PONDERACIÓN ÓPTIMA DE LAS ÁREAS EVALUADAS EN LA PRUEBAS SABER 11, PARA CALCULAR EL PUNTAJE DE INGRESO A LA UNIVERSIDAD

OPTIMAL WEIGHTING ON ASSESSED SUBJECT AREAS OF SABER 11 TEST TO COMPUTE THE ADMISSION SCORE

Luis Hernando Hurtado Tobón[1]; María Dolly García González[2]; Diego López Cárdenas[3]

[1,2] Grupo de Investigación y Asesoría en Estadística, Universidad del Quindío.
[3] Maestría en Biomatematica, Universidad del Quindío.

Recibido: 15 Noviembre de 2015
Aceptado: 15 Diciembre de 2015
*Correspondencia del autor: E-mail: lhhurtado@uniquindio.edu.co

RESUMEN

Las pruebas SABER 11 son utilizadas por muchas universidades colombianas como un mecanismo de selección de los estudiantes que aspiran a los diferentes programas académicos; para calcular el puntaje total de cada aspirante se adopta generalmente un esquema de ponderación donde los ponderadores de las diferentes áreas varían según la carrera. En este trabajo se desarrolla un método para calcular, en forma óptima, los ponderadores de las diferentes áreas contenidas en la prueba SABER 11, según la carrera, bajo la premisa de que el puntaje de ingreso debe estar relacionado con el rendimiento académico posterior del estudiante. Se hace una aplicación con datos reales por vía de ilustración.

Palabras claves: Matriz de correlaciones, Ponderación óptima, Multiplicadores de Lagrange, Pruebas SABER 11.

ABSTRACT

SABER11 are standardized tests used by many Colombian universities as a selection tool to admit first-year students at different academic programs. The total score for each applicant is generally calculated by a weighting scheme where a different weight is applied to each area in standardized test. These weights are chosen according to the academic program that the prospective student is applying to. This paper presents a method to calculate optimally the weights for each area in the SABER11 tests according to the academic program. The proposed method is based on the assumption that the admission score should be related to subsequent students academic performance. A validation with real data is also illustrated.

Keywords: Correlation Matrix, Lagrange Multipliers, optimal weighting, SABER11 standardized test.

1. INTRODUCCIÓN

Las pruebas SABER, en general, son uno de los mecanismos de que dispone el Estado para evaluar el Sistema Educativo en los diferentes ciclos. Las pruebas SABER 11, en particular, se han utilizado además por algunas universidades como mecanismo de selección de estudiantes, función que pasaron a cumplir heredada de las anteriores pruebas ICFES que a su vez reemplazaron las Pruebas de Estado.

Para su aplicacion como mecanismo de selección de estudiantes, las pruebas SABER 11 son analizadas considerando separadamente los resultados en sus distintas áreas y utilizando un sistema de ponderacion que le da un peso diferente a cada área dependiendo de la carrera a la cual el estudiante aspira.

El problema que surge es la forma como se construye el sistema de ponderación: ¿Cuál es su base científica?, ¿Las áreas que tienen mayor ponderación se supone que son garantía de rendimiento del estudiante en esa carrera?, ¿Como asignar un número a cada área que refleje su importancia en una determinada carrera y que tan diferente puede ser ese número del que se asigna a otra área para la misma carrera?. Generalmente estos interrogantes han sido resueltos en forma intuitiva por los administradores de los programas académicos, muy seguramente a través de consensos en los Consejos Curriculares o en los Consejos de Facultad de las universidades.

Lo que se propone entonces es la construcción de un sistema de ponderación, con base científica, el cual garantice que el puntaje global asignado al aspirante al ingresar a una carrera, a través de una combinación lineal de las diferentes áreas de la prueba SABER 11, tenga máxima correlación con el rendimiento académico posterior de ese estudiante. El sistema de ponderación así construido es óptimo en el sentido de maximizar la correlación entre los puntajes de ingreso con los rendimientos académicos de los estudiantes. En términos de Matemáticas el problema se reduce a encontrar un vector que maximice una función de dominio en un espacio vectorial y con imagen en los números reales. Las componentes del vector así encontrado, luego de ser normalizado, son los ponderadores de las diferentes áreas contenidas en la Prueba SABER 11.

