VALORES Y OCUPACIONES.
UN ANALISIS DISCRIMINANTE

por ELVIRA REPETTO TALAVERA
Universidad Nacional de Educación a Distancia

1. *Introducción*

La condición diferenciada de los educandos y la identificación de los múltiples factores educativos en diversas categorías, viene impuesta por el mismo fenómeno de la educación. En las investigaciones pedagógicas actuales ocupa, por tanto, un lugar destacado el estudio discriminante descriptivo de la naturaleza de las diferentes clases en que aparecen agrupados los diferentes educandos, dentro de una perspectiva sistémica. Pero aún, con ser de suma importancia la discriminación descriptiva de los grupos en la educación, estimo que hoy tiene una mayor relevancia el análisis discriminante con un fin decisional, que después de apreciar la corrección de las clasificaciones estudiadas, respecto a determinadas variables discriminantes, calcula las funciones de clasificación, con el fin de poder asignar los nuevos educandos a aquellos grupos a los que más se asemejen.

De otra parte, existen métodos de clasificación que resuelven problemas discriminantes entre grupos que no poseen equi-covarianza. Estos procedimientos de clasificación cuadrática, surgidos en fechas más recientes que los lineales clásicos, suponen una mayor adecuación de la investigación pedagógica a la realidad educativa, ya que con frecuencia no se pueden aplicar diseños de discriminación lineal porque las muestras de la que se disponen los datos no cumplen las asunciones exigidas.

Aunque la función lineal de Fisher se conoce desde hace medio siglo, sólo en fecha reciente, con la explosión tecnológica de los ordenadores, se han hecho accesibles estos métodos para su utilización en las ciencias sociales y en nuestra ciencia pedagógica. En efecto, Fisher la deriva en

Revista Española de Pedagogía
Año XLIII, n.º 168, abril-junio 1985

su artículo «The use of Multiple Measurements in Toxonomic Problems», publicado en *Annals of Eugenics*, el año 1936, para resolver problemas de antropología física y de biología. Los detalles de la teoría y de las aplicaciones del análisis discriminante están bien documentados, y no voy a tratarlos aquí. Me limito a citar algunos de los realizados por autores pioneros en el tema y otros más recientes. De esta suerte, Tatsuoka y Tiedeman (1954) y Kendall (1976) proporcionan discusiones valiosas sobre el desarrollo histórico de estas técnicas; y Morrison (1969) y (1974) realiza una certera síntesis de las mismas. Son muchas las publicaciones focalizadas hacia el empleo de la discriminación en las ciencias sociales. Así, por ejemplo, las de Romeder (1973), Cacaullos (1973), Lachenbruch (1975), Klecka (1980) y Hudson (1982), sin olvidar las clásicas de Anderson (1958) y Rao (1970, 1952).

2. *Planteamiento del diseño sobre valores y ocupaciones*

Después de la *revisión de la literatura* sobre el tema de valores y ocupaciones, me planteo un *doble tipo de problemas;* ¿las dimensiones axiológicas discriminan las ocupaciones? ¿Cabe obtener funciones de clasificación valorativa ocupacional que se aplique a los alumnos en el diagnóstico y pronóstico en los que se apoya la Orientación Educativa?

Sin que pueda ahora detenerme en la amplia fundamentación filosófica y psicológica del constructo valor que he llevado a cabo, sí quisiera formular dos *definiciones* del mismo. Siguiendo de cerca a un estudioso del tema, lo conceptualizo como «los objetivos hacia los que el hombre se dirige porque los estima como deseables». Desde un punto de vista operacional el constructo valor «reside en la valoración que un sujeto hace de la realidad expresada verbalmente».

La H_o cabe formularla diciendo: «Las áreas axiológicas medidas con el instrumento de exploración seleccionado, no discriminan significativamente los grupos ocupacionales estudiados.»

Las *variables* son doce familias ocupacionales y diez valores. Las doce familias ocupacionales se definen siguiendo la «Clasificación Internacional Uniforme de Ocupaciones» publicada por la OIT (1976) y son las siguientes: Arquitectos, Químicos, Pilotos de Aviación Civil, Médicos, Farmacéuticos, Economistas, Juristas, Profesores, Periodistas, Administrativos, Comerciantes Propietarios y Oficiales de las Fuerzas Armadas.

Los *instrumentos de exploración* empleados son el «Test de Reacción Valorativa» (T.R.V.) del Dr. García Hoz (1977) y un cuestionario ocupacional. El T.R.V., que según su ficha técnica tiene un grado de fiabilidad entre .91 y .94 y un coeficiente de validez de .81, ofrece dos índices de intensidad y sentido, proporcionando así veinte variables, dos para cada valor, una cuantitativa y otra cualitativa. Las diez áreas de valor son: dinero, fama, poder, placer, actividad-negociante, artista, técnico, intelectual, social y trascendental. El fin del cuestionario ocupacional es describir la naturaleza de las tareas realizadas por estos profesionales que realizaron el T.R.V., incluyendo la autodescripción de las funciones, la finalidad y el grado de satisfacción que obtienen en el trabajo.

La muestra se selecciona entre las Asociaciones de Padres de un Sistema Escolar Español. Después de eliminar algunos protocolos por sus irregularidades o por la escasa representación de sus ocupaciones, la muestra está constituida por un total de 1.082 profesionales agrupados en las doce ocupaciones antes citadas.

Entre los *tratamientos estadísticos* que pueden emplearse se estima que el más adecuado es el *Análisis discriminante* que pretende discriminar a los doce grupos ocupacionales en base a sus valores y clasificar nuevos sujetos en el grupo al que más se asemeje en función de sus valores. Se considera que las variables familias ocupacionales pertenecen al nivel nominal y los valores al nivel de intervalo.

De las *condiciones subyacentes al modelo se comprueban* las que se consideran más necesarias. Así, se efectúa un *análisis de varianza* para corroborar la significatividad de las diferencias de medias entre los grupos: todas son significativas excepto la «intensidad intelectual». Para contrastar la *homogeneidad de las covarianzas de los grupos,* se utiliza el test de Box (1949) que se basa en el cálculo estadístico:

$$\left\{1-\left[\frac{2n^2+3n-1}{6(n+1)\,(K-1)}\right]\left(\Sigma\frac{1}{N_t-1}-\frac{1}{N-K}\right)\right\}\left[(\qquad \hat{S}\,|-\sum_{t=1}^{K}\qquad \log|\right]$$

siendo $\hat{S}_t = \frac{N_t}{N_t - 1} S_t$ la estimación insesgada de la matriz C_t.

