Estimación de calificación del examen de admisión usando el modelo en dos etapas regular: Caso U.N.E.T.

Valera, Jorge; Sinha, Surendra; Goitía, Arnaldo

Recibido: 17-10-2011 - Revisado: 10-11-2011 - Aceptado: 26-11-2011

Valera, Jorge
Ingeniero de Sistemas.
Universidad Nacional Experimental del Táchira - Venezuela
jorgevalera39@hotmail.com

Sinha, Surendra
M.S. y Ph.D. (Major: Estadística Genética, Minor: Statistics)
Universidad de Los Andes – Venezuela
sinha32@yahoo.com

Goitía, Arnaldo
Licenciado en Matemáticas. Maestría en Estadística. Doctor en Matemáticas.
Jefe de la Sección de Docencia del Instituto de Estadística Aplicada y Computación (IEAC), Universidad de Los Andes – Venezuela
goitia@ula.ve

Se presenta un modelo en dos etapas regular como método para la estimación de la calificación obtenida en el examen de admisión de la Universidad Nacional Experimental de Táchira (U.N.E.T.) por un aspirante a ingresar a esta casa de estudios que haya cursado previamente el curso propedéutico en esta misma institución. La novedad de este tipo de modelado es que permite incorporar datos históricos provenientes de estudios previos en los cuales el común denominador es que en cada uno de ellos el interés se centra en la misma variable respuesta. Se obtienen muestras de las calificaciones de los exámenes de admisión para los períodos 2009-1 y 2010-1; así como de los cursos propedéuticos para los mismos períodos y sobre la base de ellas se realiza el ajuste del modelo. Finalmente, se obtienen intervalos de confianza para los parámetros del modelo ajustado.
Palabras clave: Modelo en dos etapas regular, estimación de parámetros, estimadores UBLUE, intervalos de confianza.

RESUMEN

We present a regular two-stage model as a method for estimating the grade obtained in the entrance examination of the National Experimental University of Tachira (UNET) by an applicant to enter this university who has completed the preparatory course prior to the same institution. The novelty of this type of modeling is that it allows incorporating historical data from previous studies in which the common denominator is that each one of them the focus is on the same response variable. Samples are obtained test scores for admission to 2009-1 and 2010-1 periods, as well as the preparatory courses for the same periods and based on them making the adjustment of the model. Finally, we obtain confidence intervals for the fitted model parameters.
Keywords: two-stage model regular, parameter estimation, UBLUE estimators, confidence intervals.

ABSTRACT

1. Introducción

Muchos trabajos han sido escritos a la fecha sobre el rendimiento estudiantil, y el motivo de ello se debe a los continuos cambios que ocurren en el proceso de enseñanza-aprendizaje como consecuencia de las variaciones que acontecen en los sistemas educativos, las cuales influyen de manera directa en el rendimiento académico a lo largo del tiempo, de allí el interés del tema para los investigadores. Desde hace tiempo en nuestras universidades se ha venido presentado el problema del bajo rendimiento estudiantil.

En su aspiración por explicar el rendimiento estudiantil, los investigadores han utilizado métodos y técnicas estadísticas diversas. González (1988), utilizó el análisis de componentes principales y el análisis de correspondencias múltiples en su estudio sobre el rendimiento estudiantil, y entre sus resultados hallaron que el bajo rendimiento estudiantil era uno de los grandes problemas de nuestras universidades, y que las notas obtenidas por los bachilleres influyen de manera positiva sobre el rendimiento universitario, y siguieren la implantación de un sistema que fortalezca los conocimientos adquiridos en el bachillerato con el fin de mejorar el rendimiento estudiantil en la universidad. Lamentablemente, después de más de 20 años el bajo rendimiento sigue siendo uno de los problemas que afectan la vida de nuestras universidades, lo cual explica aún más el porqué siempre habrá investigadores interesados en estudiar el rendimiento estudiantil y los

factores que sobre él influyen.

González (1989) y Garnica *et al*. (1991), utilizaron modelos LISREL, y Análisis Discriminante respectivamente, para estudiar el rendimiento estudiantil. González destaca que el rendimiento estudiantil es influenciado fundamentalmente por un factor aptitudinal, el cual a su vez es influenciado por la preparación del estudiante al momento de ingresar a la universidad y por su situación socioeconómica. Por su parte, Garnica *et al.* en su estudio encontraron que el rendimiento en la universidad está ligado a las destrezas numéricas del estudiante, y al promedio de bachillerato, y recomiendan la implantación de cursos de nivelación universitaria. Posteriormente Garnica (1997), en su estudio del rendimiento estudiantil, haciendo uso del Análisis de Componentes Principales en conjunto con el Análisis de Varianza (ANOVA) encontró que en los planteles tanto públicos como privados el deterioro en la calidad de la educación era evidente, siendo aun más notable en los institutos de educación públicos; y que apenas una minoría de los alumnos que ingresan a las universidades poseían una buena preparación integral.

Más recientemente, Valera *et al*. (2009), Ponsot et al. (2009), y Varela *et al*. (2009), utilizan modelos de regresión logística en sus estudios relacionados sobre el rendimiento estudiantil universitario. Valera *et al.*, al estudiar las calificaciones de los primeros dos semestres, de los estudiantes de la Facultad de de Ingeniería, de la Universidad de Los Andes, determinaron que el promedio de bachillerato tiene un efecto significativo sobre el rendimiento académico en el primer semestre. Paralelamente Ponsot *et al*. y Varela *et al.*, estudiaron las calificaciones de los estudiantes de la Facultad de Ciencias Económicas y Sociales, de la Universidad de Los Andes. Ponsot *et al.*, encontraron que aquellos estudiantes que fueron buenos alumnos en secundaria tenían casi el doble de posibilidades de obtener una eficiencia elevada en sus estudios universitarios, en relación con los alumnos con bajas calificaciones en secundaria, y Varela *et al.*, hallaron que el período académico que mejor pronostica el rendimiento estudiantil dentro de cada una de las carreras era primero y que aquellos estudiantes con rendimientos mayores a 15 puntos en el primer período académico poseen mayores posibilidades de obtener un rendimiento bueno en la carrera.

En este orden de ideas, este trabajo pretende mostrar cómo un modelo en dos etapas regular se puede emplear para predecir la calificación obtenida en el Examen de Admisión de la Universidad Nacional

Experimental del Táchira (U.N.E.T.) por un aspirante a ingresar a esta casa de estudios, a partir de su calificación en el curso propedéutico dictado por la misma universidad, y cuyo propósito es fortalecer los conocimientos obtenidos en bachillerato.

