of the evaluation of education at the University level the Statistics Institute of the School of Economic Sciences is reviewing different aspects of the "Plan 90": students' curricular performance, duration of studies, students' desertion, among others.

In this investigation we applied diverse techniques of Multivariate Analysis, aiming to find a distribution of curricular performance for the students enrolled on the School of Economic Sciences and Administration between 1992 and 1995. The model obtained corresponds to the students that opted for the orientation of Public Accountant and obtained the intermediate title.

The model presented in this study, that allows us to discriminate the students according to its performance, arises from the application of generalized discriminant functions, using the variables coming from the Analysis of Principal Components and the groups coming from the Analysis of Cluster.

Introducción

El objetivo de este trabajo es encontrar una distribución de rendimiento curricular para los estudiantes de las distintas generaciones de la Facultad de Ciencias Económicas y de Administración (FCEA). Se trata fundamentalmente de encontrar criterios que tengan en cuenta algo más que la escolaridad “tradicional” dentro de la generación a la que pertenece el estudiante. A tales efectos se aplicaron Componentes Principales, Cluster y Discriminante Logístico. Se concluye la investigación aplicando funciones discriminantes logísticas generalizadas, dado que no se puede aplicar un modelo de pendientes comunes.

Se analiza en modo específico las generaciones 1992 a 1995 de estudiantes que optaron por la carrera Contador Público, considerando que las generaciones iniciales del plan de estudio vigente (Plan 90) pueden tener un comportamiento sesgado a consecuencia de vincularse a los primeros años en que se instrumenta el Plan. Por otro lado, se considera estudiar generaciones que hubiesen tenido la posibilidad de terminar los cursos de la carrera. La información proviene de la Bedelía de la Facultad actualizada a abril de 2000. Las variables tomadas en cuenta son las notas de aprobación de las distintas asignaturas y las veces que las mismas fueron cursadas y/o rendidas por examen.

Se consideran dos distribuciones según la obtención de los distintos títulos otorgados por la Facultad: el título intermedio (Analista Universitario en Contabilidad que consta de 14 asignaturas) y el título final (Contador Público que se conforma por 33 asignaturas y trabajo monográfico); se presentan aquí los modelos obtenidos para el primer caso.

El trabajo se organiza de la siguiente forma: en primer lugar se presenta en forma sintética las consideraciones generales que refieren al plan de estudios que se analiza, sus disposiciones y particularidades, las características de las generaciones en estudio así como las especificidades del título intermedio. Más adelante se señalan los aspectos metodológicos que se aplica en esta investigación, dando una mención especial al Análisis Discriminante Logístico por ser una técnica muy rica sobre la cual existe poco desarrollo en nuestro medio. Se presenta también el detalle de la matriz de datos con la que se trabaja y la forma en que se organizan los mismos. Por último se describe y analiza los resultados obtenidos. El trabajo concluye con la obtención de un modelo que permite discriminar los estudiantes en función a su rendimiento curricular.

Consideraciones generales

Plan de Estudios 1990

En el año 1990 se pone en práctica en la FCEA el nuevo Plan de Estudios (Plan 90) que sustituye al Plan 80. El Plan de Estudios estructura una secuencia en la formación del individuo que se organiza en 5 años de cursos y un trabajo monográfico final, cuya aprobación otorga al estudiante el título que corresponda.

Los estudiantes que ingresan a la Facultad se inscriben en un Ciclo Básico que corresponde a primer año y es común a las tres carreras que se dictan (Licenciatura en Economía, Licenciatura en Administración y Contador Público). En segundo año los estudiantes optan por dos orientaciones: Económica y Administrativa-Contable, en tercero se separan las tres disciplinas. El sistema se organiza mayormente bajo un régimen de previaturas, el que determina cuales asignaturas es necesario aprobar para estar en condiciones de acceder a materias posteriores, que en general continúan y profundizan a las primeras.

Por otro lado, con la aprobación de 14 materias que se considera que brindan al individuo un primer nivel de formación, se alcanza para cada una de las orientaciones el título intermedio que corresponda¹.

Aquí se analiza específicamente el comportamiento curricular de aquellos estudiantes que pertenecen a las generaciones 1992 a 1995 del Plan 1990, optan por la carrera de Contador Público y obtienen el título intermedio de Analista Universitario en Contabilidad a abril de 2000. Esto se justifica por un lado desde el punto de vista analítico, dado que la obtención del título intermedio brinda una primera instancia de evaluación del rendimiento del estudiante en distintas áreas del conocimiento; por otro lado desde el punto de vista práctico, ya que la orientación de Contador Público es la predominante, aproximadamente un 74% de los inscriptos por generación.

Caracterización por generaciones

En el Cuadro 1 se presenta las principales características de las generaciones consideradas.

Cuadro 1_ Comportamiento lectivo por generación.

	<i>Generación</i>	<i>Generación</i>	<i>Generación</i>	<i>Generación</i>	
	<i>92</i>	<i>93</i>	<i>94</i>	<i>95</i>	<i>Gen 92_95</i>
Ingresos	1.149	1.199	1.379	1.254	4.981
Opción contador	780	822	867	830	3.299
Opción administración	143	162	152	142	599
Opción economista	127	137	168	136	568
Total opciones	1.050	1.121	1.187	1.108	4.466
Cursaron y no aprobaron materias	231	216	313	230	990
No cursaron ni aprobaron	33	36	48	49	166
Título intermedio contador	339	291	270	225	1.125
Título de grado contador	144	82	24	0	250

Elaboración propia en base a datos entregados por la Bedelía de Facultad.

¹ En el caso del estudiante que elige la opción Administrador sector privado, debe aprobar 15 materias para obtener el título de Analista Universitario en Administración.

De la observación del cuadro se desprenden algunos elementos a destacar. En primer lugar, a pesar de que el número de inscriptos tiende a aumentar², se observa un comportamiento similar en las distintas generaciones cuando se analiza la carrera elegida. Así, aproximadamente un 74% de los estudiantes elige la carrera Contador Público, 14% opta por Administración, y 12% por Economía. Interesa señalar que si bien en el cuadro anterior se resume las opciones por carrera, se debe considerar que un estudiante puede figurar en más de una opción y esto es debido a que la pertenencia a un ciclo se deduce por la inscripción a alguna de las materias del ciclo. En este sentido, Altmark et. al. (2000) analizan cuántos estudiantes optan por uno solo de los ciclos y por las distintas combinaciones que se pueden encontrar.

