

## V. MODELO DESARROLLADO

Antes de determinar un modelo que pudiese utilizarse en el análisis multivariable, fue necesario obtener varias tablas de contingencia, así como efectuar otros estudios de las distribuciones de frecuencia.<sup>14/</sup> El objetivo fue comprobar la validez de las hipótesis generales que se han propuesto y verificar que los datos fuesen adecuados, especialmente desde un punto de vista cuantitativo.<sup>15/</sup>

Como la información abarca datos de variables categóricas (sí, no); nominales (status ocupacional, etc.); escala y atributo (datos de opinión, educación del padre, etc.); el control de todas ellas en las ecuaciones de regresión se efectuó a través de variables "dummy".<sup>16/</sup>

En este estudio la calificación del estudiante es la variable dependiente, una calificación diferente es asociada con una combinación de subgrupos de las variables independientes. Para facilitar

<sup>14/</sup> Bajo este procedimiento se contó con el mayor número de variables que intervinieran en la educación y que fueran no triviales. Principalmente, a través de diferencias de medias por una "prueba-t".

<sup>15/</sup> Se sabe que la operatividad de cualquier ecuación consiste en tener un número mayor de datos observados que de coeficientes a estimar. Mayor lectura matemática se encuentra en: Johnston, J. *Econometric Methods*. Mc. Graw-Hill Ed., 1972.

<sup>16/</sup> La variable "DUMMY" es utilizada como herramienta econométrica en regresión, asigna valores 1 si la unidad de análisis (estudiante) fue miembro de una categoría dentro de cada variable en particular y cero si no. Ver Johnston, J. Op. Cit. pp. 176.

la notación en el análisis de regresión, se anticipan las siguientes tres ecuaciones:

$$(a) X_1 = \sum_{d=1}^{6,8} Z_d + \sum_{e=2}^{3,4} Z_e + \sum_{f=2}^{3,4} Z_f ; X_1 = \sum_{d,e,f} Z$$

$$(b) X_2 = \sum_{g=2}^3 Z_g + \sum_{h=2}^3 Z_h + \sum_{i=2}^2 Z_i ; X_2 = \sum_{g,h,i} Z$$

$$(c) X_3 = \sum_{j=1}^1 Z_j + \sum_{d=1}^1 Z_k + \sum_{l=1}^1 Z_l + \sum_{m=2}^2 Z_m + \sum_{n=2}^4 Z_n ; X_3 = \sum_{j..n} Z$$

donde:

$X_1$  = Factores del medio ambiente familiar

$X_2$  = Factores del medio ambiente escolar

$X_3$  = Factores personales del alumno

$Z_d$  a  $Z_n$  = Ver definición en el Cuadro 4.

De (1), (2) y (3) se obtienen las ecuaciones de regresión a estimar, sus resultados se analizarán adelante:

$$(4) Y_a = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$(5) Y_b = \alpha + \beta_1' X_1' + \beta_2' X_2' + \beta_3' X_3' + u$$

$$(6) Y_c = \alpha + \beta_1'' X_1'' + \beta_2'' X_2'' + \beta_3'' X_3'' + u$$

donde:

$Y_a$  = Calificación promedio en FACPYA

$Y_b$  = Calificación promedio en FIME

$Y_c$  = Calificación promedio en MEDICINA

$X_1$  = Factores del ambiente familiar, FACPYA

$X_2$  = Factores del ambiente escolar, FACPYA

$X_3$  = Factores personales del alumno, FACPYA

$X'_1$  = Factores del ambiente familiar, FIME

$X'_2$  = Factores del ambiente escolar, FIME

$X'_3$  = Factores personales del alumno, FIME

$X''_1$  = Factores del ambiente familiar, Medicina

$X''_2$  = Factores del ambiente escolar, Medicina

$X''_3$  = Factores personales del alumno, Medicina

Variables omitidas en el análisis de regresión.<sup>17/</sup> (Ver Cuadro 4.)

de  $X_1$  ;  $Z_d = 7$ ,  $Z_e = 1$  y  $Z_f = 1$

de  $X_2$  ;  $Z_g = 1$ ,  $Z_h = 1$  y  $Z_i = 1$

de  $X_3$  ;  $Z_j = 2$ ,  $Z_k = 2$ ,  $Z_l = 2$ ,  $Z_m = 1$  y  $Z_n = 1$

Básicamente son 11 variables independientes integradas en tres paneles ( $X_1$ ,  $X_2$  y  $X_3$ ), dentro de ellas existen 36 sub-clases definidas; 16 para el primer panel (8 para el status ocupacional y para la educación del padre y de la madre 4 cada uno); 8 para el segundo panel, (3 para utilidad de material de apoyo, 3 en opinión de bibliotecas y 2 en rigor académico); finalmente otras 12 integran el tercer panel

<sup>17/</sup> Para evitar que la matriz sea singular debido a la dependencia lineal entre las dummies usadas en cualquier tipo de variables, una variable dummy se omite en cada grupo de categorías. Ver Johnston, J. Op. Cit., pp. 176

(sexo, trabaja, satisfacción con calificación y actividades escolares, cada una con 2 y actividades extraescolares con 4).

Las ecuaciones (4), (5) y (6) contienen a " $\alpha$ ", que es el término constante en la regresión o "coeficiente de intercepción diferencial" porque explica la diferencia en la calificación proveniente de la omisión de las categorías en las variables antes mencionadas. Del mismo modo, las variables dentro de un grupo en especial, como por ejemplo educación del padre, tienen su base de comparación de la calificación media con respecto al grupo de educación específica que es incorporado en el término constante (" $\alpha$ ").

