



ANÁLISIS DE LOS RESULTADOS SABER PRO DEL PROGRAMA DE INGENIERÍA EN PRODUCCIÓN ACUÍCOLA AÑO 2016-2019

ANALYSIS OF THE SABER PRO RESULTS OF THE AQUACULTURE PRODUCTION ENGINEERING PROGRAM YEAR 2016-2019

Sandra M. Cerón-Benavides ^a, Esp. <https://orcid.org/0000-0001-7607-7731>

Manuel F. Romero-Ospina ^b, MSc. <https://orcid.org/0000-0002-6457-8827>

^a Ingeniera en Producción Acuícola, Esp. Estadística Aplicada, Docente, Departamento de Recursos Hidrobiológicos, Universidad de Nariño, Pasto, Colombia. samilbnvds@gmail.com

^b Licenciado en electrónica, Esp. Estadística Aplicada y MSc, Bioestadística y Bioinformática, Docente, Fundación Universitaria Libertadores, Bogotá, Colombia. mfromero@libertadores.edu.co

Recibido: 16-06-2021

Aceptado: 17-07-2021

RESUMEN

La presente investigación tuvo como objetivo definir las principales variables y su incidencia sobre el puntaje de las pruebas SABER PRO dentro del programa de Ingeniería en Producción Acuícola de la Universidad de Nariño en los últimos cuatro años (2016-2019), para ello se utilizaron 384 datos correspondientes a estudiantes de noveno semestre que presentaron las pruebas de estado.

Se determinó mediante diferentes técnicas multivariadas, que todas las variables inciden sobre las dos primeras componentes, con la necesidad de fortalecer Lectura crítica, Comunicación escrita y Competencia ciudadana, las cuales tienen mayor contribución sobre las tres últimas (>50) y por lo tanto impidieron que la variabilidad retenida en las primeras sea mayor al 60%.

Por otro lado, el incumplimiento del supuesto de homocedasticidad previo al Análisis de regresión múltiple no afectó la capacidad predictiva del modelo, la cual fue del 99%, el Análisis de regresión cuantílica y sus cinco modelos no presentaron diferencias significativas entre ellos, por lo que se concluye que todas las variables son indispensables para predecir el puntaje global, cada una con su grado de contribución, calidad de asociación y peso sobre las combinaciones lineales de las dos primeras componentes.

Palabras claves: Técnicas multivariadas, análisis de componentes principales, análisis de la regresión.

ABSTRACT

The objective of this research was to define the main variables and their incidence on the SABER PRO test scores within the Aquaculture Production Engineering program at the University of Nariño in the last four years 2016-2019, 384 data corresponding to ninth semester students who presented the test were used.

It was determined through different multivariate techniques that all the variables affect the first main components with the need to strengthen Critical Reading, Written Communication, and Citizen Competence, which have a greater contribution over the last three (>50), their contribution prevented the variability from being greater than 60% in the first two components.

On the other hand, failure to comply with the assumption of homoscedasticity before the Multiple Regression Analysis did not affect the predictive capacity of the model, which was 99%. Quantile Regression Analysis, and its five models did not present significant differences between them. Therefore, it was concluded that all the variables are essential to predict the global score, each one with its degree of contribution, quality of association, and weight on the linear combinations of the first two components.

Key words: Multivariate techniques, Principal component analysis, Regression analysis.

INTRODUCCIÓN

Los indicadores académicos revelan aquellas problemáticas que enfrenta la Educación Superior en nuestro país, entre ellos podemos hablar del nivel de deserción, el número de estudiantes no matriculados pero inscritos y admitidos, el nivel de desempeño con el que ingresan o egresan de su estudio de pregrado mediante el puntaje SABER 11° y SABER PRO, entre otros, que proporcionan un acercamiento al estado real de un programa, Institución o educación a nivel nacional.

Dependiendo del grado de influencia de estos indicadores se puede establecer la capacidad por parte de la institución, para solventar problemas con el transcurrir del tiempo y mejorar la calidad de su educación, lo anterior según la estructura del programa y/o facultad, medido por asignatura, semestre, periodo o cohorte académico.