En el período de análisis la suma de las opciones siempre es menor al total de ingresos. Lo anterior se deriva de que en todas las generaciones existe un número de personas que por no aprobar alguna de las materias del Ciclo Básico que establecen preiaturas, o por haber abandonado la Facultad, no están en condiciones de elegir uno o varios de los ciclos siguientes. En particular, del cuadro precedente se desprende que alrededor de un 20% de los estudiantes cursaron pero no aprobaron materias. Además, en todas las generaciones se verifica que aproximadamente un 3% de los inscriptos en Facultad, abandonan sin siquiera anotarse en alguna materia.

Se observa que a medida que se avanza en las generaciones, disminuye el número de personas que alcanzan el título intermedio y el título de grado respectivamente. En este sentido, mientras para la generación 92 hay un 30% de los inscriptos que alcanzan el título de Analista Universitario en Contabilidad, en la generación 93 son 24%, 20% en la generación 94 y 18% en la generación 95. Lo anterior permite presumir que los resultados obtenidos son dinámicos y que todavía no son los definitivos. Más adelante se analizará en mayor detalle este punto, dado que podría afectar los resultados de la presente investigación.

Respecto al título de grado la evolución es similar a la que se presentó en el párrafo anterior. Así, el título de Contador Público lo alcanzan 13% de los inscriptos en el 92, 7% en el 93, 2% en el 94 y a abril de 2000 ninguna persona de la generación 95 había alcanzado el título de Contador.

Particularidades del Título Intermedio

El título de Analista Universitario en Contabilidad se obtiene con algo más del 50% de la carga horaria total del título de grado. Se espera que los individuos que lo alcanzan sean capaces de “generar información a partir de la recopilación de datos primarios, efectuar análisis parciales de la misma y participar en procesos de gestión y supervisión en los niveles medios de las organizaciones”³.

Para cumplir con estos objetivos se incluyen materias que aportan un conocimiento global de la disciplina contable, a la vez que profundizan en ciertas técnicas que presentan el análisis de los criterios de valuación, la contabilidad de costos, el comportamiento y el funcionamiento organizacional y la administración (FCEA, 1995).

² En el documento de trabajo elaborado por Altmark et.al. (2000), se presenta una serie de los ingresos a Ciencias Económicas desde 1990 al 2000 y se verifica una tendencia creciente en la inscripción a Ciclo Básico.

³ Plan de estudios 1990 y reglamento. (FCEA, 1995).

En el cuadro precedente se observa que 1125 estudiantes de las generaciones que se estudian alcanzaron el título intermedio; de éstos se encontró que 11 personas habían revalidado una o más materias. Dado que no llegan al 1% de la población total, se decidió eliminarlas del universo a investigar.

En síntesis, este trabajo se va a centrar en analizar el comportamiento de 1114 estudiantes que alcanzaron el título intermedio y que, como fuera mencionado, pertenecen a distintas generaciones de ingreso. El no pertenecer a la misma generación influye en el grado de avance de los estudiantes. De todos modos, el título de Analista se alcanza con 14 asignaturas, en su mayoría de primer y segundo año y la base de datos es cinco años posterior al ingreso de la última generación analizada (1995).

En el Cuadro 2 se presenta algunos aspectos que sintetizan el comportamiento curricular de los individuos que alcanzan el título intermedio para dos variables de las generaciones en estudio. Las variables que se describen son la *escolaridad* y la *escolaridad corregida*⁴. En el Cuadro 1 del Anexo 1 se presenta los mismos tópicos que aquí se destacan, pero referidos a cada año en particular y se observa que si bien el comportamiento de estas variables es similar, en la medida que se avanza en las generaciones el rendimiento mejora. Esto último puede justificarse si se considera que al pasar algo de tiempo respecto al momento en que el Plan comienza a regir (1990), se podría suponer que se superen inconvenientes respecto a la implementación, adecuación del nivel de exigencia o las dificultades por materia.

Cuadro 2_ Distribución de la escolaridad y escolaridad corregida para los estudiantes con el título intermedio.

	<i>Escolaridad</i>	<i>Escolaridad Corregida</i>
Media	5.01	3.67
Modo	4	3
Varianza Muestral	2.05	2.75
Mínimo	1.92	1.04
Máximo	10.86	10.86
Primer Cuartil	4.00	2.43
Mediana	4.88	3.22
Tercer Cuartil	5.93	4.58
Número de casos	1114	1114

Elaboración propia en base a datos de la Bedelía de Facultad.

De la observación de este cuadro se desprenden algunos comentarios de interés. En primer lugar, se observa que el rendimiento de los estudiantes aproximado por su escolaridad es relativamente bajo. Particularmente si se tiene en cuenta que el rango de variación de la escolaridad va de 0 a 12 y lo más frecuente es encontrar estudiantes con una escolaridad de 4. Por otro lado, en promedio la escolaridad de los individuos con el título intermedio es 5. Además, el 75% de los mismos tiene una escolaridad menor a 6.

⁴ La variable *escolaridad* corresponde a un promedio simple entre las notas obtenidas en las asignaturas que conforman el título de Analista y las veces que debió dar el examen para aprobarlas. En el caso de que la materia se haya aprobado por curso, se considera como que el examen se hubiera dado una vez. La variable *escolaridad corregida* considera en el denominador la cantidad de veces que se dan exámenes más la cantidad de veces que se cursa una asignatura, en el entendido de que es importante diferenciar la situación del estudiante que cursa una vez una materia, de la de otro que lo haga varias veces.

En segundo lugar, este resultado se acentúa aún más cuando la variable en estudio es la escolaridad corregida, lo cual evidencia en general, una modalidad de comportamiento que no prioriza el no repetir cursos.

Aspectos metodológicos

Metodología

Las técnicas utilizadas según el orden de aplicación son Análisis de Componentes Principales, Análisis de Conglomerados y Discriminante Logístico.

La utilización del Análisis de Componentes Principales se enmarca dentro de los objetivos generales de las técnicas factoriales⁵: simplificar; reducir las dimensiones del espacio de análisis; eliminar la información redundante; construir nuevas variables no correlacionadas que describan adecuadamente el fenómeno; describir relaciones entre variables y asociaciones entre individuos.