2. Justificación

El rendimiento estudiantil siempre ha sido un tema de interés para muchos investigadores y para su estudio se han utilizado diferentes metodologías y diversas técnicas estadísticas. Los conocimientos adquiridos en el bachillerato por los estudiantes que aspiran ingresar al subsistema de educación universitaria son muy deficientes, y ello ha sido de gran preocupación para las universidades venezolanas en general, ya que constituye uno de los factores principales del bajo rendimiento académico. En el caso específico de la Universidad Nacional Experimental del Táchira (UNET, 2011) para buscar solventar esta situación se fundó la Unidad de Admisión, la cual "tiene como objetivo primordial, garantizar el ingreso de aspirantes a cursar estudios en las diferentes carreras que ofrece la universidad, de acuerdo a lo estipulado en las normas"[1], y es el ente de la universidad encargado de todo lo relacionado con el curso propedéutico, el cual se dicta todos los semestres; y según la UNET, dicho curso "cubre un conjunto de objetivos agrupados por áreas y tiene la finalidad de nivelar los conocimientos básicos del bachiller, preparándolo para que enfrente con menor dificultad las exigencias académicas del nivel básico de la universidad o del examen de admisión"[1].

En este sentido, en este trabajo nos hemos planteado dos objetivos fundamentales: fomentar el uso de los modelos lineales en dos etapas en las diferentes áreas del conocimiento, ya que hasta la fecha su uso por parte de los investigadores ha sido restringido a pesar del gran potencial que este tipo de modelos puede tener, y resolver un problema práctico en el campo de educación al explicar la calificación obtenida por un bachiller en el examen de admisión de la U.N.E.T. en función de su calificación obtenida en el curso propedéutico de la U.N.E.T a través de la aplicación de un modelo de dos etapas regular.

Para alcanzar estos objetivos, se utilizaran las expresiones de los estimadores del vector de parámetros y del vector de medias del Modelo en dos etapas regular desarrolladas por Sinha y Valera (2011);

[1] Tomado de la página web http://www.unet.edu.ve/la-docencia/admision.html

expresiones en las cuales estos estimadores se presentan como funciones de operadores de proyección.

3. El Modelo en dos etapas regular

Definición. Sea el modelo

$$\begin{aligned} \boldsymbol{y}_1 &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\ \boldsymbol{y}_2 &= \boldsymbol{D}\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \end{aligned} \tag{1}$$

tal que $\boldsymbol{y}_1$ y $\boldsymbol{y}_2$ son vectores aleatorios de dimensión $n_1 \times 1$, $n_2 \times 1$ respectivamente; las matrices $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, y $\boldsymbol{D}$ conocidas de números reales; $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ vectores de parámetros desconocidos de dimensión $p_1 \times 1$, $p_2 \times 1$; y $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$ vectores aleatorios de error, con $E(\boldsymbol{\varepsilon}_1) = \boldsymbol{0}$; $Cov(\boldsymbol{\varepsilon}_1) = \sigma_1^2\boldsymbol{V}_1$, $E(\boldsymbol{\varepsilon}_2) = \boldsymbol{0}$; $Cov(\boldsymbol{\varepsilon}_2) = \sigma_2^2\boldsymbol{V}_2$, $\boldsymbol{\varepsilon}_1$ y $\boldsymbol{\varepsilon}_2$ no correlacionados, $\sigma_1^2 > 0$, $\sigma_2^2 > 0$ parámetros desconocidos y $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ matrices conocidas de dimensiones apropiadas. En este modelo, el término $\boldsymbol{D}\boldsymbol{\beta}_1$ es el enlace entre las etapas del modelo. A este modelo se le denomina modelo lineal en dos etapas y según Kubáčk (1988) si los rangos de las matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{V}_1, \boldsymbol{V}_2$ son $r(\boldsymbol{X}_1) = p_1 < n_1, r(\boldsymbol{X}_2) = p_2 < n_2, r(\boldsymbol{V}_1) = n_1, r(\boldsymbol{V}_2) = n_2$, al modelo se le denomina modelo lineal en dos etapas regular, y en este caso es claro que $\boldsymbol{V}_1, \boldsymbol{V}_2$ son matrices definidas positivas.

En este caso, la primera etapa está relacionada al modelo $\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$, y la segunda etapa a $\boldsymbol{y}_2 = \boldsymbol{D}\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$. Obsérvese que lo que hace interesante a este tipo de modelos es que permite incorporar datos disponibles de algún estudio previo, siempre y cuando la variable respuesta en cada etapa sea la misma. Los Modelos en Dos Etapas y en general en p etapas fueron definidos por Kubáčk en (1988) y (1986) respectivamente. Tales modelos han sido estudiados por Volaufova en varios artículos, particularmente en Volaufova (1987)], (1988) y (2004), y más recientemente por Sinha y Valera (por publicar). Matricialmente al modelo lineal regular en dos etapas puede escribirse como:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{D} & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} \tag{2}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

con

$$Cov(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_1^2 \boldsymbol{V}_1 & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \boldsymbol{V}_2 \end{bmatrix} = \boldsymbol{V}_{\boldsymbol{\sigma}}$$

$\boldsymbol{X}_1$, $\boldsymbol{X}_2$ de rango completo por columnas, $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, definidas positivas (d.p.), y $\sigma_1^2 > 0$, $\sigma_2^2 > 0$ parámetros de covarianza desconocidos.

Al pre multiplicar $\boldsymbol{y}$ en el modelo (2) por la matriz $\boldsymbol{F}$ dada por

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{I}_1 & \mathbf{0} \\ -\boldsymbol{DQ} & \boldsymbol{I}_2 \end{bmatrix} \tag{3}$$

se obtiene el modelo transformado $\boldsymbol{Fy} = \boldsymbol{FX\beta} + \boldsymbol{F\varepsilon}$, el cual matricialmente se expresa como

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_1 & \mathbf{0} \\ -\boldsymbol{DQ} & \boldsymbol{I}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} \tag{4}$$

$$\boldsymbol{y}^* = \boldsymbol{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$$

con $\boldsymbol{y}_2^* = -\boldsymbol{DQy}_1 + \boldsymbol{y}_2$, $\boldsymbol{Q}$, una matriz tal que $\boldsymbol{QX}_1 = \boldsymbol{I}$, $\boldsymbol{C} = \boldsymbol{DQV}_1$, y

$$Cov(\boldsymbol{\varepsilon}^*) = \boldsymbol{V}_{\boldsymbol{\varepsilon}^*} = \begin{bmatrix} \sigma_1^2 \boldsymbol{V}_1 & -\sigma_1^2 \boldsymbol{C}' \\ -\sigma_1^2 \boldsymbol{C} & \sigma_1^2 \boldsymbol{C} \boldsymbol{V}_1^{-1} \boldsymbol{C}' + \sigma_2^2 \boldsymbol{V}_2 \end{bmatrix}$$

Volaufova (1987), estableció que el Mejor Estimador Lineal Insesgado Uniformemente (UBLUE por sus siglas en inglés) de $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)$ obtenidos a partir del modelo transformado existen si y sólo si $\mathcal{M}(\boldsymbol{D}) \subset \mathcal{M}(\boldsymbol{X}_2)$; esto es, si y sólo si el espacio vectorial de generado por las columnas de la matriz $\boldsymbol{D}$ es un subespacio del espacio vectorial generado por las columnas de la matriz $\boldsymbol{X}_2$. Sinha y Valera (por publicar) encontraron que si la condición anterior se satisface, los estimadores de $\boldsymbol{\beta}$ y $\boldsymbol{X\beta}$ están dados por