La base de datos del Instituto Colombiano para el Fomento de la Educación Superior (ICFES) y el Portal Único del Estado

Colombiano ofrecen información pública y relevante dentro del indicador académico SABER PRO, que facilita el seguimiento y el tratamiento de la misma por Institución y programa académico. Además, en el programa de pregrado existe la necesidad de evaluar el comportamiento de los indicadores académicos a través del tiempo, lo anterior mediante técnicas estadísticas multivariadas con el objetivo de facilitar la toma de decisiones en pro de una mejora continua en la educación superior.

En el presente trabajo se estableció como objetivo principal realizar un análisis de los datos generados en las pruebas de estado SABER PRO el programa de Ingeniería en Producción Acuícola del Departamento de Recursos Hidrobiológicos perteneciente a la Universidad de Nariño sede principal en el municipio de Pasto-Colombia, mediante la aplicación de diferentes técnicas estadísticas multivariadas.

MATERIALES Y MÉTODOS

Este trabajo se desarrolló con los resultados generados por estudiantes de último semestre que cumplieron con el 75% de sus créditos académicos dentro del programa en los últimos cuatro años (2016-2019), los datos se obtuvieron por medio del Instituto Colombiano

para el Fomento de la Educación Superior (ICFES) y el Portal Único del Estado Colombiano mediante su portal web ^[1], en ningún momento se publica información individualizada con el fin de respetar la confidencialidad de cada estudiante.

La base de datos se consolidó con un total de 64 registros y 576 datos distribuidos entre variables cualitativas y cuantitativas las cuales se analizaron por separado. Las variables independientes objeto del presente estudio fueron: Año, Género, Estrato socio-económico, Edad y Puntaje por cada competencia genérica, Comunicación escrita (ce), Razonamiento cuantitativo (rz), Lectura crítica (lc), Competencia ciudadana (cc) e Inglés (i), esta última también denominada predictor dentro del análisis de regresión, mientras que la variable dependiente o de respuesta estuvo dada por el puntaje global total, el cual corresponde al promedio de los puntajes de cada competencia genérica.

El modelo estadístico fue dado por un análisis descriptivo, uno de componentes principales, un análisis de regresión múltiple y otro cuantílica para caracterizar la población estudiantil y encontrar las principales variables con mayor incidencia sobre la dependiente o de respuesta (Puntaje global), la ejecución de todos los análisis estuvo dada por una matriz de datos de Excel y el programa Rstudio.

Análisis descriptivo

El análisis descriptivo se desarrolló mediante medidas de tendencia central como el cálculo de la media, mínimo, máximo, desviación estándar y porcentajes de cada característica poblacional a través del tiempo, los resultados facilitaron la determinación de datos atípicos en las diferentes observaciones y la depuración de la base de datos original.

Análisis Multivariado

Análisis de componentes principales. El análisis de componentes principales se realizó acorde al procedimiento recomendado por Kassambara ^[2]:

En primer lugar, se estandarizaron las variables, de tal forma que tengan media cero y desviación estándar uno, permitiendo

efectuar el análisis en caso de que las varianzas entre las variables no sea la misma.

En segundo lugar, se calcularon los componentes principales y los pesos de cada uno, es decir; la cantidad de información que recogen, con este valor se establecieron las combinaciones lineales de las variables originales. En tercer lugar, se determinó la varianza retenida en cada componente principal y así se apreció sobre cuales recaía el mayor porcentaje (>50%), se determinó la contribución (contrib) para determinar las variables más relevantes, la calidad o grado de asociación (cos2) y la correlación entre las mismas (Coord.). Por último, lugar, se determinó el nivel de significancia entre variables, calculando el p-valor para cada grupo de mayor incidencia dentro de las dos primeras componentes.

Análisis de Regresión y supuestos estadísticos. Previo y posterior al análisis de regresión múltiple se realizaron diferentes test para verificar el cumplimiento de cada supuesto, las hipótesis a tener en cuenta se manejaron con un nivel de significancia de 0,05 y la decisión de aceptación o rechazo recayó sobre la hipótesis nula:

Hipótesis nula (H0): Se acepta si $p\text{-valor} \geq 0,05$.

Hipótesis alterna (H1): Si $p\text{-valor} < 0,05$ se rechaza H0.

Se generaron los coeficientes de asimetría y curtosis estandarizados, donde el rango optimo debe estar entre -2 a 2 ^[3], los test de Mardia de asimetría y Mardia para curtosis, el test de Henze-Zirkler, Royston, Doornik Hansen, E-statistic y el test de Shapiro-Wilk (para residuos), para determinar el supuesto de normalidad.