El Análisis de Clusters o Conglomerados tiene por objeto conformar grupos de individuos de acuerdo a variables cuantitativas de interés. En este caso dichas variables serán las que surjan de la aplicación de la técnica factorial.

La técnica de Discriminación Logística permite construir un modelo que asigna con el mínimo error posible, los individuos a cierto número de grupos preestablecidos, dado un conjunto de variables de interés. Este análisis tiene valor predictivo pues los criterios usados para clasificar una población conocida, pueden aplicarse para nuevos individuos que se incorporen a ella.

El Discriminante Logístico es un tipo de modelización adecuado para variables de respuesta categóricas donde se desconoce la distribución de las variables explicativas, pudiendo ser estas tanto continuas como categóricas. La variable de respuesta puede tener dos o más niveles, pudiendo ser estos tanto nominales como ordinales. Los modelos logísticos de mayor difusión son aquellos de variable de respuesta dicotómica, sin embargo aquí se hará uso de los modelos de respuesta politómica.

Dentro de estos modelos se aplican los llamados logits acumulados (para modelos proporcionales) y los logits generalizados, donde se modelizan distintos logits por subpoblación.

En el caso de los *modelos proporcionales*, el ajuste de logits acumulados se basa en el supuesto de que la influencia de las variables explicativas en la variable de respuesta, es independiente del punto de corte del logit acumulado.

El modelo planteado es el siguiente⁶: $\log it(\theta_{ik}) = \alpha_k + x_i \beta$ donde i es el índice asociado a los individuos y k el índice asociado a las subpoblaciones.

Los distintos logits se pueden ver de la siguiente forma:

$$\log it(\theta_{i1}) = \log \left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3} + \dots + \pi_{iK}} \right) \quad \dots \quad \log it(\theta_{ik-1}) = \log \left(\frac{\pi_{i1} + \pi_{i2} + \dots + \pi_{ik-1}}{\pi_{iK}} \right)$$

donde π_{ik} es la probabilidad de que el individuo i pertenezca a la población k dados los valores de las variables explicativas: x_{i1}, \dots, x_{ip}

⁵ Blanco J., 1995.

⁶ Se utiliza la notación del manual del SAS "Categorical Data Analysis", 1995. Stokes, M. et al.

En el caso del *modelo logit generalizado* el modelo planteado es el siguiente:

$\log it_{ik} = \alpha_k + x_i \beta_k$, a diferencia de los modelos anteriores existe un conjunto de coeficientes de regresión para cada logit. Se ajustan los logits de cada subpoblación con respecto a la última subpoblación.

Los logits serán: $\log it_{i1} = \log\left(\frac{\pi_{i1}}{\pi_{iK}}\right)$, $\log it_{i2} = \log\left(\frac{\pi_{i2}}{\pi_{iK}}\right)$, ..., $\log it_{ik-1} = \log\left(\frac{\pi_{ik-1}}{\pi_{iK}}\right)$.

Matriz de datos

Los individuos objeto de estudio son 1114 estudiantes de las generaciones 1992 a 1995 que alcanzaron el título intermedio. Las variables que conforman la matriz de datos original se asocian al rendimiento del estudiante en las 14 asignaturas que conforman dicho título. Se consideran por un lado, las veces que el estudiante cursó y/o rindió exámenes; por otro lado, la nota con la que aprobó cada asignatura en el curso o en el examen, según corresponda. Así, cada asignatura tiene 4 variables asociadas. La matriz de datos original se estructura entonces por 1114 filas y 56 columnas.

A los efectos de resumir el rendimiento por estudiante a la vez que reducir el espacio de variables, se transformó a las variables originales conjugando en una sola a las que refieren a veces o a notas respectivamente. Concretamente, la variable “veces” sintetiza el número de veces que el estudiante rinde el examen y que cursa una asignatura⁷; la variable “nota” sintetiza la nota con la que el individuo la aprueba sin distinguir si es por curso o por examen. Con el procedimiento anterior la matriz de datos resulta de 1114 filas y 28 columnas.

En el Cuadro 1 del Anexo 2 se presenta la matriz de correlaciones entre las 28 variables de este estudio. Contrariamente a lo que en un principio se esperó, no se verifica la existencia de altas correlaciones entre asignaturas correspondientes a una misma área ni en general. En particular, en todos los casos las correlaciones son menores en magnitud a 0,5. Por lo tanto, se decide trabajar con las 28 variables que reflejan notas y veces por asignatura.

En el Cuadro 3 se detallan las asignaturas necesarias para obtener el título de Analista en Contabilidad agrupadas en cinco áreas diferentes. Más adelante se presentan los resultados del análisis del rendimiento por área.

⁷ En este caso se consideran las inscripciones a curso. Se intenta con esto diferenciar comportamientos, considerando las particularidades de los individuos que cursan una vez y aprueban de las de aquellos que requieren cursar más de una vez para aprobar.

Cuadro 3_ Asignaturas del título intermedio por áreas del conocimiento.

<i>Métodos</i>	<i>Administrativa</i>	<i>Contable</i>	<i>Economía</i>	<i>Otras</i>
Matemática I	Introducción a las Organizaciones	Introducción a la Contabilidad	Economía I	Derecho Civil
Estadística I	Organización y Métodos Administración General	Contabilidad Básica Contabilidad de Costos	Economía Descriptiva I	Sociología Derecho Comercial Introducción a la Computación

Elaboración propia en base a datos de la Bedelía de Facultad.

Análisis de los Resultados

A continuación se describen los resultados de la presente investigación referidos a las diversas técnicas aplicadas. El capítulo se organiza resumiendo en primer lugar los resultados de la aplicación de Componentes Principales a la matriz de datos; en segundo lugar, se exponen las conclusiones que se derivan de la aplicación del Análisis de Conglomerados; por último se detallan los resultados de la discriminación.

Componentes Principales

En una primera instancia se aplica Componentes Principales a la matriz de 1114 estudiantes por 28 variables de rendimiento. Se pretende estudiar la existencia de alguna relación entre las variables utilizadas o alguna asociación entre los individuos objeto de estudio, por otro lado se considera la posibilidad de reducir dimensiones. No se obtienen resultados significativos. Esto junto a la complejidad del análisis, así como la existencia de áreas de estudio diferenciadas, lleva a decidir profundizar en el estudio por áreas.