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^{*-} \boldsymbol{P}_{\boldsymbol{X}^* | \boldsymbol{V}_{\boldsymbol{\varepsilon}^*} \boldsymbol{Z}} \boldsymbol{Fy} \tag{5}$$

$$\widehat{\boldsymbol{X\beta}} = \boldsymbol{X} \boldsymbol{X}^{*-} \boldsymbol{P}_{\boldsymbol{X}^* | \boldsymbol{V}_{\boldsymbol{\varepsilon}^*} \boldsymbol{Z}} \boldsymbol{Fy} \tag{6}$$

donde

$$\boldsymbol{P}_{\boldsymbol{X}^* | \boldsymbol{V}_{\boldsymbol{\varepsilon}^*} \boldsymbol{Z}} = \begin{bmatrix} \boldsymbol{I}_1 - \boldsymbol{P}_{\boldsymbol{Z}_1 \boldsymbol{V}_1}' & \mathbf{0} \\ \boldsymbol{DQP}_{\boldsymbol{Z}_1 \boldsymbol{V}_1}' & \boldsymbol{I}_2 - \boldsymbol{P}_{\boldsymbol{Z}_2 \boldsymbol{V}_2}' \end{bmatrix}$$

$$\boldsymbol{X}^{*-} = \begin{bmatrix} \boldsymbol{X}_1^- & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2^- \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' & \mathbf{0} \\ \mathbf{0} & \left(\boldsymbol{X}_2' \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2' \end{bmatrix}$$

y las matrices $\boldsymbol{P}_{\boldsymbol{Z}_1\boldsymbol{V}_1}, \boldsymbol{P}_{\boldsymbol{Z}_2\boldsymbol{V}_2}$, son los operadores de proyección definidos por Rao (1974). Es decir, $\boldsymbol{P}_{\boldsymbol{Z}_1\boldsymbol{V}_1} = \boldsymbol{Z}_1\{\boldsymbol{Z}_1\boldsymbol{V}_1\boldsymbol{Z}_1\}^{-}\boldsymbol{Z}_1\boldsymbol{V}_1, \boldsymbol{P}_{\boldsymbol{Z}_2\boldsymbol{V}_2} = \boldsymbol{Z}_2\{\boldsymbol{Z}_2\boldsymbol{V}_2\boldsymbol{Z}_2\}^{-}\boldsymbol{Z}_2\boldsymbol{V}_2$, en las que $\boldsymbol{Z}_1 = \boldsymbol{I}_1 - \boldsymbol{P}_{\boldsymbol{X}_1}, \boldsymbol{Z}_2 = \boldsymbol{I}_2 - \boldsymbol{P}_{\boldsymbol{X}_2}$, , y donde $\boldsymbol{P}_{\boldsymbol{X}_i}$ es la matriz operador proyección perpendicular sobre $\mathcal{M}(\boldsymbol{X}_i)$ a lo largo de $\mathcal{M}(\boldsymbol{V}_i\boldsymbol{Z}_i)$ con $i = 1, 2$, y las matrices $\{\boldsymbol{Z}_1\boldsymbol{V}_1\boldsymbol{Z}_1\}^{-}$, $\{\boldsymbol{Z}_2\boldsymbol{V}_2\boldsymbol{Z}_2\}^{-}$ son inversas generalizadas en el sentido de Rao (1972).

4. Población y muestra.

Los datos utilizados en este artículo fueron obtenidos de la página web de la Unidad de Admisión de la U.N.E.T. en la sección resultados. Se obtuvieron dos conjuntos de datos. El primero correspondiente al período 2009-1 con 661 registros y el segundo al período 2010-1 con 603 registros. Cada conjunto de datos contiene solamente los resultados del examen de admisión y las notas del curso propedéutico para cada uno de los aspirantes a cursar estudios en las carreras de Ingeniería Industrial, Mecánica, Electrónica, Informática y Civil. Se excluyeron los datos de aquellos bachilleres que no tienen nota en el Examen de Admisión o en el Curso Propedéutico y cuyo ingreso fue por admisión directa o a través de la OPSU[2].

En este estudio se asume que existe independencia entre los conjuntos de datos de los períodos 2009-1, 2010-1 debido a que pertenecen a grupos diferentes de estudiantes, quienes realizaron el curso propedéutico y presentaron el examen de admisión en períodos de tiempo diferentes no consecutivos.

Para el ajuste del modelo se obtuvo de cada período muestras aleatorias en lugar de utilizar la totalidad de los datos disponibles. Dos razones motivan esta decisión, la primera está relacionada con el tipo de datos; cuando las variables son del tipo socio-económico, los datos en general son bastante heterogéneos y los modelos ajustados a partir de la totalidad de los datos resultan usualmente poco precisos. En estos casos, los resultados basados en muestras más homogéneas que el conjunto de datos en su totalidad son mucho mejores. La segunda razón es que en muchas de las investigaciones que se realizan es difícil disponer de gran cantidad de datos por diversas razones, entre ellas económicas, o

[2] Estos datos están disponibles la página web http://admision.unet.edu.ve/, de la Unidad de Admisión de la U.N.E.T.

relacionadas con la dificultad de obtener las muestras. En este sentido, se desea mostrar que no necesariamente se requiere disponer de gran cantidad de datos para realizar el ajuste de un modelo en dos etapas regular.

El modelo en dos etapas regular requiere independencia de los datos de una etapa, con respecto a los datos de otra etapa y la variable respuesta a estudiar debe ser la misma en cada una de las etapas. En este caso, de cada conjunto de datos extraeremos muestras aleatorias independientes. En este sentido, del período 2009-1 se obtendrán muestras aleatorias de las variables Calificación en el Examen de Admisión Período 2009-1 (CEA2009_1) y Calificación en el Curso Propedéutico Período 2009-1 (CP2009_1), y del período 2010-1 se obtendrán las variables Calificación en el Examen de Admisión Período 2010-1 (CEA2010_1) y Calificación en el Curso Propedéutico Período 2010-1 (CP2010_1).