De acuerdo con Salazar y Ortega et al ^[4], ^[5] una matriz de correlación puede soportar el cumplimiento de multicolinealidad entre el

predictor y la variable respuesta (el coeficiente lo más cercano a uno), y el supuesto de independencia, no autocorrelación o no multicolinealidad entre los predictores (coeficiente lo más cercano a cero).

La Relación lineal ^[4] entre predictores y variable respuesta posterior a generar el modelo por regresión se tuvo en cuenta mediante la representación de la matriz de correlación de cada residuo con su variable o competencia genérica más el intercepto, cumpliendo la condición de que los residuos deben distribuirse lo más cercano a cero.

Posterior a generar el modelo por regresión múltiple se realizó un análisis de inflación de la varianza (VIF) ^[6], y un test de hipótesis de Durbin-Watson, donde si $p > 0,05$ se aceptaba la no autocorrelación entre predictores. También se trabajó con el test de Bartlett, Levene y de Brown-Forsyth para el supuesto de homocedasticidad, este último teniendo en cuenta la mediana como unidad central, además, se manejó el test de Breusch-Pagan para los residuos.

Por último, se determinó mediante distancia de Cook, la influencia de los datos atípicos existentes, donde, según Faraway citado por Amat ^[7], valores menores a uno no influyen.

Regresión múltiple. Se establecieron las variables predictores y aquella dependiente por medio del modelo que a nuestro criterio explica

el desempeño de cada estudiante por medio del puntaje promedio de las pruebas SABER PRO para el programa de pregrado.

En segundo lugar, se seleccionaron los mejores predictores determinando la calidad del modelo mediante el método AIC y se observó si el valor de R^2 se modificó, esto con el fin de determinar si la variable excluida afectaba en gran medida la variabilidad de y ^[8].

Como último paso se determinaron los intervalos de confianza y de nuevo se validaron los supuestos o condiciones estadísticas de los residuos, de no cumplirse algún supuesto se reajustaba el modelo ^[8].

Regresión cuantílica. Este análisis se realizó teniendo en cuenta diferentes estudios ^[9,10]: se establecieron variables que se consideraron predictores, aquella dependiente y los cinco cuantiles sobre los cuales se quiso generar los modelos, 0,1, 0,25, 0,5 (mediana), 0,75 y 0,90, se generaron los coeficientes de cada variable, el p -valor asociado a cada predictor y por medio de las estimaciones del intercepto las representaciones de la pendiente de regresión para cada cuantil, con esto se determinó su influencia y significancia.

Por último, se generó ANOVAS por cada par de modelo para determinar la existencia o ausencia de diferencias significativas entre ellos.

RESULTADOS Y DISCUSIÓN

Análisis descriptivo

Variables cualitativas. En los cuatro años evaluados, el 84% (54 registros) de la población presentó edades entre los 22 a 26 años, predominando el género femenino con un 55% de participación (35 registros) y un rango entre los 21 y 26 años de edad.

Lo anterior caracteriza las edades de los estudiantes que se encontraban en noveno

semestre, pudiendo exceptuar solo seis del género masculino, con edades iguales o superiores a los 27 años, estudiantes que posiblemente reingresaron al programa años más tarde o se encontraban nivelando con el pensum académico. Los resultados en cuanto a género también predominan en los reportes nacionales según el ICFES ^[11] y estudios de otras áreas como el de Orjuela ^[12].

El 100% de los estudiantes pertenecieron a estrato I, II y III, solo en el año 2016 se presentó ausencia de estrato III y del total cuatro estudiantes no presentaron esta variable, por lo que se clasificaron como valores perdidos. El 80% de la población en los cuatro años académicos fue de estrato bajo (I y II), resultados característicos en la mayoría de estudiantes pertenecientes a Universidades públicas, y que también se observan en los documentos de análisis generados por el ICFES cuando caracterizan a los estudiantes de universidades públicas y privadas por competencia genérica ^[13].

La cantidad de estudiantes que presentaron el examen de estado varió cada año dependiendo de aquellos que lograron cumplir con el 75% de los créditos académicos del programa de pregrado, durante el 2016 se presentó la menor cantidad con tan solo ocho registros y en el 2019 la mayor con 21, la baja cantidad de estudiantes no contribuye junto a otros programas a que Nariño sea considerado por el ICFES como uno de los municipios con el mayor número de evaluados a nivel nacional^[12].