La estimación del vector de ponderación, para una determinada carrera, se puede obtener de una muestra de estudiantes a los cuales se les registra informacion sobre los puntajes en la prueba SABER 11, en sus diferentes áreas, en el momento de su ingreso además de su puntaje promedio acumulado hasta el semestre que se hace la aplicación. Por vía de ilustración se incluye la aplicación al calculo del vector de ponderación para el Programa de Medicina en la Universidad del Quindío.

2. CONSTRUCCIÓN DE PONDERADORES

Consideremos un vector $\mathbf{X}_{5\times 1}$ que tiene como componentes los puntajes obtenidos por un estudiante en las cinco (5) diferentes áreas de la prueba SABER 11: Lectura Crítica, Matemáticas, Ciencias Naturales, Ciencias Sociales e Inglés; también un vector $\mathbf{Y}_{1\times 1}$, con una única componente, que es el puntaje promedio acumulado obtenido por el estudiante en un determinado programa de una universidad, es decir

$$\mathbf{X}^T = (X_1, X_2, ..., X_5) \text{ y } \mathbf{Y} = (Y).$$

La matriz de covarianza de X y Y tiene la forma:

$$Cov(X,Y) = \left(\begin{array}{cccc|c} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{15} & \sigma_{1y} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{25} & \sigma_{2y} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{51} & \sigma_{52} & \cdots & \sigma_{55} & \sigma_{5y} \\ \hline\hline \sigma_{y1} & \sigma_{y2} & \cdots & \sigma_{y5} & \sigma_{yy} \end{array}\right)$$

y puede ser particionada en cuatro submatrices, según las líneas dobles, así que puede escribirse como

$$Cov(X,Y) = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix}$$

Donde

$$S_{XX} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{15} \\ v_{21} & v_{22} & \cdots & v_{25} \\ \vdots & \vdots & \ddots & \vdots \\ v_{51} & v_{52} & \cdots & v_{55} \end{pmatrix};$$

$$S_{XY} = S_{YX}^{T} = \begin{pmatrix} v_{1y} \\ v_{2y} \\ g \\ v_{5y} \end{pmatrix};$$

$$S_{YX} = (v_{y1}\ v_{y2}\ g\ v_{y5})\ ;$$

y $\qquad S_{YY} = (v_{yy})$.

Se tiene ademas un vector $a_{5\times 1}$ con $a_i\ !\ R^{+}$ para $i = 1, 2, ..., 5$, donde $\overset{5}{\underset{i=1}{/}} a_i = 1$ y un escalar $b\ !\ R^{+}$. [5]

El producto punto $\mathbf{a}^T\mathbf{X}$ corresponde al puntaje de ingreso a la universidad y el problema de la ponderacion óptima se convierte en encontrar un vector a tal que la correlación entre entre $\mathbf{a}^T\mathbf{X}$ y $b\mathbf{Y}$ sea máxima, esto es, el problema se convierte en resolver para el vector **a**

$$\underset{a}{máx}\ Cqrr(\boldsymbol{a}^T\boldsymbol{X}, b\boldsymbol{Y})$$

$$\underset{a}{\text{máx}} = \frac{\text{Cqv}(a^T\mathbf{X}, b\mathbf{Y})}{[\text{Var}(a^T\mathbf{X})\text{Var}(b\mathbf{Y})]^{\frac{1}{2}}}$$

$$= \underset{a}{\text{máx}} \frac{a^T S_{XY} b}{[(a^T S_{XX} a)(b S_{YY} b)]^{\frac{1}{2}}}$$

Una de las posibles soluciones, por la vía del cálculo, se logra utilizando los multiplicadores de Lagrange, estableciendo las siguientes restricciones [1]

$\mathbf{a}^T S_{XX}\,\mathbf{a} = 1$ y

$bS_{YY}\,b = b^2 S^2_{\ Y} = 1$

lo cual convierte el problema en maximizar para el vector a, la función

$$f(a,b) = a^T S_{XY} b - \frac{\lambda}{2}(a^T S_{XX} a - 1) - \frac{\mu}{2}(b^2 S_Y^2 - 1)$$