Se aplica entonces la técnica en cada una de las cinco áreas del conocimiento en las que se agrupó al conjunto de asignaturas que conforman el título intermedio, obteniéndose importantes resultados.

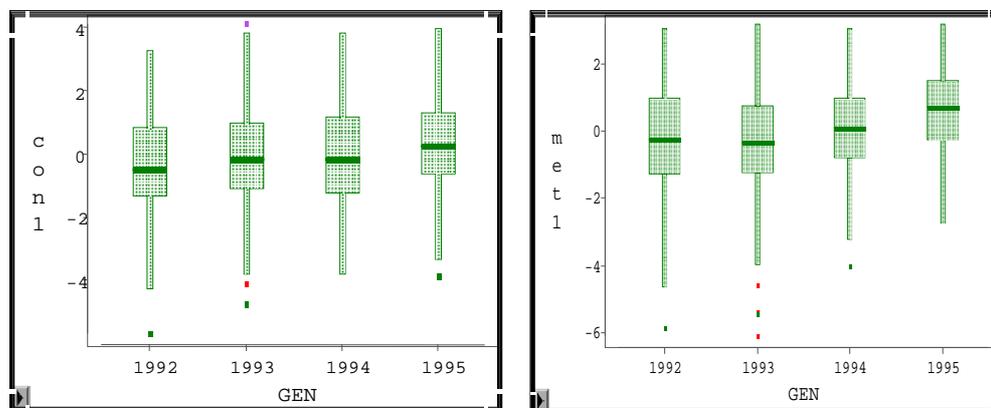
Analizando la posible reducción de dimensiones en cada una de las áreas definidas, se aprecia que en la mayoría de los casos el considerar aproximadamente el 70% de la inercia explicada, significa trabajar con dos o tres ejes. En el caso de Economía y Métodos con dos ejes se explica 70% y 72% respectivamente. En el caso de Contabilidad y Administración tres ejes explican el 68%. Como era de esperar, el área Otros necesita mayores dimensiones para explicar el mismo porcentaje de inercia, dado que está integrada por un mayor número de variables.

El resultado de la aplicación de Componentes Principales por área de estudio pone de manifiesto en todas las áreas un resultado de mucha importancia para la interpretación. **En particular el primer eje en todos los casos opone las veces a las notas.** Los valores negativos están asociados a mayor número de veces que se rinden las asignaturas y los valores positivos a mejores notas obtenidas en el área, es decir, ordena en este sentido a los estudiantes según su rendimiento. **El primer eje explica, por lo tanto, en todos los casos el rendimiento curricular por área.**

A partir de la observación del primer componente, en todas las áreas se verifica que el rendimiento por generación ha ido en aumento. Este hecho es particularmente relevante en las áreas de Métodos y Economía, donde la mediana para la generación 1992 es menor incluso que la primer cuartila de 1995; en las áreas restantes esta constatación no es tan acentuada pero también se verifica. En el gráfico 1 se presenta la evolución por generación del primer componente para las áreas Contabilidad y Métodos.

Los demás factores no tienen una interpretación tan clara como el primero, destacan las veces o la nota de alguna asignatura de las áreas. Específicamente, el segundo componente en el área Contable está influido en gran medida por las veces que se rinde la asignatura de primer año Introducción a la Contabilidad. En el área Administrativa por la nota de la asignatura de tercero Administración General. En el área de Métodos Cuantitativos la variable que tiene mayor preponderancia en el segundo eje es la que refiere a la nota de Estadística I.

Gráfico 1_ Evolución del primer componente por generaciones.



Elaboración propia en base a datos entregados por la Bedelía de Facultad.

Análisis de Conglomerados

Dadas las particularidades de los primeros ejes, se optó por utilizarlos como variables explicativas para hacer el agrupamiento de los estudiantes. Así los estudiantes se agruparán en función a su rendimiento curricular por área.

De los distintos algoritmos de unión usados se consideró que aquel que presentaba una estructura de grupos más clara era el algoritmo de Ward (ver Anexo 3).

La estructura para cinco y tres grupos se puede apreciar en el siguiente cuadro:

Cuadro 4_ Estructura de grupos.					
Con 5 clusters			Con 3 clusters		
Nº en gráfico	Calificación	Cant.Estud.	Nº en gráfico	Calificación	Cant.Estud.
1	Muy buenos	144	1	Muy Buenos	356
2	Buenos_Regulares	382	2	Buenos	615
3	Buenos_Regulares	233	3	Malos	143
4	Excelentes	212			
5	Malos	143			

Elaboración propia en base a datos de la Bedelía de Facultad.

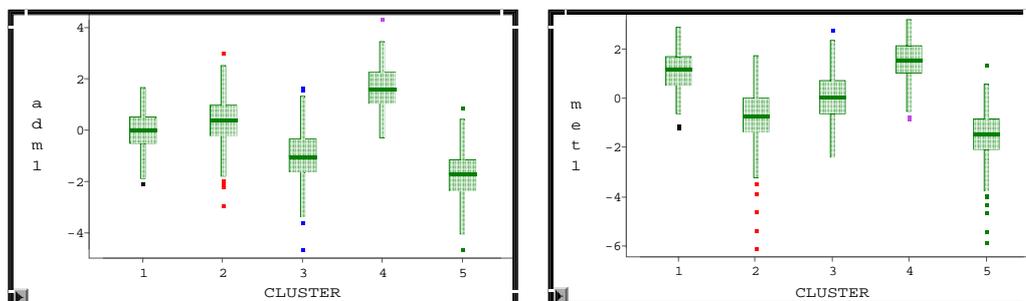
Teniendo en cuenta una estructura de cinco grupos, los estudiantes pueden ser clasificados en excelentes, muy buenos, buenos, regulares y “malos”⁸. Si se consideran tres grupos se pueden clasificar en muy buenos (un grupo de excelentes y muy buenos anterior), buenos (un grupo de buenos y regulares) y malos estudiantes.