VARIABLES CUANTITATIVAS. Se aclara que algunos valores variaron en este estudio según la cantidad de datos existentes por año en cada variable o competencia genérica. Además, sin tener en cuenta datos atípicos podemos

cuestionar los valores de las medias, medianas, desviaciones estándar y rango de puntajes en cada año. Por ejemplo, con el valor mínimo y máximo entre puntajes, se pudo determinar que la variable Razonamiento cuantitativo mantuvo el menor rango, entre 121 a 195 puntos, y junto con Comunicación escrita la mayor media con 148 y 160 puntos, puntaje por encima de la mitad del total, esto se pudo deber a que los estudiantes tienen bases sólidas en cuanto al área ingenieril, presentando entre ellos puntajes similares y acertados dentro de la competencia en los últimos cuatro años (Tabla 1).

La variable Ingles, presentó una media de 144 puntos entre los cuatro años de estudio con la mayor media en el 2019 (148 puntos), la destreza y conocimiento adquirido en esta lengua y en este año académico por 2 estudiantes con 185 y 196 puntos aportaron al resultado en mención, los valores no fueron reportarlos por Rstudio como atípicos influyentes.

Las mayores y menores desviaciones estándar se presentaron en todos los años, Competencia ciudadana y Comunicación escrita las más altas con valores ≥ 28 puntos, mientras que, Razonamiento cuantitativo, Ingles y Lectura crítica respetivamente los menores valores entre los 18 y 23.

Tabla 1. Medidas de tendencia central para variables independientes.

AÑO 2016-2019	Mínimo	Media*	Mediana	Máximo	Desviación estándar
Competencia ciudadana	79	134	133	187	28
Comunicación escrita	97	148	156	202	38
Lectura Critica	80	144	143	191	23
Razonamiento Cuantitativo	121	160	164	195	22
Ingles	97	144	145	185	18

(*) La media hace referencia al puntaje global o variable dependiente generado en cada prueba.

Análisis Multivariado

Previo a realizar el análisis multivariado, Rstudio determinó 13 observaciones atípicas, algunas coincidiendo dentro del mismo individuo o registro. Por lo tanto, la base de datos final conto con 53 registros y cinco variables de tipo cuantitativo (competencias genéricas), trabajando con un total de 265 datos para el análisis de componentes principales y con 318 observaciones para los análisis de regresión, donde se tuvo en cuenta la variable dependiente (puntaje global).

Análisis de Componentes Principales. Como se observa en la Tabla 2 las variables independientes cumplen el supuesto de normalidad al generar un p-valor mayor a 0,05, por lo que se acepta la hipótesis nula y se genera el respectivo análisis.

Tabla 2. P-valor generado entre predictores en los diferentes test de normalidad.

Test	p-valor
Mardia-skewness	0,98
Mardia-kurtosis	0,23
Henze-zirkler	0,28
Royston	0,79
Doornik-Hansen	0,77
E-statistic	0,59

El primer resultado que obtenemos por medio del comando rotation en Rstudio son los diferentes valores de peso/importancia que tiene cada variable sobre cada componente, los pesos más altos asignados en la primera componente corresponden a Razonamiento cuantitativo y lectura crítica con aproximadamente 0,5, Competencia ciudadana con 0,6 y Comunicación escrita e inglés con -0,2 y 0,2 respectivamente. En la segunda componente, la variable Comunicación escrita e inglés tienen la mayor participación con pesos de 0,6 y 0,7, con esto se establecieron

las combinaciones lineales PC1 y PC2 de su variable original.

$$PC1 = -0,20ce + 0,53rc + 0,54lc + 0,60cc + 0,18i$$

$$PC2 = 0,61ce - 0,31rc + 0,09lc + 0,18cc + 0,70i$$

De acuerdo a la Tabla 3, las dos primeras componentes explican el 58,60% de la varianza total retenida y/o observada y las tres primeras el 74,91%, mientras que las dos últimas componentes, PC4 y PC5 no superaron por separado el 14% de la misma con 14,02% y 11,10% respectivamente.