En relación con el modelo en dos etapas regular, las variables dependientes son $\boldsymbol{y}_1 = \mathrm{CEA2009\_1}$ y $\boldsymbol{y}_2 = \mathrm{CEA2010\_1}$. Las matrices $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$ están formadas por dos columnas de datos. La primera es una columna de unos correspondiente al término de la pendiente de la etapa y la segunda es la variable explicativa. En este caso CP2009_1, CP2010_1 son las variables explicativas de la etapa 1 y la etapa 2 respectivamente. Claramente $\boldsymbol{y}_1$, $\boldsymbol{X}_1$ se corresponden con el período 2009-1 y $\boldsymbol{y}_2$, $\boldsymbol{X}_2$ con el período 2010-1. En este estudio particular se decidió que el tamaño de cada una de las muestras utilizadas fuese mayor a 30 con el fin de garantizar la convergencia de las estimaciones hacia el verdadero valor de los parámetros y obtener buenos resultados en el caso de que el comportamiento de los datos se aleje del comportamiento normal, en particular se tomó $n_1 = n_2 = n = 40$. Es importante destacar que para la obtención de las muestras aleatorias, el ajuste del modelo y los resultados de este articulo, se utilizó un lenguaje de cómputos matemáticos y software estadístico. En la actualidad existen en el mercado paquetes especializados como MATLAB, Mathematica, Maple, Statgraphics, SPSS, etc. que se pueden utilizar para obtener estos resultados.

5. Ajuste del modelo en dos etapas

5.1. Análisis descriptivo y ajuste de los datos a una distribución normal.

De cada período se obtuvo una muestra aleatoria de tamaño 40. El Cuadro 1 muestra el resultado de las correlaciones entre CEA2009_1,

CP2009_1, CEA2010_1, y CP2010_1 para la muestra seleccionada. En este se destaca el valor p de significancia.

Cuadro 1
Correlaciones entre las variables bajo estudio

	CEA2009_1	**CP2009_1**	**CEA2010_1**	**CP2010_1**
CEA2009_1	**1**	0,7743 0,0000*	-0,1287 0,4288	-0.1136 0,4853
CP2009_1	0,7743 0,0000*	1	0,1099 0,4995	0,1089 0,5035
CEA2010_1	-0,1287 0,4288	0,1099 0,4995	1	0,7590 0,0000*
CP2010_1	-0.1136 0,4853	0,1089 0,5035	0,7590 0,0000*	1

Fuente: Elaboración propia. Correlación estadísticamente significativa al 1%

En este cuadro, se observa que existen correlaciones estadísticamente significativas entre las variables correspondientes a un mismo período (valores de p = 0,0000) y correlaciones no significativas entre las variables para diferente período (valores de p > 0,01), lo que nos lleva a concluir que las muestras cumplen con el supuesto de independencia de los datos entre una etapa y otra, supuesto que requiere el Modelo en dos etapas regular. El Gráfico 1 permite apreciar la correlación existente entre los datos de las muestras. Obsérvese que sólo entre las variables para un mismo período se aprecia que existe una tendencia lineal entre las variables, en los demás casos se aprecia una nube de puntos sin patrón aparente en su forma.

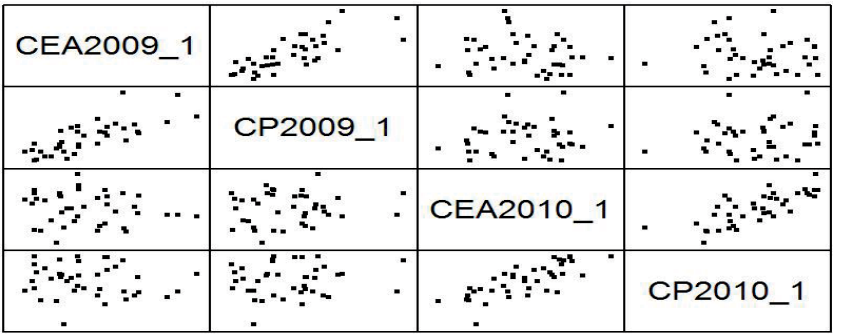

Gráfico 1. **Correlaciones entre las variables de la muestra.** Fuente: Elaboración propia

En el Cuadro 2, se presentan algunas estadísticas descriptivas de interés. En éste se aprecia que las calificaciones obtenidas por los bachilleres en ambos períodos no son altas, en general. En ambos períodos, las calificaciones promedio son inferiores a 50 pts. Menos de la mitad en la escala del 0 a 100 puntos. El segundo cuartil más grande es 53,63 puntos, lo cual indica que 75% de los aspirantes a ingresar a la U.N.E.T. en los períodos 2009-1, 2010-1, para esta muestra obtuvieron calificaciones inferiores a 54 puntos tanto en el curso propedéutico como en el examen de admisión. Esto refuerza la preocupación de docentes e instituciones en relación con la poca calidad de la educación con que llegan los bachilleres a las casas de estudios universitarios. Obsérvese que para las variables CP2009_1 y CP2010_1, que son las calificaciones alcanzadas por los alumnos de la muestra en los cursos propedéuticos de la U.N.E.T., la calificación más alta de la muestra en ambos períodos es 43,00 puntos.

Cuadro 2
Resumen estadístico de las variables

	CEA2009_1	CP2009_1	CEA2010_1	CP2010_1
Frecuencia	40,00	40,00	40,00	40,00
Media	49,45	15,52	41,25	30,08
Mediana	47,99	15,34	41,65	31,30
Desviación típica	7,10	8,96	7,04	8,30
Mínimo	39,00	2,75	25,83	7,00
Máximo	68,03	40,75	56,75	43,00
Rango	29,03	38,00	30,92	36,00
Primer cuartil	44,64	8,17	35,66	24,42
Tercer cuartil	53,63	21,00	46,95	36,09
Coef. de variación	14,36%	57,71%	17,08%	27,59%

Fuente: Elaboración propia.

En lo que respecta a la distribución de las variables calificación en el examen de admisión período 2009-1, y calificación en el examen de admisión período 2010-1, según los resultados obtenidos del ajuste de los datos de la muestra, dados en el Cuadro 3, estas variables presentan un comportamiento normal.

Cuadro 3
Ajuste de las variables respuesta a una distribución normal

Variable: **CEA2009_1**				Variable: **CEA2010_1**			
Estadístico EDF	Valor	Forma Modificada	P-Valor	Estadístico EDF	Valor	Forma Modificada	P-Valor
Kolmogorov-Smirnov D	0,1240	0,7996	>=0.10	Kolmogorov-Smirnov D	0,0907	0,5852	>=0.10
Anderson-Darling A^2	0,5269	0,5375	0,1685	Anderson-Darling A^2	0,4000	0,4081	0,3470

Fuente: Elaboración propia.

5.2. Ajuste del modelo para los datos de la muestra

A partir de las ecuaciones (5) y (6) se procede a realizar el ajuste del Modelo en dos etapas. Para el ajuste del modelo se considerará que las matrices $\boldsymbol{V_1}$, $\boldsymbol{V_2}$ poseen la forma de simetría compuesta. En otras palabras, se asumirá que los elementos seleccionados en cada una de las muestras están igualmente correlacionados. De esta manera, las matrices $\boldsymbol{V_1}$, $\boldsymbol{V_2}$ son de la forma

$$\boldsymbol{V}_i = \begin{bmatrix} 1 & \rho_i & \dots & \rho_i \\ \rho_i & 1 & \dots & \rho_i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \dots & 1 \end{bmatrix}$$

para $i = 1, 2$ y $\rho_1 = 0.3$, $\rho_2 = 0.3$.