Los estudiantes excelentes y muy buenos se destacan por ser los mejores en todas las áreas (las medianas de estos grupos están por encima de las de los demás). Por otro

⁸ La denominación de malos estudiantes debe relativizarse en la medida que se está considerando sólo aquellos que han terminado el título intermedio. En cada generación existe un grupo importante de estudiantes que aún no ha logrado finalizar los requerimientos de dicho título.

lado, los excelentes se destacan de los muy buenos fundamentalmente por tener mejor rendimiento en las áreas de administración, contabilidad y economía. Las medianas de este grupo (grupo 4 en el gráfico) se encuentran por encima a las del grupo de muy buenos (grupo 1).

Gráfico 2 Rendimiento por grupos.



Elaboración propia en base a datos entregados por la Bedelía de Facultad.

Los grupos intermedios (2 y 3) registran un comportamiento heterogéneo. Según el área que se tome en cuenta los mismos pueden ser buenos o regulares. Si se consideran las áreas de Métodos y Contabilidad, el grupo 3 es el de mejores rendimientos. En las áreas de Administración y Derecho el grupo 2 tiene un rendimiento mejor.

El grupo de los “malos” estudiantes se caracteriza por presentar los peores rendimientos en todas las áreas, las medianas en cada una de ellas se encuentran por debajo a las del resto de los grupos.

Si se consideran tres grupos se tiene los “malos” con un muy bajo rendimiento, los “buenos” con un rendimiento medio y los “muy buenos” con un mayor rendimiento en todas las áreas. Del análisis realizado anteriormente se desprende que estos dos últimos grupos presentan un comportamiento diferenciado en sus integrantes según el área que se considere.

A los efectos de realizar el Análisis Discriminante se consideró trabajar con 3 grupos: muy buenos (MB), buenos (B) y malos (M) estudiantes. En el Cuadro 1 del Anexo 1 se presenta el histograma de frecuencias para los grupos identificados, del conjunto de estudiantes de las generaciones en estudio y el de la generación 92.

Análisis Discriminante

A partir de los grupos formados en Análisis de Clusters y las variables de rendimiento que surgen del análisis de Componentes Principales, se aplicó el Discriminante Logístico buscando concretar dos objetivos: por un lado, construir un modelo que permita discriminar lo mejor posible entre grupos; por otro, generar una regla de clasificación con valor predictivo, en el sentido de que los individuos que se incorporen en el futuro puedan ser asignados a uno de estos grupos.

En primer lugar se realizó un test de multinormalidad. Luego de rechazar la hipótesis nula (se rechazó la simetría) se optó por trabajar con un discriminante de tipo logístico.

El análisis planteado tiene como variable de respuesta los grupos formados en cluster (MB, B, M) y como variables explicativas las 5 variables de rendimiento por área encontradas en el análisis de Componentes Principales.

Una primera aproximación consistió en la utilización de logits acumulados, asumiendo modelos proporcionales⁹. En este caso: $\log it(\theta_{ik}) = \alpha_k + x_i \beta$

donde $k = 1, 2, 3$ es la cantidad de grupos e $i = 1, 2, 3, 4, 5$ es la cantidad de variables explicativas.

siendo el primero el logit $\log it(\theta_{i1}) = \log\left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}}\right)$ de los muy buenos estudiantes versus el resto y el segundo el logit $\log it(\theta_{i2}) = \log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3}}\right)$ de los muy buenos y buenos versus los malos estudiantes.

La hipótesis nula de proporcionalidad $H_0) \beta_k = \beta$ es rechazada, al obtener un chi-cuadrado de 95.1212 y un pvalue < 0.0001 (ver Anexo 4), por lo que los modelos logit acumulados de tipo proporcional no son los adecuados.

Se consideró entonces, una segunda aproximación a través de modelos logit generalizados¹⁰. En este caso concreto se estudian dos funciones $\log it_{ik} = \alpha_k + x_i \beta_k$

donde $\log it_{i1} = \log\frac{\pi_{i1}}{\pi_{i3}}$, $\log it_{i2} = \log\frac{\pi_{i2}}{\pi_{i3}}$ siendo el primer logit el de los muy buenos

contra los malos estudiantes y el segundo el de los buenos estudiantes contra el de los malos. Recordar que π_{i1} es la probabilidad de que el estudiante i sea clasificado en el grupo *muy bueno* dado los rendimientos obtenidos en las distintas áreas (variables explicativas).

Del análisis de varianza del modelo se desprende la significación global del mismo así como la significación de las 5 variables utilizadas (ver Anexo 4).

Los dos logits ajustados son:

$$\log it_{i1} = 6.769 + 3.8279met1_i + 2.4387cont1_i + 3.053ecol_i + 1.8499adm1_i + 4.1861der1_i$$

$$\log it_{i2} = 10.6528 + 1.1415met1_i + 1.5364cont1_i + 1.1213ecol_i + 1.9109adm1_i + 2.4618der1_i$$

El conjunto de variables es significativo para todos los grupos en cada uno de los logits ajustados (ver Anexo 4). Los signos positivos de los parámetros estimados estarían indicando que un incremento en una unidad de las variables asociadas a los mismos (mejora en el rendimiento por área) aumentaría la probabilidad de un estudiante de ser clasificado como muy bueno en el primer caso y bueno en el segundo caso. Sin embargo, se puede apreciar el “peso” diferenciado de las distintas variables. En el primer logit se aprecia la gran importancia de las áreas de derecho y de métodos, seguidas del área de economía. En el segundo logit, el área de mayor peso es nuevamente la de derecho seguida en este caso por administración y contabilidad.

⁹ Los modelos se ajustaron usando SAS a través del proc logistic.

¹⁰ Los modelos se ajustaron usando SAS a través del proc catmod.

Si se desea obtener el logit de los muy buenos versus los buenos estudiantes se realiza la resta de los logits antes encontrados.

$$\text{logit}_{i1} - \text{logit}_{i2} = -3.8838 + 2.6864\text{met}1_i + 0.902\text{cont}1_i + 1.9317\text{ecol}_i - 0.061\text{adm}1_i + 1.7243\text{der}1_i$$

Observando el Cuadro 5 se aprecia el poder predictivo del Discriminante Logístico realizado a través de logits generalizados.

Cuadro 5. Cluster versus predicción.