En la Tabla 3 también se puede apreciar que la variable Comunicación escrita e inglés presentaron un alto grado de asociación (Cos2), contribución (Contrib.) y correlación (Coord.) sobre la segunda componente con p-valor de 2,87e-08 y 4,52e-12, mas no son variables predictores relevantes en la primera, mientras que Razonamiento cuantitativo, Lectura crítica y Competencia ciudadana si lo son con p-valor de 1,60e-08, 3,57e-09 y 7,90e-12 respectivamente.

Lo anterior se pudo confirmar con los coeficientes de correlación que resultaron ser superiores a 0,7, con p-valor<0,05 (Tabla 3) y con el trabajo de Martínez & Mendoza ^[14] donde se aprecia la relación entre variables dentro de cada componente, este resultado es considerado normal debido a que existen según el ICFES ^[15] subcompetencias en cada módulo, por ejemplo, Razonamiento cuantitativo implica la interpretación y argumentación de sus resultados y es apenas normal que esté relacionado con lectura crítica y Competencia ciudadana mas no Inglés o Comunicación escrita, los cuales son evaluados por niveles de desempeño.

Tabla 3. Contribución en porcentaje de las variables sobre los componentes principales

		Comunicación escrita	Razonamiento cuantitativo	Lectura crítica	Competencia ciudadana	Inglés
PC1	Contrib	4,16	27,6	29,36	35,58	3,31
	Cos2	0,07	0,47	0,5	0,6	0,06
	Coord	-0,27	0,68	0,71	0,78	0,24
PC2	Contrib	37,01	9,31	0,84	3,19	49,645
	Cos2	0,46	0,11	0,01	0,04	0,61
	Coord	0,68/	-0,34	0,1	0,2	0,78/

Contribución (Contrib), Calidad o grado de asociación (Cos2), Coeficiente de correlación (Coord)

Supuestos estadísticos previos a la regresión. La media entre variables presentó valores entre los 136 a 162 puntos, la mediana de 134 a 165 y varianzas desde los 118 a 625,57 puntos con una desviación estándar mínima de 11 y máxima de 25 puntos.

Todos los supuestos estadísticos se cumplieron a excepción de la homocedasticidad, el cual mediante el test de Bartlett generó un p-valor de $3,25e-09$, el de Levene al igual que el de Brown-Forsyth un p-valor de $4,59e-09$, todos menores a 0,05, confirmando la ausencia de homogeneidad entre varianzas y rechazando la hipótesis nula, más sin embargo la media y mediana mantuvieron baja variación entre competencias, comparable con los resultados presentados por Rodríguez Manrique et al^[16] en su estudio.

Como se observa en la Tabla 4, los coeficientes de asimetría y curtosis y los p-valor obtenidos en los diferentes test fueron superiores a 0,05, por lo que se acepta una distribución normal y por ende H_0 .

Tabla 4. Test multivariados para normalidad.

Test	p-valor
Asimetría	0,16 (1,15)
Curtosis	0,23 (1,70)
Mardia Skewness	0,99
Mardia kurtosis	0,24
Henze-Zirkler	0,45
Royston	0,58
Doornik-Hansen	0,12
E-statistic	0,84

Cada par de variables predictores generaron valores de correlación cercanos a cero (-0,2 a 0,3), pero tendientes a 1 con relación al puntaje global, competencias ciudadanas con 0,7, lectura crítica con 0,6, razonamiento cuantitativo e inglés con 0,5 y comunicación escrita con 0,4, por lo tanto, se acepta la Hipótesis nula para el supuesto de multicolinealidad.

Regresión lineal múltiple. Obviando el incumplimiento del supuesto de homocedasticidad se decidió determinar el modelo por mínimos cuadrados sin previa transformación de los datos, y el modelo cumplió un R^2 de 0,99, lo que implica que más del 90% de la variabilidad del puntaje global está explicada en el modelo, similar a lo ocurrido con Marquín^[17] y con un p-valor significativo de $2,2 e-16$ (Tabla 5), se puede decir que al menos uno de los coeficientes parciales es diferente de cero, que las variables

explicativas tienen asociación con la variable dependiente [18] y que aquella parte del promedio global o variable que dejamos por

explicar corresponde a 0,28 como error residual estándar.

Tabla 5. Coeficientes, error estándar, t value y Pr para cada competencia genérica.