En relación con la matriz $\boldsymbol{D}$ se sabe que debe seleccionarse de tal manera que satisfaga la condición $\mathcal{M}(\boldsymbol{D}) \subset \mathcal{M}(\boldsymbol{X_2})$. En el Modelo en dos etapas $\boldsymbol{D\beta_1}$ es el término de enlace entre las etapas, por consiguiente el objetivo es proponer una matriz $\boldsymbol{D}$ en la que la influencia de la etapa 1 se refleje en la etapa 2. De la condición anterior, podemos escribir $\boldsymbol{D = X_2 A}$, donde $\boldsymbol{A}$ es una matriz de dimensión $p_2 \times p_1$. En consecuencia, se sigue que la segunda etapa puede expresarse como

$$\boldsymbol{y}_2 = \boldsymbol{X}_2\boldsymbol{A\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$$

$$= \boldsymbol{X}_2(\boldsymbol{A\beta}_1 + \boldsymbol{\beta}_2) + \boldsymbol{\varepsilon}_2$$

$$= \boldsymbol{X}_2\boldsymbol{\beta}_2^* + \boldsymbol{\varepsilon}_2$$

De esta manera se aprecia que la influencia de la etapa 1 en la etapa 2 está relacionada a la estructura de la matriz $\boldsymbol{A}$. En el caso particular en que $\boldsymbol{A}$ sea

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{7}$$

$$\boldsymbol{\beta}_2^* = \boldsymbol{A}\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 = \begin{bmatrix} \beta_{21} \\ \beta_{12} + \beta_{22} \end{bmatrix}$$

luego, la segunda etapa puede escribirse como

$$\boldsymbol{y}_2 = \beta_{21}\mathbf{J} + (\beta_{12} + \beta_{22})\boldsymbol{X}_{22} + \boldsymbol{\varepsilon}_2$$

$$\boldsymbol{y}_2 = \beta_{21}^*\mathbf{J} + \beta_{22}^*\boldsymbol{X}_{22} + \boldsymbol{\varepsilon}_2 \tag{8}$$

donde $\mathbf{J}$ es un vector de unos de $n_2 \times 1$ y $\boldsymbol{X}_{22}$ la variable independiente de la etapa 2; esto es, $\boldsymbol{X}_{22} = \text{CP2010\_1}$. Otras opciones para la matriz $\boldsymbol{A}$ son

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

En particular, estudiaremos el caso donde la matriz $\boldsymbol{A}$ está dada por (7). Los resultados obtenidos a partir de la muestra seleccionada se señalan en el Cuadro 4.

Cuadro 4
Resultados de ajustar el modelo en dos etapas

	Coeficientes de la Etapa 1		Coeficientes de la Etapa 2			
	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{21}^*$	$\hat{\beta}_{22}^*$
$\boldsymbol{D} = \boldsymbol{X}_2\boldsymbol{A}$	39,929	0,61373	21,876	0,030331	21,876	0,64406

Fuente: Elaboración propia.

Del cuadro anterior, vemos que $\hat{\beta}_{21} = \hat{\beta}_{21}^* = 21{,}88$ y $\hat{\beta}_{22}^* = \hat{\beta}_{12} + \hat{\beta}_{22} = 0.64$. Recuerde que las variables dependientes son las calificaciones en los exámenes de admisión en cada período, y las variables independientes las calificaciones en los cursos propedéuticos.

La calificación obtenida por un estudiante en el curso propedéutico depende de muchos factores, los cuales se han dividido en dos grupos. Los factores inherentes al estudiante como lo son su ambiente familiar,

su condición económica, y sobre todo los conocimientos adquiridos en el bachillerato, entre otros, y los factores relacionados con el curso propedéutico entre los que destacan la orientación profesoral con base en un material educativo seleccionado para reforzar sus conocimientos del bachillerato, la interrelación con otros compañeros en el curso propedéutico que tienen como meta común ingresar al sistema de educación superior, lo cual fomenta en el estudiante una sana competencia por adquirir conocimientos, así como otros factores relacionados.

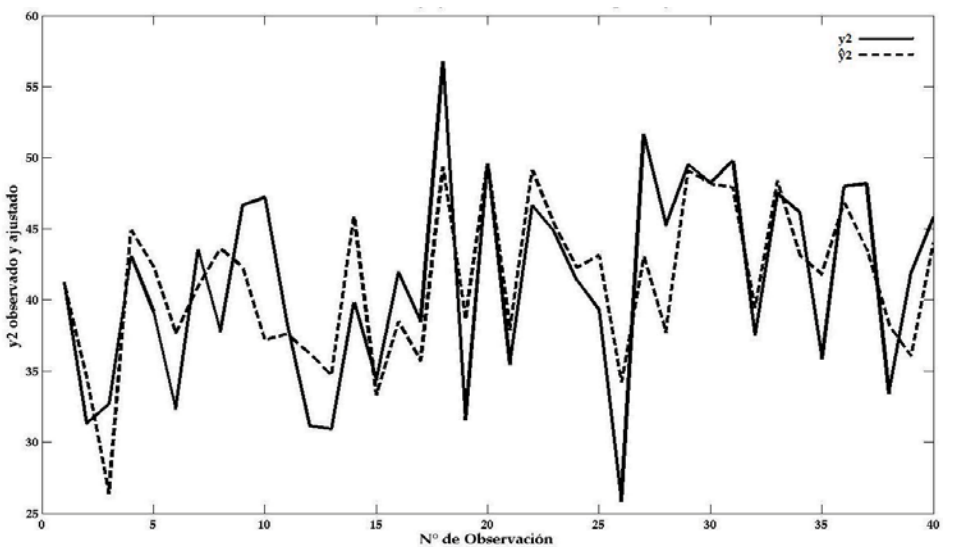

Grafico 2. **Valores observados y predichos a partir del modelo ajustado.** Fuente: Elaboración propia

En este sentido, podemos interpretar la influencia de la etapa 1 sobre la 2, representada por el coeficiente $\hat{\beta}_{12} = 0{,}614$, como la influencia de los factores relacionados con el curso propedéutico, la cual cambia en la medida que se lleven a cabo correctivos que mejoran la eficiencia de dicho curso. De esta manera, la influencia de la calificación del curso propedéutico del período 2010_1 (CP2010_1) sobre la calificación en el examen de admisión para el mismo período (CEA2010_1), puede expresarse como la suma de dos componentes. El componente relacionado con los factores propios del curso propedéutico 0,614*CP2010_1 y el componente correspondiente a los factores inherentes al estudiante 0,03*CP2010_1. La influencia total de la calificación en el curso propedéutico del período