<i>Cluster\predict.</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>Total</i>
1- muy buenos	324	32	-	356
2- buenos	33	567	15	615
3- malos	-	17	126	143
Total	357	616	141	1114

Elaboración propia en base a datos entregados por la Bedelía de Facultad.

El poder predictivo total no supera el 10%. Si se observa al interior de los grupos la re-clasificación fue aproximadamente del 12% en el grupo de los malos estudiantes, de 9% en el de los muy buenos y de 8% en el de los buenos estudiantes. En todos los casos los re-clasificados fueron “enviados” al grupo contiguo, por ejemplo los re-clasificados en el grupo de muy buenos fueron enviados al grupo de buenos.

Conclusiones y perspectivas

Conclusiones

1. A partir de conocer el rendimiento curricular de un individuo en las áreas consideradas, se obtiene una regla de asignación para los estudiantes de las distintas generaciones. El modelo adecuado es un logit generalizado.
2. Se verifica la existencia de grupos de comportamiento diferenciado en función del rendimiento curricular por áreas, los cuales resultan definidos por las primeras Componentes Principales de cada área.
3. A los efectos de ordenar los estudiantes dentro de cada grupo, resulta muy factible considerar la combinación lineal óptima definida por los primeros autovectores de cada área.
4. La presente investigación ha permitido avanzar en el conocimiento del comportamiento curricular de las generaciones.

Perspectivas

1. El trabajo continúa en la profundización del ordenamiento de los estudiantes por generación de acuerdo a su rendimiento curricular, utilizando los resultados obtenidos al aplicar Componentes Principales por área.
2. La investigación se completa con el análisis del rendimiento curricular de los graduados de Contador Público, así como del de los Licenciados en Economía, tanto para el título intermedio como para los graduados.
3. En una etapa posterior del análisis, se considera complementar los resultados obtenidos respecto al rendimiento curricular, con las características socioeconómicas de los estudiantes.

Referencias bibliográficas

Agresti A., *Categorical Data Analysis*, John Wiley & Sons, New York, 1990.

Altmark S., Biurrun J., Nalbarte L., “Análisis de las generaciones del plan 90”. Documento de trabajo del Instituto de Estadística, Montevideo, 2000.

Blanco J., *Introducción al Análisis Multivariado*, Montevideo, 1995.

FCEA, *Plan de estudios 1990 y Reglamento*, Montevideo, 1995.

Hair J., Anderson R., Tatham R., Black W., *Análisis Multivariante*, 5.^a ed., Prentice Hall, Madrid, 1999.

Kleinbaum D., *Logistic Regression. A self learning text*, Springer_Verlag New York, 1994.

McLachlan G., *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, USA, 1992.

Stokes M., Davis Ch., Koch G., Manual del SAS, “Categorical Data Analysis. Using the SAS System”, Cary: NC: SAS Institute Inc., 1995.

Anexo 1

Escolaridad y Escolaridad corregida en las generaciones 1992 a 1995

	<i>Escolaridad</i>				<i>Escolaridad corregida</i>			
	<i>1992</i>	<i>1993</i>	<i>1994</i>	<i>1995</i>	<i>1992</i>	<i>1993</i>	<i>1994</i>	<i>1995</i>
Media	4,75	4,82	5,13	5,52	3,29	3,5	3,7	4,4
Modo	4	5	5	5	2	3	3	4
Varianza Muestral	1,80	2,33	1,73	2,05	2,07	3,08	2,31	3,14
Mínimo	2,00	1,92	2,46	2,04	1,04	1,16	1,44	1,42
Máximo	8,71	10,86	9,86	9,36	8,71	10,9	9,86	9,36
Primer Cuartil	3,80	3,74	4,13	4,47	2,22	2,30	2,60	3,06
Mediana	4,53	4,53	5,00	5,40	2,88	2,96	3,23	4
Tercer Cuartil	5,79	5,73	6,04	6,54	4,15	4,27	4,47	5,53
Desviación Estándar	1,34	1,53	1,32	1,43	1,44	1,76	1,52	1,77
Número de Casos	335	289	267	223	335	289	267	223

Frecuencias de estudiantes en los 3 grupos para la generación 1992 y 1992_1995

<i>Grupos</i>	<i>Generación 1992</i>		<i>Generaciones 1992_1995</i>	
	<i>n(x)</i>	<i>h(x)</i>	<i>n(x)</i>	<i>h(x)</i>
1	146	0.44	356	0.32
2	128	0.38	615	0.55
3	61	0.18	143	0.13
	335	1	1114	1