Los puntos críticos de esta función se obtienen de resolver el sistema [3]

$$\frac{\delta(a,b)}{\delta b} = a^T S_{XY} - \mu b S_{YY} = 0 \qquad (1)$$

$$\frac{\delta f(a,b)}{\delta a} = S_{XY} b - \lambda S_{XX} a = 0. \qquad (2)$$

Multiplicando ahora la ecuación (1) a la derecha por b y la ecuación (2) a la izquierda por $\mathbf{a}^T$ se tiene el sistema

$$a^T S_{XY} b - \mu b S_{YY} b = 0$$

$$a^T S_{XY} b - \text{m} a^T S_{XX} a = 0$$

a partir del cual, teniendo en cuenta las restricciones y simplificando, se llega a que

$$\lambda = \mu = \mathbf{a}^T{}_{SXY}\, b \qquad (3)$$

y reemplazando λ por μ en la ecuación (2) se tiene que

$$S_{XY}\, b = \mu S_{XX}\, \text{a}.$$

Los términos de la ecuación (1) son escalares así que la ecuación puede escribirse también como [2]

$$S_{YX}\,\mathbf{a} - \mu S_{YY}\, b = 0$$

y multiplicando esta última ecuación a la izquierda por $S_{XY}S_{YY}^{-1}$ se tiene

$$S_{XY}S_{YY}^{-1}S_{XY} a - \text{n} S_{XY}S_{YY}^{-1}S_{YY} b = 0$$

que luego de hacer los reemplazos y las simplificaciones correspondientes se convierte en

$$S_{XY}S_{YY}^{-1}S_{YX} a - \text{n}^2 S_{XX} a = 0$$

lo cual puede escribirse como

$$(S_{XY} v_Y^2 S_{YX} - \mu^2 S_{XX})\, a = 0$$

o también

$$(S_{XX}^{-1} S_{XY} S_{YX} - \text{n}^2 I)\text{a} = 0 -;$$

ya que la variable Y puede estandarizarse de tal manera que $v_Y^2 = 1$.[5]

El problema se reduce entonces a hallar los valores propios μ^2 de la matriz

$$S_{XX}^{-1} S_{XY} S_{YX}$$

y según la ecuacion (3) el valor de μ esta asociado a la covarianza entre el puntaje de ingreso y el rendimiento académico, por lo tanto el vector propio correspondiente al mayor valor propio, una vez normalizado, es el vector que hace maxima la asociación entre el puntaje de ingreso y el rendimiento académico, esto es el vector de ponderación óptima.

3. ESTIMACIÓN DEL VECTOR a

En la actualidad el ICFES, por medio de la Resolución 000503 de julio 22 del 2014, [4] ha establecido cinco (5) áreas de la Prueba SABER 11 a partir de las cuales calcula un puntaje global a cada estudiante que presenta la prueba; en el momento las áreas a considerar para el puntaje global son: Lectura Crítica (LEC), Matematicas (MAT), Ciencias Naturales (NAT), Ciencias Sociales y competencias ciudadanas (SOC) e Inglés (ING). Este puntaje global tiene, como uno de sus propositos establecer una ordenación descendente que permita seleccionar los estudiantes que serían acreedores a algunos de los estímulos que ofrece el gobierno nacional; las mismas áreas escogidas para el cálculo del puntaje global se pueden considerar para calcular el puntaje de ingreso, pero con una ponderación que tenga relación con la carrera a la cual se aspira a ingresar.

La estimación del vector **a**, para un programa académico, se puede entonces conseguir a partir de una muestra de estudiantes del Programa, que tenga registrada información sobre los puntajes obtenidos en cada una de las áreas de la prueba SABER 11 con que hicieron su ingreso a la universidad y su puntaje promedio acumulado (PRO) en el momento en que se hace el estudio.

4. APLICACIÓN

Con el propósito de ilustrar el procedimiento anterior, se presenta una aplicación con la información de los estudiantes que ingresaron al programa de Medicina de la Universidad del Quindío en 2009.

En las primeras cinco columnas de la Tabla 1 aparecen los resultados de las pruebas SABER11 en las diferentes áreas, y en la última columna aparece el puntaje promedio acumulado de las materias cursadas por el estudiante.

Tabla 1. Notas estudiantes de Medicina 2009.