2010_1 sobre su calificación en el examen de admisión está dada por 0.644*CP2010_1. Esta manera de interpretar los coeficientes refleja la realidad en lo que respecta a la mala preparación con la que egresan la mayoría de los estudiantes de hoy día de los liceos del país. De esta manera, el modelo correspondiente a la etapa 2 se expresa como

$$\text{CEA2010\_1}_i = 21{,}88 + 0.644\text{CP2010\_1}_i + \varepsilon_{2i}, i = 1, 2, \ldots, 40 \quad (9)$$

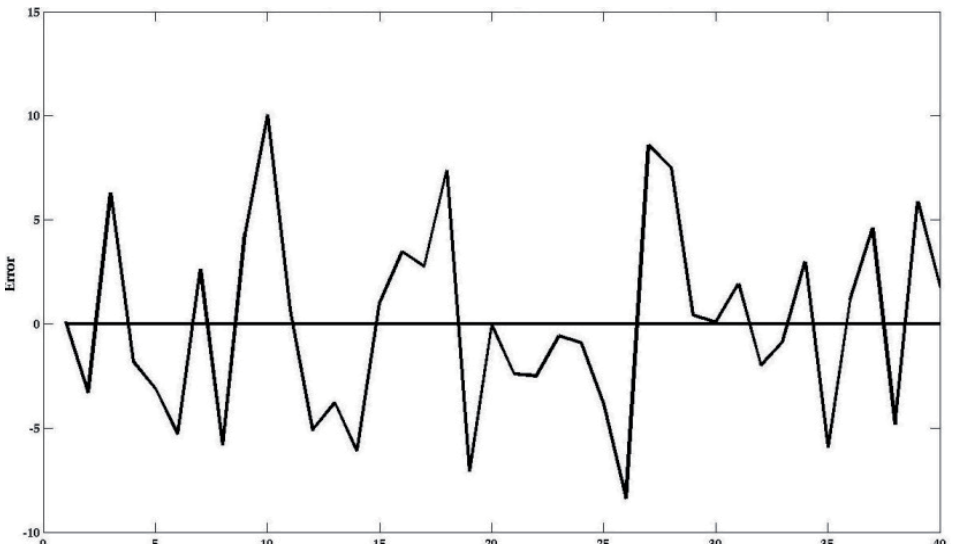

Grafico 3. **Errores entre el valor observado y predicho de la variable CEA2010-1.**
Fuente: Elaboración propia

Como una manera empírica de validar el modelo se obtuvo otra muestra aleatoria de tamaño 40 del período 2010-1, y se ajustó el modelo anterior obteniendo el vector de valores predichos $\widehat{\text{CEA2010}}_1 = \hat{\boldsymbol{y}}_2$ (vector de valores ajustados). Luego se procedió a obtener el vector de $error = \boldsymbol{y}_2 - \hat{\boldsymbol{y}}_2 = \text{CEA2010}_1 - \widehat{\text{CEA2010}}_1$. Los resultados se muestran en los Gráficos 2 y 3. En el grafico 2, se aprecia que los valores predichos por el modelo (línea segmentada) son bastante cercanos a los observados (línea continua), lo cual se aprecia mejor en el grafico 3 donde ninguna diferencia entre los valores observados y los valores predichos de CEA2010_1 obtenidos a partir del modelo es mayor a 10 puntos. Resultados más formales relacionados con el ajuste del modelo no se darán dado que no forman parte de los objetivos de este artículo.

Cuadro 5
Promedios, mínimos y máximos de los parámetros por grupo.

	β_{11}			β_{12}		β_{21}				β_{22}			β_{22}^{*}		
Gr	Prom	Min	Max	Prom	Min	Max	Prom	Min	Max	Prom	Min	Max	Prom	Min	Max
1	38,75	35,00	41,26	0,62	0,52	0,81	22,17	13,09	25,78	0,01	-0,19	0,32	0,64	0,50	0,94
2	39,82	37,17	43,17	0,60	0,44	0,72	24,62	19,78	26,96	-0,03	-0,11	0,15	0,57	0,48	0,73
3	40,57	39,25	42,50	0,55	0,44	0,66	21,23	11,97	26,13	0,11	-0,08	0,35	0,67	0,51	0,94
4	40,76	36,85	42,93	0,56	0,45	0,77	22,46	16,58	25,24	0,06	-0,19	0,24	0,62	0,55	0,81
5	39,24	35,82	42,58	0,60	0,49	0,73	22,23	14,05	26,09	0,03	-0,21	0,29	0,63	0,52	0,89
6	39,95	38,15	41,47	0,57	0,45	0,67	21,41	16,98	24,70	0,10	-0,03	0,26	0,67	0,59	0,79
7	40,28	37,97	42,01	0,58	0,52	0,65	23,40	19,31	26,45	0,01	-0,09	0,12	0,59	0,52	0,71
8	39,48	38,35	41,44	0,58	0,51	0,65	21,71	15,66	25,29	0,06	-0,07	0,28	0,64	0,56	0,82
9	39,49	38,47	40,32	0,60	0,55	0,66	21,37	18,62	24,00	0,05	-0,05	0,19	0,65	0,57	0,74
10	40,13	38,85	41,78	0,57	0,50	0,62	21,88	18,81	25,05	0,08	0,01	0,27	0,65	0,54	0,77
11	39,58	38,10	41,34	0,58	0,51	0,69	23,22	20,47	26,44	0,04	-0,12	0,15	0,62	0,53	0,71
12	39,47	36,79	41,21	0,60	0,52	0,74	21,59	17,11	24,91	0,04	-0,11	0,23	0,64	0,55	0,76
13	39,08	35,97	41,27	0,61	0,50	0,75	23,43	20,10	28,61	0,00	-0,16	0,15	0,61	0,47	0,70
14	38,81	35,08	40,84	0,63	0,55	0,78	21,94	16,37	26,25	0,02	-0,27	0,26	0,64	0,51	0,85
15	39,15	37,73	41,59	0,63	0,53	0,72	20,29	13,84	24,80	0,06	-0,03	0,25	0,69	0,56	0,88
16	39,11	35,99	41,70	0,62	0,52	0,76	22,11	16,55	25,70	0,00	-0,14	0,12	0,62	0,53	0,80
17	39,22	35,54	42,61	0,63	0,48	0,83	21,05	16,45	26,83	0,03	-0,08	0,22	0,66	0,50	0,82
18	39,04	37,14	41,62	0,63	0,52	0,77	21,35	16,42	26,65	0,03	-0,15	0,25	0,66	0,51	0,78
19	39,61	37,76	42,82	0,61	0,49	0,69	21,31	17,45	27,99	0,05	-0,18	0,24	0,66	0,48	0,79
20	38,68	34,65	40,96	0,64	0,52	0,79	21,37	18,97	25,63	0,01	-0,21	0,18	0,65	0,54	0,75

Fuente: Elaboración propia.