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	-0,09	0,631	-0,16	0,875
Com. Escrita	0,201	0,002	125,43	<2e-16***
Raz. cuantitativo	0,202	0,003	77,22	<2e-16***
Lect. critica	0,200	0,002	95,60	<2e-16***
Comp. ciudadanas	0,197	0,002	111,40	<2e-16***
Ingles	0,200	0,003	74,04	<2e-16***

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1 Residual standard error: 0,2799

En la selección de predictores e intervalo de confianza se estableció que al tener solo cinco predictores todos son esenciales para el modelo lineal, además, el método de *Akaike* (AIC) arrojó una sola opción, igualmente se

pudo determinar que el rango y/o intervalo de valores donde se localiza la media poblacional para cada coeficiente esta entre 0,19(2,5%) a 0,21(97,5%).

Tabla 6. Calidad del modelo lineal con el valor AIC = 129,34.

	DF	Sum of Sq	RSS	AIC
(none)			3,68	-129,34
Ingles	1	429,42	433,11	121,34
Raz. cuantitativo	1	467,20	470,88	125,77
Lect. critica	1	715,93	719,61	148,25
Comp. ciudadanas	1	971,53	975,21	164,35
Com. Escrita	1	1232,40	1236,09	176,92

El mejor modelo resultante con los coeficientes aproximados a dos cifras decimales fue:

$$\text{Puntaje global} = 0,20ce + 0,20rc + 0,20i + 0,20cc$$

Supuestos estadísticos posteriores a la regresión. Los residuos se distribuyeron aleatoriamente en torno a cero con una variabilidad constante a lo largo del eje X y la posible presencia de algunos datos atípicos cumpliendo el supuesto de linealidad entre predictores y variable respuesta.

El test de Shapiro-Wilk arrojó un p-valor de 0,093 < 0,05, por lo que se acepta la hipótesis

nula al establecer una distribución normal de los residuos al encontrarse lo más alineados entorno a la recta. Mediante el test de Breusch-Pagan se pudo establecer un p-valor igual a 0,12 > 0,05, por ende, también se acepta la Ho para el supuesto de homocedasticidad, y los coeficientes de correlación entre cada par de variables fueron menores a uno con valores entre -0,2 a 0,4, por lo que se deduce la no multicolinealidad entre residuos.

Mediante el análisis de inflación de varianza (FIV) no se presentaron predictores que muestren una correlación lineal muy alta, todos los valores de FIV estuvieron entre 1,06 a 1,3 por lo que al no ser mayor a 10 no se evidencia

un grado de colinealidad preocupante. Además, no se evidenció autocorrelación entre los residuos al presentar un p-valor de $0,7 > 0,05$ y un coeficiente Durbin-Watson aproximado a dos (1,89), por lo tanto, se aceptó la Hipótesis nula para este supuesto.

Por último, la distancia de Cook determinó siete observaciones atípicas (4, 16, 17, 22, 27, 32 y 35) con valores entre 0,02 a 0,12, al ser menores a uno se dedujo que ninguno es preocupante o significativamente influyente.

Regresión cuantílica. Los puntajes de las medianas presentaron valores entre 134 a 165 puntos con una variación de ± 31 en la mediana entre predictores y la variable dependiente.

En este análisis se generaron 5 modelos con diferentes cuantiles, entre ellos 0,1, 0,25, 0,5 (mediana), 0,75 y 0,9, todos demostraron coeficientes aproximados a 0,20 sobre cada predictor, por lo que al realizar los ANOVAS no se encontró diferencias significativas entre ellos con p-valores entre 0,22 a 0,69, mayores a 0,05.

Modelos Cuantílicos 0,1/0,25/0,5/0,75/0,9:

$$\text{Puntaje global} = 0,20ce + 0,20rc + 0,20lc + 0,20cc + 0,20i$$

Adicional, podemos decir que las regresiones presentaron un crecimiento desde el cuantil 0,1 al 0,9 para Comunicación escrita, Lectura crítica y Competencia ciudadana. Las variaciones de cada regresión se presentaron dentro del intervalo de confianza entre 0,19 a 0,21 establecido para cada competencia y cuantil, lo anterior a excepción de Comunicación escrita en el cuantil 0,8 y Competencia ciudadana en el cuantil 0,4, donde sus coeficientes mínimamente se establecieron por fuera del intervalo (Figura 1), estos cambios en los intervalos son comparables con los publicado en el estudio de Pérez-Pulido et al ^[19] donde el intercepto si presentó crecimiento a medida que aumento el cuantil en cada regresión.