LEC	MAT	NAT	SOC	ING	PRO
56,42	64,43	65,55	53,94	51,16	3,8
66,34	64,43	66,55	32,21	49,43	4,1
59,83	53,45	59,66	48,35	25,75	3,8
55,38	57,98	65,96	38,29	43,4	4
64,56	57,98	63,66	52,27	67,36	3,6
58,68	56,55	65,93	53,16	58,32	3,6
56,21	67,33	59,57	53,18	64,54	4,1
56,21	60,24	66,3	54,12	57,91	3,6
62,32	51,21	73,87	53,96	67,69	4,1
52,07	53,92	63,63	54,02	67,69	4
58,91	56,55	60,26	54,93	64,89	4,2
55,33	61,77	65,7	52,98	67,69	4
58,36	67,33	62,87	50,3	54,19	4,2
55,7	69,93	61,17	53,17	50,67	4,2
61,12	51,11	59,17	45,45	56,01	4,1
60,01	48,67	62,35	48,02	59,92	3,6
66,47	75,91	65,21	40,54	45,31	4,1
61,89	60,24	65,39	51,28	59,92	4,1
61,33	72,76	64,5	53,18	67,36	4
58,93	75,91	61,31	43,39	56,01	4,1
50,52	53,45	69,32	43,66	56,01	4,2
53,28	51,11	67,3	48,45	48,92	3,8
62,3	70,07	64,13	53,94	75,67	4,1
56,55	53,92	65,85	60,85	71,1	3,8

La matriz de varianzas y covarianzas S_{XY} construida con los datos de la Tabla 1 o la matriz de correlaciones, porque las variables han sido estandarizadas, se muestra a continuación.

$$\begin{pmatrix} 1 & 0{,}24 & 0{,}01 & -0{,}2 & 0{,}4 & -0{,}24 \\ 0{,}24 & 1 & -0{,}13 & -0{,}05 & 0{,}06 & 0{,}35 \\ 0{,}01 & -0{,}13 & 1 & -0{,}03 & 0{,}26 & -0{,}08 \\ -0{,}2 & -0{,}05 & -0{,}03 & 1 & 0{,}48 & -0{,}24 \\ 0{,}04 & 0{,}06 & 0{,}26 & 0{,}48 & 1 & 0 \\ -0{,}024 & 0{,}35 & -0{,}08 & -0{,}24 & 0 & 1 \end{pmatrix}$$

A partir de esta matriz se identifica la matriz de correlaciones de las a´reas evaluadas en la prueba SABER 11.

$$S_{XX} = R_{XX} = \begin{pmatrix} 1 & 0,24 & 0,001 & -0,2 & 0,04 \\ 0,24 & 1 & -0,13 & -0,05 & 0,06 \\ 0,01 & -0,13 & 1 & -0,03 & 0,26 \\ -0,2 & -0,05 & -0,03 & 1 & 0,48 \\ 0,04 & 0,06 & 0,26 & 0,48 & 1 \end{pmatrix}$$

cuya matriz inversa es

$$S_{XX}^{-1} = R_{XX}^{-1} = \begin{pmatrix} 1,124 & -0,242 & 0,012 & 0,296 & -0,176 \\ -0,242 & 1,094 & 0,184 & 0,079 & -0,142 \\ 0,012 & 0,184 & 1,143 & 0,252 & -0,429 \\ 0,297 & 0,079 & 0,252 & 1,443 & -0,775 \\ -0,176 & -0,142 & -0,429 & -0,775 & 1,499 \end{pmatrix}$$

también de la Tabla 1 se obtiene la matriz de covarianza

$$S_{XY} = R_{XY} = \begin{pmatrix} -0,24 \\ 0,35 \\ -0.08 \\ -0,24 \\ 0,00005 \end{pmatrix}$$

además

$$S_{YY} = S_Y^2 = 1.$$

Al realizar el producto

$$S_{XX}^{-1} S_{XY} S_{Y\,X}$$

se obtiene la matriz

$$S_{XX}^{-1} S_{XY} S_{YX} = \begin{pmatrix} 0,102 & -0,149 & 0,041 & 0,102 & -0,00002 \\ -0,097 & 0,142 & -0,032 & -0,098 & 0,00002 \\ 0.021 & -0,031 & 0,0072 & 0,021 & -0,000004 \\ 0,098 & -0,143 & 0,033 & 0,098 & -0,00002 \\ -0,05 & 0,074 & -0,017 & -0,051 & 0,00001 \end{pmatrix}$$