5.3. Ajuste del modelo para varias muestras

En la sección 5.1 se mostró que las variables $\mathrm{CEA2009\_1}$ y $\widehat{\mathrm{CEA2010}}_1 = \widehat{\boldsymbol{y}}_2$ poseen un comportamiento normal, y de la ecuación (5) es claro que $\widehat{\boldsymbol{\beta}}$ es una función lineal de vector $\boldsymbol{y} = \left(\boldsymbol{y}_1^{'}, \boldsymbol{y}_2^{'}\right)^{'}$, por lo tanto el vector aleatorio $\widehat{\boldsymbol{\beta}}$ también se distribuye en forma normal. En consecuencia, $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$ y cada elemento $\hat{\beta}_{ij}$ $i = 1, 2$ individualmente se distribuye también en forma normal. Por otra parte, un resultado bastante conocido en estadística dado por Mood et al. (1974, p. 250) en expresa que si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2. La variable aleatoria

$$T = \frac{(\bar{X} - \mu)\left(\sigma/\sqrt{n}\right)}{\sqrt{(1/\sigma^2)\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}}$$

$$= \frac{\sqrt{n(n-1)}(\bar{X} - \mu)}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene distribución t con $n-1$ grados de libertad, donde S^2 es la varianza de la muestra dada por

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Entonces un intervalo de confianza del $(1-\alpha)\times 100\%$ para $E(X) = \mu$ en el caso en que la varianza σ^2 sea desconocida esta dado por

$$\left(\bar{X} - t_{\alpha/2}\, S/\sqrt{n}\,, \bar{X} + t_{\alpha/2}\, S/\sqrt{n}\right) \qquad (10)$$

Cuadro 6
Intervalos de confianza para los parámetros del modelo por grupo de muestras

Parámetro	β_{11}		β_{12}		$\beta_{21} = \beta_{22}^*$		β_{22}		β_{22}^*	
Grupo	Li	Ls	Li	Ls	Li	Ls	Li	Ls	Li	Ls
1	37,293	40,197	0,562	0,679	19,591	24,748	-0,082	0,110	0,552	0,719
2	38,477	41,158	0,545	0,651	23,112	26,132	-0,092	0,028	0,513	0,619
3	39,868	41,267	0,508	0,600	18,329	24,128	0,018	0,208	0,576	0,758
4	39,534	41,987	0,501	0,625	20,679	24,251	-0,032	0,147	0,565	0,676
5	37,775	40,697	0,552	0,652	19,595	24,869	-0,081	0,137	0,542	0,717
6	39,063	40,830	0,518	0,626	19,858	22,967	0,023	0,175	0,630	0,713
7	39,413	41,138	0,540	0,612	21,775	25,026	-0,036	0,064	0,545	0,634
8	38,747	40,207	0,545	0,622	19,694	23,718	-0,013	0,127	0,585	0,696
9	38,983	39,999	0,575	0,621	20,172	22,574	0,010	0,094	0,613	0,686
10	39,345	40,906	0,543	0,600	20,219	23,533	0,018	0,136	0,597	0,700
11	38,816	40,345	0,532	0,632	21,744	24,687	-0,026	0,106	0,582	0,661
12	38,512	40,427	0,538	0,658	19,692	23,488	-0,056	0,132	0,579	0,694
13	37,623	40,531	0,542	0,672	21,437	25,424	-0,071	0,068	0,550	0,662
14	37,363	40,266	0,575	0,677	19,764	24,107	-0,093	0,124	0,568	0,715
15	38,297	40,007	0,595	0,666	17,717	22,855	-0,001	0,127	0,613	0,775
16	37,617	40,595	0,570	0,673	20,181	24,034	-0,063	0,064	0,566	0,679
17	37,236	41,211	0,542	0,710	18,526	23,565	-0,040	0,105	0,589	0,728
18	37,971	40,117	0,573	0,681	19,240	23,455	-0,061	0,117	0,589	0,722
19	38,630	40,588	0,567	0,651	19,040	23,578	-0,042	0,144	0,595	0,725
20	37,120	40,235	0,576	0,707	19,677	23,054	-0,063	0,089	0,601	0,707

Fuente: Elaboración propia.

Tomando en consideración estos resultados se procedió a obtener 20 grupos de muestras, cada uno formado por 10 muestras aleatorias e independientes de tamaño 40. Luego para cada muestra, en cada uno de los grupos se obtuvo el valor de los parámetros $\boldsymbol{\beta}_1 = (\beta_{11},\ \beta_{12})$, $\boldsymbol{\beta}_2 = (\beta_{21},\ \beta_{22})$ y $\boldsymbol{\beta}_2^* = (\beta_{21}^*, \beta_{22}^*)$. Posteriormente, se obtuvieron algunas estadísticas descriptivas e intervalos de confianza para cada uno de los parámetros en cada uno de los grupos. Los promedios y los valores mínimos y máximos se dan en el Cuadro 5. Los intervalos de confianza se dan en el Cuadro 6. Al comparar los valores de $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21} = \hat{\beta}_{21}^*, \hat{\beta}_{22}, \hat{\beta}_{22}^*$ dados en el Cuadro 4 con los valores del Cuadro 5 se observa que en general ellos se ubican entre

los valores mínimo y máximo de cada grupo de muestras. Por ejemplo, $\hat{\beta}_{11} = 39{,}93$ se encuentra entre 35,00 y 41,26 que son los valores mínimo y máximo de los $\hat{\beta}_{11}$ del grupo 1.

Esto muestra de manera empírica que los valores del Cuadro 4 utilizados para ajustar el modelo de la etapa 2 son similares en orden de magnitud a sus pares obtenidos en cada una de las 10 muestras, en cada uno de los 20 grupos. En el Cuadro 6 se muestran los intervalos de confianza para cada uno de los parámetros y para cada uno de los 20 grupos, obtenidos a partir de la expresión (10).