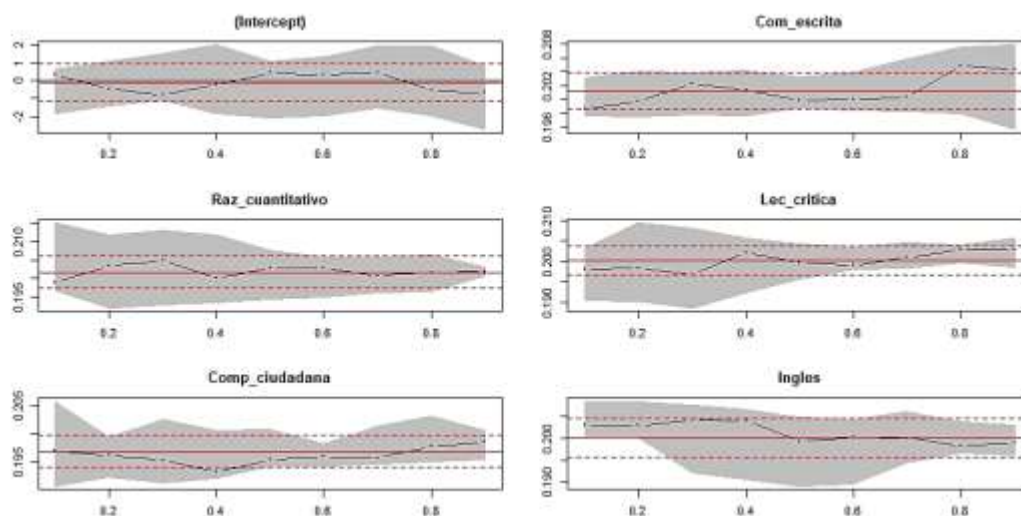


Figura 1. Coeficientes de regresión en cada cuantil sobre la pendiente en cada variable (punteada negra), su recta de regresión (roja continua) e intervalos de confianza (punteada roja).

CONCLUSIONES Y RECOMENDACIONES

Todas las variables independientes son indispensables para predecir el puntaje global, cada una con diferente grado de contribución, calidad de asociación y peso sobre las combinaciones lineales de las dos primeras componentes, de igual forma confirma la alta capacidad predictiva de los modelos generados por regresión lineal múltiple y cuantílica.

Se recomienda aumentar el rendimiento de los estudiantes en las pruebas SABER PRO para evitar la posible influencia de factores externos en el promedio anual del puntaje global, lo anterior se podría lograr fortaleciendo aquellas competencias que no presentaron Contribución mayor a 50, como lo fue Lectura crítica, Comunicación escrita y Competencia ciudadana, de esta forma la variabilidad retenida en las dos componentes iniciales también sería mayor al 60%.

Aunque la variación entre medias y medianas no influyo en el cumplimiento de la normalidad y multicolinealidad entre variables, se recomienda la práctica de otras técnicas multivariadas con previa transformación de los datos para suplir el incumplimiento de la homocedasticidad entre predictores.

Es necesario establecer estudios con los diferentes indicadores académicos (Nivel de deserción, Matriculados, Puntaje Saber 11, entre otros) y la existencia de nuevas variables independientes en los modelos de regresión (Estrato socioeconómico, Edad, Genero, Estado civil, etc.), facilitando la corrección de aquellas falencias que se presentan desde un inicio de la carrera de pregrado y por ende la toma de decisiones en pro de una mayor calidad académica en los futuros profesionales.