Como la $H_0 \neq C_j \neq \ldots C_k$ no se rechaza, el análisis de discriminación lineal al no cumplir los requisitos del modelo puede proporcionarnos resultados defectuosos, mientras que de la discriminación cuadrática cabe obtener datos válidos. No obstante, como según Lachenburch (1973), esta técnica tolera algunas violaciones de las asunciones, me decido a aplicar primero un análisis discriminante lineal.

3. *Discriminación lineal descriptiva*

Los diseños de discriminación lineal descriptivos pretenden la reducción de las múltpiles variables P a K combinaciones ponderadas lineales que tengan el potencial máximo para distinguir entre los grupos.

Se denomina poder discriminante del eje factorial u_1 a la cantidad λ_1. U_1 contiene los coeficientes ponderados que definen la primera función discriminante canónica, es decir, la primera variable canónica o combinación lineal de las medidas múltiples que explican la mayor covarianza λ_1 entre los centroides de los grupos. La segunda variable canónica se defiene por el compuesto ponderado de los P coeficientes del vector u_2 asociado con la mayor covarianza entre los grupos λ_2 después de que la covarianza λ_1 se ha eliminado: $\lambda_2 < \lambda_1$. El máximo número de K posibles funciones corresponde a las únicas soluciones matriciales (no ceros) de esta ecuación determinante $|T^{-1} B - \lambda| = 0$, siendo igual al número de grupos menos 1 (G-1) o al número de variables discriminantes si son menores que los grupos.

Los coeficientes u pueden emplearse en las funciones discriminantes, pero si se ajustan sus valores llegaremos a contar con otros coeficientes que dan a la función mejores propiedades. A estos últimos coeficientes los denominamos C y se definen como $C_i = u_i \sqrt{n - G}$; $C_o = tc_i g_i$. Para tipificarlos se transforman mediante $Z_i = C_i \sqrt{\dfrac{SS\ W_i}{N - g}}$

A continuación presento los nueve coeficientes de las funciones discriminantes canónicas obtenidas y las constantes.

TABLA 1.—*Coeficientes de las funciones discriminantes canónicas*

Variable	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
1	0.01138	0.01239	—0.07676	0.01702	—0.01504	—0.01121	0.06395	0.08130	—0.02853
2	0.00679	—0.02203	0.05237	0.01702	—0.00615	0.01087	0.09501	—0.04827	—0.02969
3	—0.01115	—0.06273	0.06404	—0.03116	0.06082	—0.06080	—0.01067	—0.01149	—0.00628
4	—0.02030	0.01203	—0.02673	0.00562	0.03795	0.06918	—0.04087	—0.03128	—0.00981
5	0.04071	0.04147	—0.00272	—0.01671	—0.07926	—0.04932	0.01464	—0.03012	—0.01897
6	—0.30021	—0.02700	—0.00294	0.00527	—0.01617	—0.02332	—0.05604	—0.00947	—0.05725
7	0.01507	—0.01586	—0.00655	—0.00668	—0.00352	0.03039	—0.03037	—0.03692	0.03690
8	—0.04639	0.00309	—0.02530	0.00078	0.03084	0.05453	0.00952	0.01348	—0.02354
9	—0.00315	—0.01370	—0.02953	0.01506	0.05675	—0.03988	—0.01300	0.05612	0.10554
10	0.31954	0.03818	0.02135	—0.01475	0.01307	0.00006	0.01293	—0.00343	—0.03416
11	—0.02282	0.03500	—0.03472	0.01043	0.03157	—0.00984	0.02534	—0.06312	—0.00126
12	0.00552	0.02538	—0.01351	—0.01509	—0.02589	0.01597	—0.00116	0.02983	—0.01600
13	—0.01328	—0.01849	0.00303	—0.07208	—0.00105	—0.00439	0.01053	0.02023	0.02336
14	—0.70352	—0.03795	—0.00405	—0.00902	—0.00802	—0.02770	0.03100	0.01225	0.01714
15	0.07327	—0.00749	—0.00001	0.04218	0.03737	—0.02390	—0.01731	0.01950	0.01410
16	—0.07106	0.02723	—0.01243	—0.01981	—0.02996	—0.03127	—0.01079	0.01060	0.01654
17	0.01380	—0.01354	—0.02092	—0.02356	—0.01764	0.01862	0.00241	0.00229	—0.00558
18	0.01222	—0.00228	0.03776	0.00503	—0.01192	0.02770	0.01235	0.02206	0.01889
19	—0.02808	0.02893	0.02752	0.01624	—0.03993	0.00864	0.02480	—0.06609	0.00100
20	—0.01025	0.02546	0.00940	—0.00894	—0.03364	—0.00260	—0.03941	0.00153	0.01020
Z_o	0.03191	—0.75805	0.13687	0.77583	—0.88524	0.88436	0.36998	—0.00185	0.47591
	1.41492								

Para calcular las funciones discriminantes canónicas aplico la fórmula fi $(X_{jk}) = \Sigma C_i \; X_{jk} + C_o$; con coeficientes tipificados fi $(X_{jk}) = \Sigma Z_i \; X_{jk} + Z_o$ siendo $f_i (x_{jk})$ = valor de la función discriminante canónica i para un individuo x; X_{jk} = valor del individuo x en la variable discriminante j en el grupo k; C_i = coeficiente que produce la característica deseada en la función; $Z_j = C_i$ tipificado; C_o = constante: es el ajuste para las medias, de tal modo que la puntuación discriminante media sea cero para todos los casos $Z;_o = C_o$ tipificado. De este modo se hallan las nueve funciones que mejor discriminan a los grupos entre sí.

