

FACTORES DEMOGRÁFICOS, SOCIOECONÓMICOS Y ACADÉMICOS QUE
INFLUYEN SOBRE LOS RESULTADOS DEL COMPONENTE GENÉRICO DE LA
PRUEBA SABER PRO

CASO: ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS DE LA
UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, SECCIONAL
DUITAMA

DORA PAOLA PACHECO MORENO

Código: 201120139

Trabajo de grado para optar al título de
Licenciatura en Matemáticas y Estadística

DIRECTORA DE LA INVESTIGACIÓN
M.Sc. CARMEN HELENA CEPEDA

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA
LICENCIATURA EN MATEMÁTICAS Y ESTADÍSTICA
DUITAMA

2018

Duitama, mayo 30 de 2018

A Dios por brindarme la oportunidad de obtener este triunfo y darme salud, sabiduría y entendimiento para lograr esta meta.

“Tan solo por la educación puede el hombre llegar a ser hombre. El hombre no es más que lo que la educación hace de él”, Immanuel Kant.

“Lo más difícil de aprender en la vida es qué puente hay que cruzar y qué puente hay que quemar”, Bertrand Russel.

AGRADECIMIENTOS

En primera instancia agradezco a Dios, a mis padres que con su amor, motivación y trabajo me educaron, a la Universidad por brindarme el espacio y la oportunidad en mi formación personal, académica y profesional, a mi esposo y mi adorada hija por ser mi fuente de motivación e inspiración para poderme superar cada día y luchar para que la vida nos conceda un futuro mejor y a todas aquellas personas que me guiaron y se esforzaron por ayudarme a llegar al punto en el que me encuentro, en especial a mi tutora M.Sc Carmen Helena Cepeda a quien le agradezco por su paciencia y dedicación.

RESUMEN ANALÍTICO ESPECIALIZADO – RAE

1. **Titulo.** FACTORES DEMOGRÁFICOS, SOCIOECONÓMICOS Y ACADÉMICOS QUE INFLUYEN SOBRE LOS RESULTADOS DEL COMPONENTE GENÉRICO DE LA PRUEBA SABER PRO. CASO: ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA SECCIONALES DUITAMA
2. **Autor (a).** PACHECO MORENO, Dora Paola
3. **Director (a).** CEPEDA ARAQUE, Carmen Helena
4. **Publicación.** Duitama. Universidad Pedagógica y Tecnológica de Colombia. 2.018
5. **Unidad patrocinante.** Universidad Pedagógica y Tecnológica de Colombia. Facultad seccional Duitama. Escuela de matemáticas y estadística.
6. **Palabras clave.** Investigación, Saber Pro, calidad de la educación, factores académicos, demográficos y socioeconómicos, modelo estadístico, Regresión Logística, GAMLSS (Modelos aditivos generalizados para localización escala y forma).
7. **Objetivo.** Determinar los factores demográficos, socioeconómicos y académicos que influyen en los resultados (desempeño y puntaje) del componente genérico de la prueba Saber Pro en el programa de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama.
8. **Descripción.** En este trabajo de investigación se presenta un análisis de las pruebas Saber Pro de los estudiantes de últimos semestres del programa de Administración de Empresas Agropecuarias desde el segundo semestre de 2011 hasta el segundo semestre de 2015. Como principal herramienta para este análisis se tuvo en cuenta los modelos aditivos generalizados para localización escala y forma, con el fin de modelar el puntaje de las pruebas Saber Pro en función de algunas variables que describen el entorno demográfico, socioeconómico y académico del estudiante.
9. **Fuentes.** Para llevar a cabo esta investigación se utilizó la información académica del programa de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Seccional Duitama reportada por el Instituto Colombiano para la Evaluación de la Educación ICFES, contenida en las bases de datos de la Prueba Saber Pro, en el periodo comprendido entre el segundo periodo de 2011 a 2015 y el diccionario de las variables suministrado también por el ICFES. Se dispuso de 222 filas por 31 columnas. Se empleó el software estadístico R 3.1.1 y el programa Microsoft Excel para el procesamiento de los datos. Además, se tomaron en cuenta varias investigaciones que sirvieron como objeto de referencia y que muestran estudios relacionados con el tema tratado, tales como; teoría sobre evaluación de la educación superior, teoría sobre modelos estadísticos, teoría sobre los GAMLSS, Modelo logístico, Regresión Logística Ordinal, entre otros.
10. **Contenido.** En el documento se presentan cinco capítulos, el primero presenta el marco general el cual explica la finalidad y el propósito de la investigación, el segundo es la planeación y construcción de la teoría, el tercero consiste en la metodología con la cual se trabajó, el cuarto trata de la descripción de los datos de estudio, el quinto se basa en el análisis de la información basados en los modelos Aditivos

Generalizados para Localización Escala y Forma, Regresión Logística Ordinal y finalmente las recomendaciones y conclusiones.