Anexo 2

Matriz de correlaciones

<u>_NAME_</u>	veces 110	Nota 110	veces 111	Nota 111	veces 112	Nota 112	veces 113	Nota 113	veces 114	Nota 114	veces 115	Nota 115	veces 116	Nota 116	veces 210	Nota 210	veces 214	Nota 214	veces 217	Nota 217	veces 218	Nota 218	veces 314	Nota 314	veces 316	Nota 316	veces 318	Nota 318
veces110	1.00	-0.45	0.39	-0.33	0.28	-0.29	0.13	-0.13	0.25	-0.20	0.11	-0.08	0.42	-0.30	0.20	-0.18	0.21	-0.15	0.20	-0.17	0.31	-0.21	0.12	-0.10	0.25	-0.28	0.18	-0.12
nota110	-0.45	1.00	-0.30	0.44	-0.30	0.44	-0.11	0.18	-0.23	0.24	-0.12	0.10	-0.34	0.40	-0.34	0.32	-0.20	0.23	-0.22	0.21	-0.39	0.30	-0.21	0.17	-0.26	0.32	-0.16	0.21
veces111	0.39	-0.30	1.00	-0.43	0.33	-0.30	0.13	-0.06	0.19	-0.17	0.13	-0.07	0.37	-0.27	0.25	-0.15	0.19	-0.13	0.15	-0.20	0.40	-0.21	0.16	-0.08	0.25	-0.24	0.14	-0.13
Nota 111	-0.33	0.44	-0.43	1.00	-0.29	0.38	-0.11	0.13	-0.18	0.21	-0.09	0.14	-0.30	0.30	-0.25	0.24	-0.25	0.19	-0.19	0.23	-0.44	0.33	-0.17	0.15	-0.21	0.27	-0.18	0.18
Veces112	0.28	-0.30	0.33	-0.29	1.00	-0.23	0.22	-0.16	0.37	-0.23	0.15	-0.10	0.37	-0.22	0.28	-0.13	0.25	-0.12	0.26	-0.15	0.27	-0.16	0.16	-0.12	0.19	-0.18	0.16	-0.11
Nota 112	-0.29	0.44	-0.30	0.38	-0.23	1.00	-0.16	0.25	-0.28	0.33	-0.14	0.11	-0.34	0.41	-0.30	0.28	-0.24	0.25	-0.19	0.20	-0.29	0.29	-0.24	0.13	-0.23	0.35	-0.17	0.18
veces113	0.13	-0.11	0.13	-0.11	0.22	-0.16	1.00	-0.02	0.23	-0.16	0.14	-0.07	0.19	-0.13	0.13	-0.04	0.18	-0.07	0.14	-0.10	0.12	-0.06	0.06	-0.07	0.05	-0.07	0.04	-0.07
Nota 113	-0.13	0.18	-0.06	0.13	-0.16	0.25	-0.02	1.00	-0.17	0.16	-0.11	0.09	-0.14	0.14	-0.13	0.11	-0.12	0.15	-0.11	0.09	-0.09	0.09	-0.14	0.12	-0.11	0.16	-0.09	0.03
veces114	0.25	-0.23	0.19	-0.18	0.37	-0.28	0.23	-0.17	1.00	-0.45	0.24	-0.11	0.32	-0.19	0.20	-0.08	0.22	-0.19	0.24	-0.10	0.15	-0.12	0.17	-0.12	0.14	-0.19	0.16	-0.09
nota114	-0.20	0.24	-0.17	0.21	-0.23	0.33	-0.16	0.16	-0.45	1.00	-0.19	0.20	-0.20	0.25	-0.18	0.10	-0.23	0.22	-0.13	0.17	-0.16	0.15	-0.19	0.09	-0.16	0.19	-0.13	0.15
veces115	0.11	-0.12	0.13	-0.09	0.15	-0.14	0.14	-0.11	0.24	-0.19	1.00	-0.01	0.23	-0.06	0.11	-0.03	0.20	-0.07	0.18	-0.08	0.11	-0.02	0.12	-0.08	0.10	-0.04	0.09	-0.07
Nota 115	-0.08	0.10	-0.07	0.14	-0.10	0.11	-0.07	0.09	-0.11	0.20	-0.01	1.00	-0.08	0.12	-0.04	0.05	-0.17	0.10	-0.14	0.07	-0.07	0.11	-0.03	0.05	-0.08	0.11	-0.06	0.07
Veces116	0.42	-0.34	0.37	-0.30	0.37	-0.34	0.19	-0.14	0.32	-0.20	0.23	-0.08	1.00	-0.30	0.26	-0.20	0.27	-0.19	0.28	-0.15	0.34	-0.19	0.25	-0.13	0.24	-0.21	0.18	-0.12
Nota 116	-0.30	0.40	-0.27	0.30	-0.22	0.41	-0.13	0.14	-0.19	0.25	-0.06	0.12	-0.30	1.00	-0.28	0.24	-0.18	0.17	-0.17	0.15	-0.27	0.22	-0.20	0.10	-0.18	0.27	-0.17	0.15
Veces210	0.20	-0.34	0.25	-0.25	0.28	-0.30	0.13	-0.13	0.20	-0.18	0.11	-0.04	0.26	-0.28	1.00	-0.24	0.23	-0.16	0.19	-0.18	0.34	-0.26	0.20	-0.13	0.25	-0.31	0.18	-0.16
Nota 210	-0.18	0.32	-0.15	0.24	-0.13	0.28	-0.04	0.11	-0.08	0.10	-0.03	0.05	-0.20	0.24	-0.24	1.00	-0.13	0.18	-0.13	0.18	-0.22	0.22	-0.17	0.17	-0.22	0.26	-0.13	0.14
Veces214	0.21	-0.20	0.19	-0.25	0.25	-0.24	0.18	-0.12	0.22	-0.23	0.20	-0.17	0.27	-0.18	0.23	-0.13	1.00	-0.15	0.25	-0.14	0.24	-0.20	0.15	-0.05	0.21	-0.16	0.16	-0.12
Nota 214	-0.15	0.23	-0.13	0.19	-0.12	0.25	-0.07	0.15	-0.19	0.22	-0.07	0.10	-0.19	0.17	-0.16	0.18	-0.15	1.00	-0.16	0.13	-0.14	0.13	-0.23	0.15	-0.15	0.15	-0.14	0.16
veces217	0.20	-0.22	0.15	-0.19	0.26	-0.19	0.14	-0.11	0.24	-0.13	0.18	-0.14	0.28	-0.17	0.19	-0.13	0.25	-0.16	1.00	-0.02	0.28	-0.14	0.24	-0.14	0.18	-0.19	0.20	-0.17
nota217	-0.17	0.21	-0.20	0.23	-0.15	0.20	-0.10	0.09	-0.10	0.17	-0.08	0.07	-0.15	0.15	-0.18	0.18	-0.14	0.13	-0.02	1.00	-0.17	0.12	-0.14	0.12	-0.14	0.11	-0.14	0.09
veces218	0.31	-0.39	0.40	-0.44	0.27	-0.29	0.12	-0.09	0.15	-0.16	0.11	-0.07	0.34	-0.27	0.34	-0.22	0.24	-0.14	0.28	-0.17	1.00	-0.33	0.18	-0.16	0.26	-0.26	0.21	-0.19
nota218	-0.21	0.30	-0.21	0.33	-0.16	0.29	-0.06	0.09	-0.12	0.15	-0.02	0.11	-0.19	0.22	-0.26	0.22	-0.20	0.13	-0.14	0.12	-0.33	1.00	-0.14	0.12	-0.18	0.27	-0.14	0.17
veces314	0.12	-0.21	0.16	-0.17	0.16	-0.24	0.06	-0.14	0.17	-0.19	0.12	-0.03	0.25	-0.20	0.20	-0.17	0.15	-0.23	0.24	-0.14	0.18	-0.14	1.00	-0.28	0.22	-0.17	0.21	-0.20
nota314	-0.10	0.17	-0.08	0.15	-0.12	0.13	-0.07	0.12	-0.12	0.09	-0.08	0.05	-0.13	0.10	-0.13	0.17	-0.05	0.15	-0.14	0.12	-0.16	0.12	-0.28	1.00	-0.12	0.03	-0.09	0.13
veces316	0.25	-0.26	0.25	-0.21	0.19	-0.23	0.05	-0.11	0.14	-0.16	0.10	-0.08	0.24	-0.18	0.25	-0.22	0.21	-0.15	0.18	-0.14	0.26	-0.18	0.22	-0.12	1.00	-0.33	0.23	-0.14
nota316	-0.28	0.32	-0.24	0.27	-0.18	0.35	-0.07	0.16	-0.19	0.19	-0.04	0.11	-0.21	0.27	-0.31	0.26	-0.16	0.15	-0.19	0.11	-0.26	0.27	-0.17	0.03	-0.33	1.00	-0.21	0.17
veces318	0.18	-0.16	0.14	-0.18	0.16	-0.17	0.04	-0.09	0.16	-0.13	0.09	-0.06	0.18	-0.17	0.18	-0.13	0.16	-0.14	0.20	-0.14	0.21	-0.14	0.21	-0.09	0.23	-0.21	1.00	-0.04
nota318	-0.12	0.21	-0.13	0.18	-0.11	0.18	-0.07	0.03	-0.09	0.15	-0.07	0.07	-0.12	0.15	-0.16	0.14	-0.12	0.16	-0.17	0.09	-0.19	0.17	-0.20	0.13	-0.14	0.17	-0.04	1.00