4. *Procedimientos de clasificación lineal y cuadrática*

Los métodos de clasificación lineal y cuadrática constituyen el núcleo fundamental de esta comunicación, ya que en la actualidad, la mayor parte de los autores le atribuyen un puesto prioritario. Hay, por ejemplo, en «*Discriminant Analysis*» (1982), sostiene que la aplicación usual del análisis discriminante es un problema de asignación de la teoría de la decisión. Según Hudson (1982), Lachenbruch (1975), Thorndike (1978) y Romeder (1973), el proceso de discriminación es esencialmente decisional. Hawkin en *Topics in Applied Multivariate Analysis* (1982), subraya que este análisis pretende la clasificación de un sujeto de origen desconocido en uno de dos o más grupos distintos sobre la base de las observacioneso btenidas. Es decir, se necesita partir de la evidencia científica de los grupos para después asignar el individuo nuevo a la clase con la que tenga una distancia menor o mayor semejanza.

Aquí sólo voy a tratar de los sistemas de clasificación lineal y cuadrática mediante las funciones discriminantes de clasificación.

4.1. *Funciones de clasificación lineal*

Las *funciones de clasificación* de dos grupos bajo la hipótesis de normalidad y equicovarianza de Fisher las generaliza a multigrupos, Lachenbruch (1975) y Romeder (1973). Se asigna el sujeto x al grupo y_i si su distancia medida por $(x - \bar{y}_i)' \, \Sigma^{-1}(x - \bar{y}_i) < (x - \bar{y}_j)' \;\; \Sigma^{-1}(x - \bar{y}_j)$ es menor que la misma distancia referida a los restantes grupos. Una derivación análoga es la aplicada por el programa BMDP7M: $f(i) = \Sigma B_i x_{ij} + \alpha$, resultando los coeficientes de las funciones de clasificación lineal.

Tabla 2.—*Coeficientes de las funciones de clasificación lineal*

Variable / Grupo	Arqui.	Quimi.	Pilot.	Medic.	Farma.	Econo.	Juris.	Profe.	Period.	Admin.	Comer	Ofici.
1 A1	0.24980	0.21175	0.24714	0.24414	0.17312	0.25305	0.19111	0.19106	0.19381	0.28179	0.24179	0.16756
2 A2	0.07208	0.07208	0.09292	0.06475	0.06272	0.09936	0.05404	0.08137	0.09198	0.01694	0.08690	—0.09449
3 A3	0.01896	0.07198	0.02342	—0.00255	0.04072	0.05997	0.08280	0.05315	0.14115	0.02668	0.00496	0.14108
4 A4	0.07527	0.07332	0.06229	0.12847	0.10493	0.04940	0.05892	0.11452	0.07607	0.11077	0.05645	0.08785
5 A5	0.08868	0.03602	0.84909	—0.01066	0.06948	0.04816	0.03984	—0.00640	—0.05204	0.01348	0.09164	0.03483
6 A6	0.00208	0.07278	0.10286	0.02008	0.05373	0.04823	0.07022	0.05948	0.07335	0.05752	0.03931	0.03146
7 A7	—0.02887	—0.00659	—0.01152	—0.02929	—0.02976	—0.05875	—0.05808	—0.04125	—0.04639	—0.03698	—0.03219	—0.01315
8 A8	0.01129	0.00988	0.07392	0.10007	0.05994	0.04714	0.05037	0.09562	0.08734	0.08740	0.01516	0.01985
9 A9	0.01445	0.04190	—0.01427	0.03317	—0.02955	0.01499	0.03933	0.02409	0.07625	0.07881	0.04085	0.03192
10 A10	0.10969	0.11397	0.04367	0.11147	0.14505	0.12274	0.11873	0.09849	0.05379	0.09664	0.12158	0.13427
11 B1	—0.04424	—0.06230	—0.06691	—0.02547	0.01198	—0.02654	—0.03739	0.00298	—0.02184	0.02103	—0.01262	—0.04763
12 B2	0.03128	0.00589	0.02851	0.03642	0.03442	0.01488	0.02391	0.00345	—0.02444	0.02195	0.02559	0.00851
13 B3	—0.07457	—0.07630	0.00071	—0.05615	—0.06677	—0.05511	—0.04049	—0.05953	—0.04298	—0.04657	—0.08056	0.02446
14 B4	—0.01246	—0.00333	0.04589	—0.01537	—0.00641	0.00813	0.00200	—0.01187	0.05694	0.00565	—0.00871	0.01347
15 B5	0.01380	0.11677	—0.04351	—0.01810	—0.02526	0.03367	0.00326	—0.02882	—0.04032	0.01475	0.05656	—0.00174
16 B6	0.07213	—0.03277	0.02718	0.03269	0.02222	0.01292	0.04407	0.04107	0.03433	0.04132	0.02000	0.03626
17 B7	—0.01657	—0.00020	0.04436	—0.01430	—0.01490	—0.01434	—0.02723	—0.03033	—0.03549	—0.00480	—0.02592	—0.00412
18 B8	0.03745	0.06644	0.04080	0.06896	0.05535	0.05525	0.05538	0.04021	0.04213	0.00052	0.05441	0.02175
19 B9	—0.01145	—0.06711	—0.02033	—0.01456	0.01017	—0.01456	—0.02001	0.01861	—0.01210	—0.05489	—0.00438	—0.03497
20 B10	0.01282	0.01532	—0.01467	0.02613	0.01189	0.01862	0.04887	0.05099	0.00731	0.03576	0.01851	0.03654
Constante	—12.95126	—12.52949	—11.73568	—12.09709	—12.07790	—12.34907	—12.65112	—12.56970	—12.50254	—13.26179	—12.32662	—13.98882

Una vez que se han construido las doce funciones de clasificación, el nuevo sujeto se asigna a la función que tenga el valor más alto, ya que este valor máximo nos indica que el individuo pertenece a ese grupo.