Específicamente, los coeficientes de regresión de las variables independientes omitidas, son iguales a cero. Los coeficientes de regresión de las restantes variables independientes ( $\beta$ 's) son leídos en términos de puntos obtenidos por el estudiante (1 a 100) y son aditivos. Este número absoluto representa un grado de influencia, mientras que el signo asociado a él ( $\pm$ ) muestra la dirección positiva o negativa de dicha influencia. El error de regresión es " $u$ ".

Los grupos base incorporados en el término constante son los que se indican con asterisco en el Cuadro 4 (ejemplo, en el status ocupacional sería la categoría 7. "Obreros") y son los mismos para todas las ecuaciones que tratan en forma separada a FACPYA, FIME y Medicina.

Un supuesto primordial de este enfoque estadístico es que todas las categorías tienen la misma ponderación, es decir, las categorías ocupacionales y educación del padre, de la madre, etc. se supone que son idénticos entre escuelas (igual número de individuos en cada categoría). Es obvio que estos supuestos no son reales y por lo tanto los resultados del análisis deben ser interpretados con cautela. Sin embargo, la característica de igual ponderación de este proceso estadístico tiene la ventaja de indicar el efecto "puro" de cada variable sobre la calificación del estudiante en cada escuela, en el sentido de que un coeficiente de regresión para un nivel de educación del padre, 7 a 9 años, por ejemplo, indicará la contribución de ese grado de educación (puede ser secundaria o comercial) a la calificación del estudiante, sin tomar en cuenta el número de individuos (padres de estudiantes) que hayan alcanzado o no ese nivel educativo.

CUADRO 4  
VARIABLES SELECCIONADAS PARA EL ANALISIS DE REGRESION

Nombre de la Variable	Categoría en la Variable	Identificación de la Variable
<u>Aprovechamiento Escolar</u>	Dato continuo	
FACPYA		$Y_a$ a = 1 a 100
FIME		$Y_b$ b = 1 a 100
MEDICINA		$Y_c$ c = 1 a 100
<u>Factores del Medio Ambiente Familiar</u>		
Status ocupacional	1. Profesional y técnico afín 2. Funcionario público y otros 3. Personal administrativo 4. Comerciante 5. Servicios personales 6. Agricultores y ganaderos *7. Obreros 8. Insuficientemente esp.	$Z_d$ d = 1 a 8
Educación del padre	*1. 1 a 6 años 2. 7 a 9 años 3. 10 y 11 años 4. 12 y más	$Z_e$ e = 1 a 4
Educación de la madre	*1. 1 a 6 años 2. 7 a 9 años 3. 10 y 11 años 4. 12 y más	$Z_f$ f = 1 a 4
<u>Factores del Medio Ambiente Escolar</u>		
Utilidad del material de apoyo	*1. Nada (1 a 5) 2. Poca (6 a 10) 3. Mucha (11 a 25)	$Z_g$ g = 1 a 3
Opinión de la Biblioteca	*1. Mala (1 a 27) 2. Regular (28 a 36) 3. Buena (37 a 45)	$Z_h$ h = 1 a 3
Rigor Académico	*1. Quede igual (10 a 18) 2. Aumente (19 a 27)	$Z_i$ i = 1 y 2
<u>Factores Personales del Alumno</u>		
Sexo	1. Hombre *2. Mujer	$Z_j$ j = 1 y 2
Trabaja	*1. Sí 2. No	$Z_k$ k = 1 y 2

CONTINUACION.....

Nombre de la Variable	Categoría en la Variable	Identificación de la Variable
Satisfacción con la calificación	1. Sí *2. No	$Z_1$ 1 = 1 y 2
Horas semanales dedicadas a actividades escolares	*1. 1 a 40 2. 41 y más	$Z_m$ m = 1 y 2
Horas semanales dedicadas a actividades extraescolares	*1. 1 a 10 2. 11 a 20 3. 21 a 30 4. 31 y más	$Z_n$ n = 1 a 4

\* Variables omitidas en el análisis de regresión

## VI.- ANALISIS DE RESULTADOS

La finalidad de esta sección es verificar las hipótesis de trabajo que se han expuesto con anterioridad (Ver sección IV). Los procedimientos estadísticos empleados, análisis multivariable y pruebas de diferencias de medias se presentan en forma separada.

## A.- MODELO DE REGRESION MULTIPLE.

Los resultados derivados del modelo de regresión especificado, concuerdan en forma general con lo esperado, salvo algunas excepciones que resultan interesantes. El Cuadro 5 presenta el coeficiente de regresión ( $\beta$ ) asociado a las categorías de las variables integradas en: factores del medio ambiente familiar, escolar y características personales del alumno, para FACPYA, FIME y Medicina. También se incluye la desviación estándar ( $S_{bx}$ ) para cada conjunto de betas ( $\beta$ 's); el estadístico  $R^2$ , medida de la bondad del ajuste y el tamaño (n) de la muestra.

Se recuerda que algunos coeficientes son iguales a 0.00, estos datos son incorporados en el término constante  $\alpha$ . Los coeficientes de regresión de las restantes variables independientes deben ser leídos en términos de puntos de calificación obtenidos, y son aditivos. Un ejemplo de este procedimiento se elabora en el Cuadro 6, el objetivo como puede observarse es tener un modelo predictivo del rendimiento académico.