cuyo valor propio máximo es 0,56626 y su correspondiente vector propio es

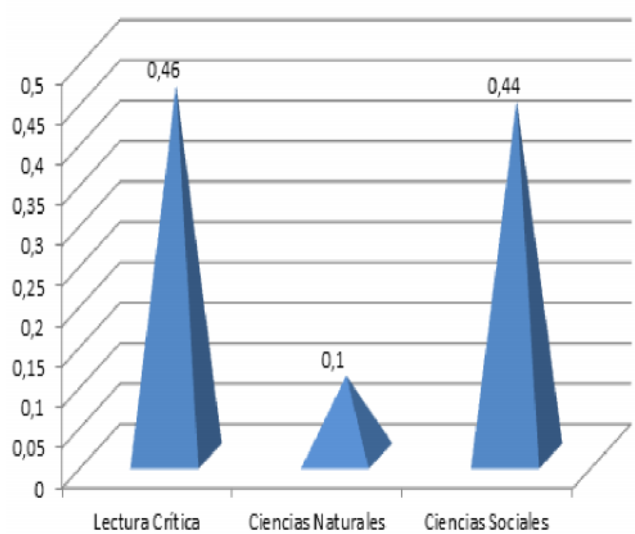
$$\begin{pmatrix} 0,5656 \\ -0,5390 \\ 0,1192 \\ 0,5421 \\ -0,2829 \end{pmatrix}.$$

Como se puede observar, hay dos áreas que no aportan positivamente al puntaje total; haciendo entonces un proceso de depuracion, donde se elimina primero una de las áreas y luego la otra, se llega a que las áreas de la prueba SABER 11 que aportan positivamente al rendimiento académico de los estudiantes de Medicina son: Lectura crítica, Ciencias naturales y Ciencias sociales.

Las componentes del vector propio que se ob- tienen se muestran en la segunda columna de la Tabla 2 y en la tercera columna aparece la pon- deracion normalizada. La ponderación óptima produce una correlacion del 59,24 % con el rendimiento académico (obtenida de la raiz cuadrada del mayor valor propio que resulta cuando el análisis se reduce a las tres a´reas mencionadas) y en este caso es 0,351. Las componentes del vector propio se muestran en la segunda columna de la Tabla 2 y en la tercera columna la ponderación normalizada

Tabla 2. Ponderacion de las áreas.

Area del conocimiento	Vector propio	Ponderadores normalizados
Lectura crítica	0,565	0, 46
Ciencias Naturales	0,11	0, 10
Ciencias Sociales	0,542	0, 44



En la figura 1 se muestra el vector propio nor- malizado, donde aparecen los ponderadores de las áreas que finalmente permiten optimizar la selección de los estudiantes de medicina.

5. CONCLUSIÓN

La utilización de las pruebas SABER 11 como mecanismo para la selección de estudiantes a los programas académicos de Educacion Superior,debe acompañarse de un esquema de ponderación óptimo para las diferentes áreas que contiene la prueba, de tal forma que exista una correlación positiva entre los puntajes de ingreso y el rendimiento académico de los estudiantes; así se garantiza que el mecanismo utilizado sea realmente de selección mas que una herramienta de exclusión.

De todas formas, aún utilizando un esquema de ponderación óptimo, con la selección de los estudiantes a partir de la prueba SABER 11 se pueden presentar situaciones donde los puntajes de selección no muestren correlación con el rendimiento de los estudiantes seleccionados, lo cual es facilmente explicable si se piensa que las pruebas SABER 11 han sido diseñadas mas para evaluar el Sistema Educativo en el ciclo de secundaria que para seleccionar estudiantes en su ingreso a la Educación Superior.

BIBLIOGRAFÍA

1. Apostol, T. Análisis Matemático. Addison Wesley. 1957.
2. Asmar, Abraham. Tópicos en Teoría de Matrices. 1995.
3. Bartle, R. The Elements of Real Analysis. Se- cond Edition. Wiley. 1976.
4. Instituto Colombiano para la evaluación de la Educacion, ICFES, Resolucion N. 000503 de julio 22, 2014.
5. Diaz, Luis. Estadística Multivariada: Inferencia y Métodos. Panamericana Formas e Impresos S.A. 2002.