4.2. Método de clasificación cuadrática para grupos con matrices de covarianza diferentes

Cuando los datos no poseen equicovarianza se aconseja utilizar la discriminación cuadrática. Smith en su artículo «Some examples of discrimination» publicado en *Annals of Eugenics* (1947) es el primero en considerar dos poblaciones con matrices de covarianza diferentes bajo la hipótesis de normalidad y mostrar que la función discriminante mejor es la cuadrática. En fechas más recientes merecen destacarse, entre otros, los estudios de Hubarty (1978), Lachenbrunch (1975), Sebestyen (1973) i Romeder (1973), quienes generalizan la discriminación cuadrática bajo la hipótesis de normalidad de dos grupos al multigrupo, de tal modo que x se asigna al grupo i si se verifica que

$$(x - \bar{y})'\Sigma^{-1}(x - \bar{y}_i) + \ln |\Sigma^{-1}| < (x - \bar{y}_i)' \Sigma^{-1} (x - \bar{y}_i) \ln |\Sigma^{-1}|$$

Romeder propone para el programa SEB-2, que se aplica en nuestra investigación, la ecuación fundamental de la semejanza (1973).

$$S(x,y) = |\Sigma|^{4} [(x - \bar{y})' \Sigma^{-1} (x - \bar{y}) + P]$$

Interesa señalar que el enfoque de Romeder tiene en cuenta variables discretas y continuas y admite funciones secundarias curvilíneas e irregulares. Se emplea la validación de los resultados, mediante la división de la muestra en dos subtests. Siguiendo Sebestyen (1973), Cover (1965) y Romeder (1973), se elige el 75 % para la muestra base y el 25 % para la muestra test. Con los datos así divididos se confecciona el algoritmo que clasifica a los sujetos en los grupos de dos en dos, y en todos los grupos en conjunto.

5. El problema de la selección de las variables. Bondad de los clasificadores

Como en la investigación se poseen muchas variables conviene *seleccionar* sólo aquellas que contribuyen a la discriminación entre los grupos. Un medio de eliminar las variables innecesarias y seleccionar las variables discriminantes más útiles cada vez, es usar el procedimiento paso a paso.

La mayor parte de los programas de selección del modelo *paso a paso* utilizados en la discriminación lineal requieren que cada variable que vaya a incluirse posea ciertas condiciones mínimas antes de que se seleccionen. Estas condiciones son la prueba de tolerancia (t_r) para garantizar la corrección del cálculo y un estadístico F parcial para asegurar que el incremento de la disciminación exceda cierto nivel establecido previamente. En el programa se asigna un umbral de $F = 4$ y de $t_r = 0.10$.

El método paso a paso combina el modelo hacia adelante que tiene en cuenta el valor de la «F to enter» y el hacia atrás que trabaja con la «F to remove». La *«F to enter»* es una prueba parcial multivariante que contrasta la significatividad de la discriminación adicional introducida por la variable que está siendo considerada, después de tener en cuenta la discriminación alcanzada por las variables ya introducidas. *La F to remove* es otro estadístico F parcial que prueba si hay algunas variables introducidas que ya no proporcionan una amplia contribución única a la discriminación, de tal modo que convenga suprimirlas de la selección de las variables ya realizada. El estadístico *U lambda de Wilks* mide el poder de discriminación de las variables seleccionadas. En la tabla 3 se presentan la selección de las variables de la discriminación lineal siguiendo el método paso a paso. Para ilustrar el proceso en su conjunto muestro el paso n.º 1.6 y 20.

TABLA 3.—*Selección de variables por el método paso a paso en la discriminación lineal*

Paso número 1
Variable introducida 15 B5

Variable	F to remove DF=11 1069	Fuerza del nivel
15 85	16.296	1

Variable	F to enter DF=11 1068	Fuerza del nivel	Tolerancia
1 A1	2.806	1	0.973617
2 A2	1.889	1	0.957076
3 A3	5.074	1	0.956746
4 A4	4.145	1	0.924680
5 A5	1.225	1	0.745308
6 A6	6.047	1	0.937837
7 A7	3.766	1	0.926937
8 A8	4.815	1	0.873924
9 A9	4.934	1	0.900271
10 A10	5.089	1	0.939177
11 B1	5.816	1	0.925159
12 B2	2.527	1	0.904981
13 B3	5.791	1	0.825842
14 B4	4.885	1	0.926094
16 B6	* 11.372	1	0.920870
17 B7	3.452	1	0.827694
18 B8	5.752	1	0.813408
19 B9	9.790	1	0.795000
20 B10	1.017	1	0.889681

ESTADISTICO U DE WILKS' LAMBDA 0.8563944
ESTADISTICO APROXIMADO DE F 16.296
GRADOS DE LIBERTAD 1 11 1069
GRADOS DE LIBERTAD 11.00 1069.00

Paso número 6
Variable introducida 19 B9

Variable	F to remove DF=11 1064	Fuerza del nivel
3 A3	5.501	1
11 B1	6.068	1
15 B5	24.613	1
16 B6	6.928	1
19 B9	9.863	1
20 B10	5.371	1

Variable	F to enter DF=11 1063	Fuerza del nivel	Tolerancia
1 A1	3.337	1	0.692267
2 A2	0.918	1	0.643346
4 A4	2.761	1	0.720834
5 A5	2.910	1	0.567232
6 A6	1.768	1	0.559364
7 A7	4.675	1	0.811740
8 A8	1.942	1	0.672025
9 A9	2.542	1	0.549756
10 A10	1.944	1	0.570666
12 B2	1.619	1	0.789008
12 B3	4.163	1	0.681029
14 B4	1.010	1	0.790650
17 B7	* 5.147	1	0.766143
18 B8	2.432	1	0.598329

ESTADISTICO U DE WILKS' LAMBDA 0.6135413
ESTADISTICO APROXIMADO DE F 8.264
GRADOS DE LIBERTAD 6 11 1069
GRADOS DE LIBERTAD 66.00 5698.75

Paso número 20
Variable introducida 6 A6

Variable	F to remove DF=11 1050	Fuerza del nivel
1 A1	2.718	1
2 A2	1.826	1
3 A3	4.232	1
4 A4	1.750	1
5 A5	4.119	1
6 A6	1.461	1
7 A7	1.465	1
9 A9	1.566	1
8 A8	3.297	1
10 A10	2.693	1
11 B1	3.928	1
12 B2	1.923	1
13 B3	3.378	1
14 B4	1.820	1
15 B5	11.579	1
16 B6	3.894	1
17 B7	2.087	1
18 B8	1.984	1
19 B9	3.589	1
20 B10	1.492	1

Variable	F to enter DF=11 1049	Fuerza del nivel	Tolerancia

ESTADISTICO U DE WILKS' LAMBDA 0.4279578
ESTADISTICO APROXIMADO DE F 4.245
GRADOS DE LIBERTAD 20 11 1069
GRADOS DE LIBERTAD 220.00 10195.37

En el paso n.º 1 se ha introducido la variable 15 con un valor de «F to enter» = 16.296 que conserva en la «F to remove» con una U lamb de Wilks = 0.8563944 y el estadístico aproximado de F = 16.296. Se marca con asterístico la variable 11 cuya «F to enter» = 11.372 que le sigue en valor y se introducirá en el paso n.º 2. El paso n.º 6 ilustra cómo la variable 15 al combinarse con otras incrementa su poder discriminante hasta 24.013. En el paso 20 se han intoducido todas las variables siendo la última introducida la 20 con una «F to enter» = 1.492 y un estadístico aproximado de F = 4.245. En la tabla 7 muestro el resumen final.