Anexo 3

Análisis de cluster: Método Ward.

Cluster History

NCL	----	Clusters	Joined	----	FREQ	SPRSQ	RSQ	ERSQ	CCC	PSF	PST2
30	CL33	CL51			98	0.0037	.780	.837	-20	132	19.2
29	CL46	CL62			80	0.0038	.776	.835	-20	134	24.7
28	CL36	CL90			71	0.0039	.772	.832	-20	136	21.0
27	CL73	CL69			64	0.0039	.768	.829	-20	138	29.7
26	CL47	CL70			36	0.0042	.764	.826	-20	141	19.4
25	CL52	CL85			72	0.0043	.759	.823	-20	143	28.9
24	CL34	CL57			38	0.0045	.755	.820	-21	146	12.2
23	CL44	CL41			60	0.0050	.750	.816	-21	149	21.8
22	CL35	CL68			106	0.0055	.744	.813	-21	152	32.9
21	CL43	CL61			80	0.0056	.739	.809	-21	155	34.3
20	CL29	CL25			152	0.0056	.733	.805	-21	158	27.5
19	CL53	CL49			75	0.0060	.727	.800	-21	162	30.2
18	CL27	CL21			144	0.0061	.721	.796	-22	167	28.5
17	CL72	CL26			60	0.0064	.715	.791	-22	172	21.0
16	CL38	CL45			45	0.0064	.708	.785	-21	178	18.7
15	CL16	CL24			83	0.0071	.701	.779	-21	184	14.7
14	CL32	CL42			92	0.0075	.694	.772	-21	192	30.2
13	CL55	CL14			128	0.0080	.686	.765	-21	200	27.6
12	CL20	CL56			184	0.0083	.678	.757	-21	210	35.1
11	CL31	CL28			148	0.0096	.668	.748	-21	222	44.6
10	CL19	CL30			173	0.0109	.657	.738	-20	235	43.9
9	CL23	CL15			143	0.0113	.646	.726	-20	252	24.7
8	CL22	CL11			254	0.0141	.632	.712	-20	271	55.6
7	CL17	CL10			233	0.0149	.617	.696	-19	297	44.3
6	CL12	CL67			212	0.0158	.601	.676	-18	334	59.0
5	CL8	CL13			382	0.0372	.564	.651	-18	358	116
4	CL6	CL18			356	0.0433	.520	.617	-16	401	141
3	CL5	CL7			615	0.0536	.467	.567	-14	486	130
2	CL3	CL9			758	0.1013	.366	.467	-10	641	201
1	CL2	CL4			1114	0.3655	.000	.000	0.00	.	641

Anexo 4

Discriminante logístico: modelos proporcionales.

The LOGISTIC Procedure

Model Information

Response Profile

Ordered Value	CLUSTER	Total Frequency
1	1	356
2	2	615
3	3	143

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
95.1212	5	<.0001

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2134.088	550.126
SC	2144.119	585.236
-2 Log L	2130.088	536.126

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1593.9614	5	<.0001
Score	835.2358	5	<.0001
Wald	266.6771	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-3.3625	0.2466	186.0040	<.0001
Intercept2	1	7.9794	0.4851	270.5768	<.0001
met1	1	1.5784	0.1374	132.0582	<.0001
con1	1	1.0170	0.1190	73.0558	<.0001
adm1	1	0.6951	0.1087	40.9273	<.0001
eco1	1	1.3143	0.1445	82.6876	<.0001
der1	1	1.7109	0.1452	138.7784	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
met1	4.847	3.703 6.345
con1	2.765	2.190 3.491
adm1	2.004	1.620 2.479
eco1	3.722	2.804 4.941
der1	5.534	4.163 7.356

Discriminante logístico: logit generalizados.

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	2	202.95	<.0001
met1	2	124.67	<.0001
con1	2	68.09	<.0001
adm1	2	57.09	<.0001
eco1	2	86.69	<.0001
der1	2	120.74	<.0001
Likelihood Ratio	2E3	430.18	1.0000

Analysis of Maximum Likelihood Estimates

Effect	Parameter	Standard Estimate	Chi-Error	Square	Pr > ChiSq
Intercept	1	6.7610	1.1861	32.49	<.0001
	2	10.6528	1.1291	89.01	<.0001
met1	3	3.8279	0.3505	119.27	<.0001
	4	1.4150	0.2417	34.26	<.0001
con1	5	2.4387	0.3017	65.35	<.0001
	6	1.5364	0.2563	35.95	<.0001
adm1	7	1.8499	0.2924	40.03	<.0001
	8	1.9109	0.2533	56.92	<.0001
eco1	9	3.0530	0.3323	84.43	<.0001
	10	1.1213	0.2225	25.39	<.0001
der1	11	4.1861	0.3891	115.75	<.0001
	12	2.4618	0.3219	58.50	<.0001