REFERENCIAS BIBLIOGRÁFICAS

- [1]. Portal Único del Estado Colombiano. Gov.co. Resultados Saber Pro [Internet]. Bogotá, D.C.; [cited 2020 Mar 26]. Available from: <https://www.datos.gov.co/browse?q=saber-pro&sortBy=relevance%3E>
- [2]. Kassambara A. Practical Guide to Principal Component Methods in R. In: Multivariate Analysis II. STHDA; 2017. p. 170.
- [3]. Boccardo-Bosoni G, Ruiz-Bruzzone F. Estadística descriptiva con Rstudio/Forma de una distribución: simetría, curtosis y normalidad. In: RStudio para Estadística Descriptiva en Ciencias Sociales Manual de apoyo docente para la asignatura Estadística Descriptiva [Internet]. Segunda ed. RStudio/Bookdown; 2019 [cited 2020 Mar 20]. Available from: <https://bookdown.org/gboccardo/manual-ED-UCH/estadistica-descriptiva-con-rstudio.html#estimacion-puntual-de-estadisticos-descriptivos-usando-r>
- [4]. Salazar C. Cálculo de correlaciones [Internet]. RPbus.com. 2018 [cited 2020 Apr 4]. Available from: <https://rpubs.com/camilamila/correlaciones>
- [5]. Ortega-Martínez RM, Tuya-Pendás LC, Martínez-Ortega M, Pérez-Abreu A, Cánovas AM. El coeficiente de correlacion de los rangos de spearman caracterizacion. Revista Habanera de Ciencias Medicas. 2009;8(2): 19.
- [6]. Calderon S. Inflación de la varianza (VIF) [Internet]. RPbus.com. 2018 [cited 2020 Apr 15]. Available from: <https://rpubs.com/scalderon/434714>

- [7]. Amat-Rodrigo J. Apuntes varios(miscellaneous)/ Identificación de outliers, observaciones con alto leverage y observaciones influyentes. [Internet]. Correlación lineal y Regresión lineal simple. 2016 [cited 2020 Apr 15]. Available from: https://www.cienciadedatos.net/documentos/24_correlacion_y_regresion_lineal#identificación_de_outliers,_observaciones_con_alto_leverage_y_observaciones_influyentes
- [8]. Amat-Rodrigo J. RPubS - Introducción a la regresión lineal múltiple [Internet]. Rpubs.com. 2016 [cited 2020 Apr 3]. Available from: https://rpubs.com/Joaquin_AR/226291
- [9]. S-Cade B, R-Noon B. A gentle introduction to quantile regression for ecologists. *Front Ecol Environ.* 2003;1(8): 412–20.
- [10]. Das K, Krzywinski M, Altman N. Points of significance Quantile regression/Quantile regression robustly estimates the typical and extreme values of a response. *Nat Methods.* 2019;16(6): 451–2.
- [11]. ICFES. Informe nacional Saber Pro 2016 - 2018 1. Bogotá, D.C.; 2018.
- [12]. Orjuela J. Análisis del Desempeño Estudiantil en las Pruebas de Estado para Educación Media en Colombia mediante Modelos Jerárquicos Lineales. *Ingeniería.* 2013;18(2): 54–67.
- [13]. ICFES. Las diferencias y el contexto en los módulos específicos. Documento de análisis de resultados Examen SABER PRO 2017. Módulo de Análisis Económico. Colombia; 2018.
- [14]. Martínez CY, Mendoza LF. Análisis de los resultados de la evaluación en competencias genéricas de las pruebas SABER PRO 2014 en programas de Licenciatura en el área de las ciencias naturales y educación ambiental de cuatro universidades del País. Universidad Distrital Francisco José de Caldas; 2016.
- [15]. (ICFES) IC para la E de la E. Guía de Orientación. Saber Pro Módulos de Competencias Genéricas. Bogotá, D.C.; 2017. p. 90.
- [16]. Rodríguez Manrique JA, Ruiz Escorcía RR, Cohen Manrique CS. Análisis Multivariado Aplicado a la Evaluación de Competencias Saber-Pro en el Departamento de Sucre, Colombia. Lima, Peru; 2018. Report No.: 16.
- [17]. Marquín MJ. Predicción del rendimiento académico mediante técnicas del análisis multivariado en la asignatura de Álgebra Lineal [Tesis de Maestría en investigación de operaciones y estadística]. Universidad Tecnológica de Pereira UTP; 2017.
- [18]. Rodríguez Ayán MN. Análisis multivariado del desempeño académico de estudiantes universitarios de Química. Universidad Autónoma de Madrid; 2007.
- [19]. Pérez-Pulido MO, Aguilar-Galvis F, Orlandoni-Merli G, Ramoni-Perazzi J. Análisis estadístico de los resultados de las pruebas de estado para el ingreso a la educación superior en la Universidad de Santander, Colombia. *Rev Científica.* 2016;27: 328–39.