El cuadro resumen nos ofrece el orden de las variables introducidas, el valor de la «F to enter or remove», el del estadístico U, la F aproximada y su comparación con la tabla de Snedecor. Cuando todas las variables están incluidas la «F to enter or remove» se usa para calcular el orden del rango del poder disciminante único proporcionado por cada una de las variables seleccionadas. Los estadísticos F obtenidos al compararlos con las F de la tabla de Snedecor con sus grados de libertad correspondientes resultan ser todos significativos al nivel de confianza del 0.05. En el programa se han incluido las veinte variables aunque el umbral de F = 4. Si se hubiese mantenido este nivel sólo se habían introducido ocho variables, en concreto, la B_5, B_6, B_1, B_{10}, A_3, B_9, B_7 y B_3 que en la tabla se marca con una línea.

La *discriminación cuadrática* realizada con el programa SEB-2, Romeder (1973) sigue el método paso a paso eligiendo en cada paso la variable que maximiza la semejanza de cada sujeto a cada uno de los centroides de los grupos y que maximiza el porcentaje de los individuos bien clasificados en su grupo correspondiente: después de elegir la variable que mejor discrimina, se selecciona entre las 19 restantes la que más diferencia y se repite todo el proceso, recalculando una función cuadrática que dependerá de las matrices de covarianza de las dos variables más discriminantes. Análogamente, se sigue con todas las variables hasta que estén todas introducidas (paso 20). En la investigación, este proceso se sigue con todas las parejas de grupos, es decir, 66 veces, y se completa con un estudio global de los 12 grupos a la vez.

La bondad de los clasificadores depende de la significatividad de las funciones discriminantes canónicas y de clasificación. Para contrastar su significación pueden utilizarse diversas pruebas entre las que destacan el *porcentaje relativo de los autovalores*, la correlación canónica y el porcentaje de bien clasificados. La función de mayor autovalor sabemos que es la variable canónica con mayor poder discriminador. Pero los autovalores no pueden interpretarse directamente, sino que deben hallarse sus porcentajes relativos mediante la fórmula

TABLA 4.—*Resumen*

Número de pasos	Variable introducida	F value to enter or remove	Número de variables incluidas	Estadístico U	Estadístico aproximado de F
1	15 B5	10.2460	1	0.6564	16.296>F 11, 1069 = 1.80
2	16 B6	11.3722	2	0.7606	13.799>F 22, 2136 = 1.58
3	11 B1	6.3207	3	0.7197	11.254>F 33, 3144 = 1.46
4	20 B10	5.7855	4	0.6792	9.869>F 44, 4080 = 1.40
5	3 A3	5.2240	5	0.6444	8.933>F 55, 4939 = 1.35
6	19 B9	4.9631	6	0.6135	8.254>F 66, 5698 = 1.35
7	17 B7	5.1470	7	0.5825	7.816>F 77, 6377 = 1.28
8	13 B3	4.1605	8	0.5584	7.363>F 88, 6973 = 1.24
9	1 A1	3.0308	9	0.5414	6.880>F 99, 7493 = 1.24
10	5 A5	2.4304	10	0.5281	6.433>F 110, 7945 = 1.24
11	8 A8	3.5026	11	0.5096	6.172>F 121, 8337 = 1.24
12	4 A4	2.6403	12	0.4960	5.880>F 132, 8676 = 1.24
13	10 A10	2.3332	13	0.4842	5.608>F 148, 8970 = 1.24
14	18 A8	2.4811	14	0.4720	5.388>F 154, 8925 = 1.24
15	9 A9	1.6377	15	0.4641	5.137>F 165, 9447 = 1.24
16	2 A2	1.5947	16	0.4565	4.910>F 176, 9639 = 1.17
17	12 B2	1.6803	17	0.4486	4.726>F 187, 9807 = 1.17
18	14 B4	1.5447	18	0.4415	4.549>F 198, 9954 = 1.17
19	7 A7	1.5320	19	0.4345	4.391>F 209, 10082 = 1.17
20	6 A6	1.4612	20	0.4280	4.245>F 220, 10195 = 1.11

$$\% \text{ rel. } \lambda_i = \frac{(\lambda_i)}{\sum^{k} \lambda_i} \qquad \lambda_i = \text{autovalor de la función i}$$

Otra forma de juzgar la aportación sustantiva de las funciones discriminantes consiste en aplicar los *coeficientes de correlación canónica.* Estos coeficientes, que ofrecen el grado de asociación entre los grupos y las funciones discriminantes, se calculan con $R_{ci} = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}$. En la tabla 8 muestro los autovalores correspondientes a cada función discriminante, los porcentajes relativos de los autovalores y las correlaciones canónicas. Como puede observarse sólo las dos primeras funciones discriminantes merecen tenerse en cuenta.