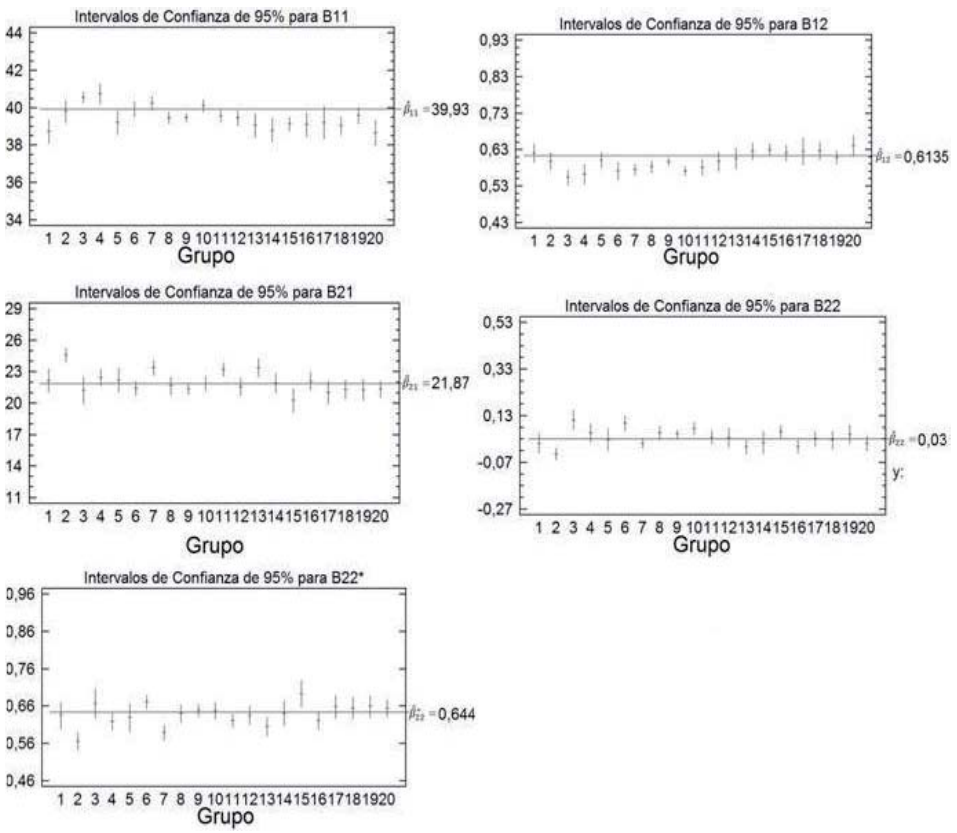

Grafico 4. **Intervalos de confianza para los parámetros del modelo en dos etapas.** Fuente: Elaboración propia

En el Gráfico 4 se visualizan los intervalos de confianza dados en el Cuadro 6 y una línea de referencia que señala el valor del parámetro correspondiente del Cuadro 4. En él se aprecia que los valores del

Cuadro 4 se encuentran en muchos de los intervalos de confianza, lo cual valida estos resultados como representativos de los valores reales de los parámetros que ellos estiman, mostrando que no necesariamente se debe disponer de gran cantidad de datos para ajustar un modelo si la muestra obtenida para ajustar los datos es lo suficiente representativa, de modo que las variables independientes sean buenas variables predictivas de la variable dependiente de interés a estimar.

6. Conclusiones

Al ajustar un modelo en dos etapas regular para predecir las calificaciones obtenidas por los bachilleres en el examen de admisión de la U.N.E.T. a partir de su calificación obtenida en el curso propedéutico previo a dicho examen, hemos mostrado la potencialidad que este tipo de modelado puede tener en otras áreas del saber. Las bondades del modelo de permitir incorporar información correspondiente a los períodos 2009-1 y 2010-1 en un solo modelo se hicieron evidentes al poder separar la influencia de la calificación obtenida en el propedéutico del 2010-1 en dos componentes. Un componente representado por lo que se ha denominado la influencia histórica del curso propedéutico, en donde se refleja la técnica, los materiales, métodos y sobre todo la experiencia adquirida en el curso propedéutico a lo largo de los años, como ente dedicado a la nivelación y al apuntalamiento del conocimiento adquirido por los jóvenes en el bachillerato, representada por el coeficiente β_{12} y un segundo componente relacionado con los factores inherentes al estudiante señalados anteriormente representados en el coeficiente β_{22}.

En este artículo se ha mostrado que la implantación del modelo en dos etapas es un procedimiento relativamente sencillo. Especialmente resulta de utilidad cuando se le emplea para predecir la calificación en el examen de admisión 2010-1 a partir de muestras representativas (de tamaño 40 o superiores), seleccionadas en forma aleatoria e independiente de cada período escolar estudiado. Finalmente, se espera con esta investigación haber contribuido en fomentar el uso de los modelos en etapas, en particular el uso de los modelos en dos etapas regular como herramientas útiles para el análisis de datos.

7. Referencias

Garnica, E.; González, P.; Díaz, A. y Torres, E. (1991). "*Análisis discriminante. Estudio del rendimiento estudiantil.*" Revista Economía Nº 6. Universidad de los Andes. Mérida, Venezuela.

Garnica, E. (1997). "*El rendimiento estudiantil: Una metodología para su medición*". Revista Economía Nº 13. Universidad de los Andes.

Mérida, Venezuela.

González, P. (1988). *Indicadores sintéticos del rendimiento estudiantil.* Revista Economía No 2. Universidad de los Andes. 10(1). Mérida, Venezuela. pp. 69-84.

González, P. (1989). *Aplicación del Lisrel al análisis del rendimiento estudiantil.* Revista Economía No 4.Universidad de los Andes. 10(1). Mérida, Venezuela. pp. 55-73.

Kubáčk, L. (1986). *Multistage regression model*. Applications of Mathematics, Vol. 31, No. 2, pp. 89-96.

Kubáčk, L. (1988). *Two-stage regression model.* Mathematica Slovaca, Vol. 38, No. 4, pp. 383-393

Mood, A., Graybill, F. & Boes D. (1974). *Introduction to the theory of statistics*, 3era edicion. New York: McGraw-Hill.

Ponsot, E.; Sinha, S.; Varela, L. y Valera, J. U*n Modelo de Regresión Logística en los Estudios Universitarios: Caso FACES ULA*. Actualidad Contable FACES Año 12 Nº 18, Enero-Junio 2009. Mérida. Venezuela. pp. 81-102.

Rao, C.R. (1972). *Linear Statistical Inference and its applications.* Jhon Wiley & sons.

Rao, C.R. (1974). *Projectors, Generalized Inverses and the BLUE's.* Journal of the Royal Statistical Society. Series B (Methodological), Vol. 36, No. 3, pp. 442-448.

Sinha, S.; Goitia, A. y Valera, J. (2011). *Ublue for the regular two-stage linear model from the perspective of projection operators.* Electronic Journal of Applied Statistical Analysis. (Entregado para publicación). Universidad Nacional Experimental del Táchira (2011) [Página Web en Línea]. Disponible: http://www.unet.edu.ve/la-docencia/admision.html. [Consulta: 2011, agosto 15]

Valera, J.; Sinha, S.; Varela, L. y Ponsot, E. (2009). *Una explicación del rendimiento estudiantil universitario mediante modelos de regresión logística.* Revista Visión Gerencial FACES. Año 8 N° 2; Julio - Diciembre 2009; Mérida-Venezuela. pp. 415-427.

Varela, J.; Sinha, S.; Ponsot, E.; Valera, J. (2009). *Valor pronóstico del k-ésimo período inicial sobre rendimiento de los estudiantes de la FACES-ULA.* Actualidad Contable FACES Año 12 Nº 19, Julio-Diciembre 2009. Mérida. Venezuela. pp. 133-146

Volaufova, J. (1987). *Estimation of parameters of mean and variance in two-stage linear models.* Applications of Mathematics, Vol. 32, No. 1, pp. 1-8.

Volaufova, J. (1988). *Note on the estimation of parameters of the mean and the variance in n-stage.* Applications of Mathematics, Vol. 33, No. 1, pp.41-48.

Volaufova, J. (2004). *Some estimation problems in multistage.* Linear Algebra and its Applications, 388, pp. 389-397.