TABLA 5.—*Bondad de clasificadores. Discriminación lineal*

Autovalores, sus porcentajes relativos y la correlación canónica

Funciones discriminantes canónicas	Autovalores	Porcentaje relativo	Correlación canónica
1	0'41869	44'89 *	0'54325 *
2	0'14953	16'03 *	0'35067 *
3	0'11199	12	0'31735
4	0'07304	7'83	0'26090
5	0'05884	6'30	0'23572
6	0'05153	5'52	0'22138
7	0'03167	3'39	0'17521
8	0'01814	1'94	0'13347
9	0'01108	1'18	0'10467
10	0'00812	0'87	0'08974
Total →	0'93263		

$$\% \lambda_i = \frac{\lambda_i}{\sum^{k} \lambda_k}, \quad \% \lambda_1 = \frac{0.41869}{0.93263} = 44.89$$

$$R_{ci} = \sqrt{\frac{\lambda_i}{1+\lambda_i}}, \quad R_{c1} = \sqrt{\frac{0.41869}{1+0.41869}} = 0.54325$$

En cuanto al *porcentaje de bien clasificados* es otra forma de validar el método de discriminación para el que se ha definido un proceso de clasificación. A cada individuo se le asigna al grupo a cuyo centroide tiene la mínima distancia cuadrática. La Proporción de los casos correctamente clasificados indica la validez del procedimiento. En la tabla 9 muestro la matriz de clasificación lineal. Como puede observarse, el porcentaje de bien clasificados asciende sólo al 31 %.

TABLA 6.—*Matriz de clasificación lineal*

Grupo	Porcentaje correcto	Número de casos casificados dentro de cada grupo											
		Arqui.	Quim.	Pilot.	Medic.	Farma.	Econo.	Juris.	Profe.	Period.	Admin.	Comer.	Ofici.
Arqui.	26.8	11	2	5	2	2	0	3	3	7	0	4	2
Quimi.	48.1	2	74	12	6	0	7	3	2	4	6	15	15
Pilot.	34.5	0	2	10	1	1	0	1	0	4	5	0	1
Medic.	20.5	4	5	7	16	0	1	8	10	3	8	6	2
Farma.	20.0	4	3	3	2	7	2	3	3	8	3	1	1
Econo.	14.2	9	12	9	11	7	16	12	4	9	6	11	7
Juris.	20.2	13	2	6	5	7	3	20	8	9	1	7	16
Profe.	20.6	10	8	11	12	13	7	20	37	23	11	8	14
Period.	57.5	2	4	3	5	1	3	2	7	42	1	1	2
Admin.	31.5	2	5	7	6	4	8	2	7	10	34	11	6
Comer.	31.3	7	14	6	8	6	9	3	3	3	20	40	9
Ofici.	51.3	2	1	3	0	0	2	3	1	2	4	1	20
TOTAL	31.0	74	132	90	80	64	58	80	85	119	99	105	95

En la tabla 10 presento el porcentaje de bien clasificados mediante las funciones de clasificación cuadrática en todos los grupos juntos que alcanza al 83.49 %.

Tabla 7.—*Matriz de la clasificación cuadrática de todos los grupos*

PASO NUMERO: 20

VARIABLES INTRODUCIDAS									
15	11	8	13	16	7	3	19	20	17
14	4	2	1	9	12	5	10	6	18

TABLA DE CLASIFICACION DE LA MUESTRA BASE

Grupo de origen		Grupo de afectación 1	2	3	4	5	6	7	8	9	10	11	12	% parciales
Arqui.	1	30	0	0	1	0	0	0	0	0	0	0	0	96.77 *
Quim.	2	0	108	0	2	1	0	3	0	0	0	2	0	93.10 *
Pilot.	3	0	0	24	0	0	0	1	0	0	0	0	0	96.00 *
Med.	4	3	0	0	50	0	0	2	1	1	0	1	1	84.75 *
Farm.	5	1	0	0	0	25	0	1	0	0	0	0	0	92.59 *
Econom.	6	14	0	8	10	3	20	24	0	1	0	4	1	23.53
Juris.	7	8	0	0	0	1	0	65	0	0	0	0	0	87.84 *
Profes.	8	10	5	4	5	5	10	10	20	3	2	4	2	59.26 *
Period.	9	2	0	0	3	0	0	0	0	50	0	0	0	94.55 *
Admin.	10	2	0	0	2	2	3	10	1	0	70	0	0	86.42 *
Emp. prop.	11	4	2	0	0	0	0	0	0	0	0	90	1	93.75 *
Ofic.	12	1	0	0	0	0	0	0	0	0	0	1	28	93.33 *

PORCENTAJE DE CASOS BIEN CLASIFICADOS: 83.49

Se advierte que todos los grupos tienen un porcentaje muy elevado de sujetos bien clasificados excepto los economistas (23.53) y profesores (52.26).

La tabla 11 resume los porcentajes de bien clasificados calculados mediante las funciones de clasificación para cada pareja de grupos. Puede apreciarse que los porcentajes oscilan entre el 72.32 % de bien clasificados en los arquitectos y los oficiales de las fuerzas armadas y el 100 % entre los arquitectos y los pilotos.

Se juzga la magnitud del porcentaje de bien clasificados en relación al porcentaje esperado de las clasificaciones correctas si la asignación se hubiera hecho de forma aleatoria. Si tuviéramos dos grupos esperaríamos obtener el 50 % de predicciones adecuados por pura aleatoriedad. Como en el caso al que he hecho referencia contamos con 12 grupos, la asignación aleatoria correcta a priori vendrá dado por $pi = \frac{100}{12}$ $= 8'33$ %. La *reducción proporcional de los errores* nos la proporciona la medida de asociación τ (tau). La compara la predicción proporcional aleatoria de un índice con la predicción proporcional condicional basada en el conocimiento de otro índice. El valor máximo de τ es 1 y ocurre cuando no existen errores en la predicción. Un valor de 0 indica que no hay incremento predictivo. Los resultados negativos indicarían una situación degenerativa. En el caso estudiado para la clasificación de la discriminación lineal

$$\tau = \frac{n_c - \sum^{6} pini}{N - \sum^{6} pini} = \frac{(11 + 74 + 18 + 16 \ldots + 20) - \sum^{12}[(1/12 \cdot 41) + (1/12 \cdot 154) + \ldots (1/12 \cdot 20)]}{1081 - \sum^{12}[(1/12 \cdot 41) + (1/12 \cdot 154) + \ldots + (1/12 \cdot 20)]} = 0 \cdot 25$$

Esto significa que la *clasificación lineal comete 25 % menos* de errores de los que se esperarían por la asignación aleatoria. En el estudio realizado con la métrica cuadrática tendremos

$$\tau \text{ (todos los grupos)} = \frac{n_c - \sum^{6} pini}{N - \sum^{6} pini} = \frac{(30 + 108 + \ldots + 90 + 28) - (1/12 \cdot 31 + 1/12 \cdot 116 + \ldots 1/12 \cdot 30)}{814 - (1/12 \cdot 31 + 1/12 \cdot 116 + \ldots 1/12 \cdot 30)}$$

Es decir, que se cometen el *77 % menos de errores con la discriminación cuadrática aplicada a todos los grupos.*

τ (cada dos grupos):

$\tau_{1\,y\,2} = 0.57$; $\tau_{1\,y\,3} = 1.00$; $\tau_{1\,y\,4} = 0.87$; $\tau_{1\,y\,5} = 0.83$; $\tau_{1\,y\,6} = 0.63$; $\tau_{1\,y\,7} = 0.84$; $\tau_{1\,y\,8} = 0.44$; $\tau_{1\,y\,9} = 0.60$; $\tau_{1\,y\,10} = 0.45$; $\tau_{1\,y\,11} = 0.54$; $\tau_{1\,y\,12} = 0.96$.

Tabla 8.—*Matriz de clasificación cuadrática de cada dos grupos*

Grupos nº	Paso	Variables introducidas	Muestra base N	N	T	Muestra test N	N	T
1 y 2	6	11, 13, 2, 9, 16, 15	87 %	77 %	78 %	70 %	76 %	75 %
1 y 3	13	17, 5, 9, 7, 10, 11, 13, 2, 1, 8, 14. 3, 20	100 %	100 %	100 %	70 %	50 %	61 %
1 y 4	10	5, 20, 18, 16, 12, 3, 11, 10, 13, 7	93’5 %	93’2 %	93’33 %	60 %	84’5 %	79’31 %
1 y 5	9	1, 6, 16, 19, 3, 10, 2, 1, 5, 4	96’7 %	85’19 %	91’38 %	90 %	25 %	61’11 %
1 y 6	12	19, 10, 3, 2, 20, 19, 14, 16, 7, 11, 13	100 %	75 %	81’90 %	10 %	85’7 %	65’79 %
1 y 7	18	15, 12, 10, 20, 19, 18, 7, 14, 11, 3, 13, 4, 16, 5, 6, 17, 9, 8	100 %	89’19 %	92’38 %	80 %	68 %	71’43 %
1 y 8	3	3, 20, 19	16’13 %	85’93 %	72’89 %	20 %	91 %	78’18 %
1 y 9	7	15, 13, 12, 17, 2, 14, 3	93’55 %	72’73 %	80’23 %	90 %	61’11 %	71’43 %
1 y 10	1	12	19’35 %	92’59 %	72’32 %	10 %	96’30 %	72’97 %
1 y 11	18	8, 2, 9, 3, 20, 4, 14, 15, 12, 6, 7, 13, 10, 11, 17, 5, 1, 9	100 %	69’79 %	77’17 %	80 %	78’13 %	78’57 %
1 y 12	10	13, 15, 12, 10. 3, 7, 6, 4, 8, 9	100 %	96’67 %	98’36 %	90 %	77’78 %	84’21 %

Es decir, que se cometen desde el *44 % menos de errores hasta incluso ningún error en la discriminación cuadrática* aplicada a cada dos grupos ocupacionales.

6. Conclusiones y sugerencias

Del análisis discriminante aplicado al estudio de los veinte índices de valor evaluados por el «Test de Reacción Valorativa» en las doce ocupaciones, caben deducir las siguientes conclusiones:

1) Del análisis discriminante lineal se obtienen nueve funciones discriminantes pero sólo las dos primeras son sustantivas para la discriminación de los doce grupos según las áreas axiológicas, como muestran los porcentajes relativos de los autovalores (44.894 y 16.03) y las correlaciones canónicas entre los grupos y las funciones discriminantes (0.54 y 0.35).

2) Las funciones de clasificación lineal presentan sólo el 31 % de sujetos bien clasificados, número que es evidentemente insuficiente.

3) La correlación τ nos ofrece que la clasificación lineal comete el 25 % menos de errores de los que se esperarían por asignación aleatoria.

4) De estos datos cabe deducir que el análisis discriminante lineal está distorsionado por la violación de la asunción de la equicovarianza de los grupos.

5) Las doce ocupaciones estudiadas se discriminan en función de los valores medidos por el «Test de Reacción Valorativa» si se aplica un análisis de discriminación cuadrática.

6) Los resultados hallados con el análisis discriminante cuadrático, al no exigir la equicovarianza de los grupos, muestra un elevado poder de clasificación, alcanzando el 84.49 % y el 100 % de los sujetos bien clasificados.

7) La correlación τ nos proporciona que la clasificación cuadrática comete el 77 % menos de errores que los que se esperarían por asignación aleatoria si se emplea con todos los grupos e incluso ningún error si se hace uso de ella con cada pareja de grupos.

8) Se recomienda que el diagnóstico axiológico se integre en el diagnóstico y pronóstico vocacional y, por tanto, en el proceso de la Orientación Educativa.

No quisiera terminar sin aludir a otros problemas pedagógicos que pueden resolverse mediante los diseños discriminantes. Es evidente que no voy a hacer ahora una enumeración exhaustiva de los mismos, sino sólo mencionar los grandes campos de la Orientación escolar y vocacional a los que pueden aplicarse.

Concebida la Orientación Educativa desde un enfoque sistémico y con un carácter individualizado dentro de un contexto colectivo, son diversos los subsistemas que interaccionan con el subsistema pedagógico y condicionan a los educandos, así como los paradigmas de tratamientos orientadores y los grupos diferenciados de alumnos que existen en la realidad pedagógica y que pueden estudiarse en los análisis discriminantes lineal y cuadrático. A título indicativo, los diseños de discriminación cabe focalizarlos en el análisis de las estructuras del subsistema sociológico y en el acoplamiento o asignación de los alumnos con las intervenciones orientadoras. En efecto, la naturaleza y la dinámica del subsistema sociológico, por las limitaciones que imponen en la conducta de los educandos o las facilidades que le proporcionan, son tópicos relevantes para la investigación con estas técnicas. De esta suerte, cabe emplear el análisis discriminante descriptivo para conocer los perfiles diferenciados y las pautas orientadoras de los grupos socio-culturales, étnicos, familiares, de sexo, u ocupaciones, entre otros. Pero también, la concepción individualizada de la educación en un contexto colectivo conlleva la identificación de los modelos orientadores y de las clases diversificadas de educandos para obtener reglas de asignación que resuelva el tema permanente de la clasificación de alumnos. Clasificación entendida no como selección y promoción de los mejores, sino como estímulo para el aprendizaje máximo de todos los escolares por muy diversos que sean sus modelos de aptitudes o los rasgos de su personalidad.

Si nos apoyamos en las investigaciones con las técnicas de clasificación lineal y sobre todo cuadrática, podremos aplicar intervenciones orientadoras de objetivos múltiples a los grupos de educandos que posean, por ejemplo, diferentes estilos cognitivos, locus de control, motivaciones, estrategias para la toma de decisiones o constructos personales, por aludir a algunas de las nuevas áreas de investigación dentro de la Orientación escolar y vocacional.

Como investigadores pedagógicos estamos comprometidos en realizar trabajos que mejoren la praxis educativa. Estimo que, una de las vías para hacer realidad este compromiso reside en el estudio discriminante de las categorías de los modelos orientadores y de los perfiles de diagnóstico de los educandos con objeto de acoplar las intervenciones orientadoras que interaccionen más eficazmente con las aptitudes y rasgos personales de los alumnos y que, por tanto, conlleven los cambios de conductas deseadas y el desarrollo positivo de su personalidad.

BIBLIOGRAFIA

ANDERSON, T. W. (1958) *An introduction to multivariate statistical analysis* (New York, John Wiley).

BOX, G. E. P. (1949) A general distribution theory for a class of likelihood criteria, *Biometrika,* 36, pp. 317-346.

BOX, G. E. P. y COX, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society,* B, 26, pp. 211-252.

CACOULLOS, D. T. (1973) *Discriminant Analysis and Applications* (New York, Academic Press).

COVER, T. M. (1965) Geometrical and statistical properties of systems of linear inequalities with applications in Pattern Recognition, *IEE Transactions on electronic computers,* pp. 326-334.

HAWKINS, D. M. (1976) The subset problem in multivariate analysis of variance, *Journal of the Royal Statistical Society,* B, 38, pp. 132-139.

— (1980) *Identification of Outlers* (London, Chapman and Hall).

— (1981) A new test for multivariate normality and homosced sticity, *Technometrics,* 23, pp. 105-110.

— (1983) *Topics in Applied Multivariate Analysis* (Cambridge, Cambridge Universiting Press).

LACHEMBRUCH, P. A. (1967) An almost unbiased method of obtaining confidence inter vals for the probability for misclassification in discriminant analysis, *Biometrics,* 23, pp. 639-645.

— (1968) On expected probabilities of misclassification in discriminant analysis, necessary sample, size, and a relation with the multiple correlation coefficient *Biometrics,* pp. 823-834.

— (1975) *Discriminant Analysis* (New York, Hafner Press).

MAY, R. W. (1982) «Discriminant Analysis» en Hudson, H. C.: *Classifyng Social Data* (San Francisco, Jossey Bass Publications).

MORRISON, D. G. (1969) On interpretation of discriminant analysis, *Journal of Marketing Research,* 6, pp. 156-163.

— (1974) Discriminant analysis, pp. 2.442-2.457 en R. FEBER (ed.) *Handbook of Marketing Research* (New York, John Wiley).

RAO, C. R. (1952) *Advanced Statistical Methods in Biometric Research* (New York, John Wiley).

ROMEDER, J. M. (1973) *Methodes et programmes d'analyses dîscriminante* (París, Dunod).

SEBESTYEN, G. S. (1962) *Decision marking processes in pattern recognition* (New York, The MacMillan Company).

— (1973) Programme SEB-2 en ROMEDER, J. M. (1973), o.c., pp. 134-154.

TATSUOKA, M. M. y TIEDEMAN, D. V. (1954) Discriminant analysis, *Review of Educational Research,* 24, pp. 402-420.

THORNDIKE, R. M. (1978) *Correlation Procedures for Research* (New York, Gardner Press).

TIMM, N. H. (1975) *Multivariate Analysis with Applications in Education and Psychology* (Monterrey. Calif. Brooks/Cole.).

SUMARIO: La condición diferenciada de los educandos y la identificación de los múltiples factores educativos en diversas categorías, viene impuesta por la misma realidad de la educación. De aquí que en las investigaciones sobre la Orientación Educativa ocupe un lugar destacado el estudio discriminante descriptivo de las diferentes clases en que aparecen agrupados los orientandos y más aún la clasificación de los nuevos orientandos a aquellos grupos a los que más se asemejan. Este método es el que se emplea en el presente trabajo para resolver un doble tipo de problema: ¿las dimensiones axiológicas discriminan las ocupaciones?; ¿cabe obtener funciones de clasificación valorativa ocupacional que se aplique a los alumnos en el diagnóstico y pronóstico profesional en los que se apoya la Orientación Educativa? Las variables son doce familias ocupacionales y diez áreas de valor. Las familias ocupacionales, clasificadas según los criterios de la O.I.T. son: Arquitectos, Químicos, Pilotos, Médicos, Farmacéuticos, Economistas, Juristas, Profesores, Periodistas Administraivos Comerciantes propietarios y Oficiales de las Fuerzas Armadas. Las diez áreas axiológicas son: dinero fama, poder, placer, actividad negociante, artística, técnica, intelectual, social y trascendental. Los instrumentos de exploración empleados son el «Test de Reacción Valorativa» del Dr. García Hoz y un cuestionario ocupacional. De la aplicación del análisis discriminante lineal se deduce que dado que no cumplen los datos la asunción del modelo sobre la equicovarianza de los grupos, su poder de descriptivo y de clasificación es bajo. Por el contrario, al aplicar la discriminación cuadrática se aprecia un alto poder de clasificación. Se concluye que el diagnóstico axiológico se integre en el diagnóstico y pronóstico profesional, y por tanto en la Orientación Educativa.

Descriptores: Occupations, Values, Vocational Choice, Vocational guidance, Discriminant analysis, Work and values.