11. Metodología. En este trabajo de investigación se realizó un análisis descriptivo con enfoque cuantitativo sobre los resultados del componente genérico de la prueba Saber Pro de Administración de Empresas Agropecuarias desde el segundo semestre de 2011 hasta el segundo semestre de 2015, utilizando información suministrada por el ICFES. El análisis es de tipo descriptivo ya que se utilizaron distribuciones de frecuencia, inferencias, pruebas de hipótesis e independencia entre variables. Posteriormente, se utilizó el modelo logístico ordinal y los modelos GAMLSS para determinar los factores demográficos, socioeconómicos y académicos que más influyen sobre los resultados del componente genérico de la prueba Saber Pro. El proceso para llevar a cabo esta investigación fue la planeación y delimitación del problema de investigación; revisión y elaboración del marco teórico; aplicación de técnicas estadísticas a la base de datos y finalmente la elaboración del informe.
12. Conclusiones. Los factores demográficos que influyen en los resultados de Razonamiento Cuantitativo de la prueba Saber Pro en el programa de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, son la edad, el género y el ser cabeza de familia. En Lectura Crítica influyen la edad, el número de personas que conforma el hogar y número de personas de la que el estudiante se encuentra a cargo. En Comunicación Escrita influye solamente la edad. Ningún factor demográfico de los estudiados influye en los puntajes de Inglés.

Los factores socioeconómicos que influyen en los resultados de Comunicación Escrita de la prueba Saber son el valor anual de la matrícula del año anterior, la ocupación del padre, la ocupación de la madre, el estrato socioeconómico, tener o no celular, contar o no en el hogar con conexión a internet, televisión, teléfono fijo y lavadora, los Ingresos familiares mensuales familiares representado en salarios mínimos mensuales y el sí o no trabaja actualmente. En Inglés influye el valor anual de la matrícula del año anterior, la ocupación del padre, la ocupación de la madre, tener o no celular, contar o no en el hogar con conexión a internet, televisión, teléfono fijo y los Ingresos familiares mensuales familiares representado en salarios mínimos mensuales. En Lectura Crítica influyen el valor anual de la matrícula del año anterior, los Ingresos familiares mensuales familiares representado en salarios mínimos mensuales y el sí o no trabaja actualmente. En Razonamiento Cuantitativo influyen la ocupación de la madre y tener o no celular.

Los factores académicos que influyen en los resultados de Razonamiento Cuantitativo son el máximo nivel educativo alcanzado por la madre y máximo nivel educativo alcanzado por el padre. El Lectura Crítica influye el máximo nivel educativo alcanzado por la madre. En Inglés influye el máximo nivel educativo alcanzado por el padre. En Comunicación Escrita Influye el sí o no tomo curso de preparación.

En el caso de los desempeños, en el aspecto demográfico, la situación del hogar actual es la única variable que explica los desempeños Comunicación Escrita e Inglés. En cuanto al factor socioeconómico el desempeño en Comunicación Escrita es explicado, por la ocupación del padre y el desempeño en Inglés es explicado por el valor de la matrícula y sí o no cuenta con servicio con lavadora. Respecto a factor académico el desempeño en Inglés es explicado por la educación de la madre y del padre del estudiante y sí o no tomo curso de preparación para la prueba. Se considerada como un factor influyente en la determinación del rendimiento académico, así como generalmente a mayor nivel educativo de los padres mayor es el rendimiento académico de sus hijos.

13. Autor del RAE. Dora Paola Pacheco Moreno.

CONTENIDO

	Pág.
1. PRESENTACIÓN DEL PROYECTO.....	14
2. REFERENTES TEÓRICOS.....	20
2.1 LA EVALUACIÓN EN LA EDUCACIÓN SUPERIOR	20
2.2 EXAMEN SABER PRO	22
2.3 MODELOS LINEALES	25
2.4 MODELO LINEAL GENERALIZADO.....	26
2.5 MODELOS ADITIVOS GENERALIZADOS	28
2.6 MODELOS ADITIVOS GENERALIZADOS DE LOCALIZACIÓN, ESCALA Y FORMA.....	29
2.7 REGRESIÓN LOGÍSTICA ORDINAL	39
3. DISEÑO METODOLÓGICO	46
3.1 VARIABLES DE ESTUDIO	47
3.1.1 Factores demográficos	47
3.1.2 Factores socioeconómicos	48
3.1.3 Factores académicos.....	49
3.1.4 Variables del componente genérico	50
3.2 MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS	52
4. DESCRIPCIÓN DE LOS DATOS.....	53
4.1 DESCRIPCIÓN UNIVARIADA POR COMPONENTE.....	53
4.2 RESULTADOS EN COMPETENCIAS GENÉRICAS DE LA PRUEBA SABER PRO AÑO 2016 DEL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS.....	58

4.3	DESCRIPCIÓN UNIVARIADA DE ALGUNAS CARÁCTERÍSTICAS DE LOS ESTUDIANTES DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS	61
4.4	DESCRIPCIÓN BIVARIADA DE LAS CARACTERÍSTICAS DE LOS ESTUDIANTES	67
5.	FACTORES ASOCIABLES RESULTADOS DEL COMPONENTE GENÉRICO	78
5.1	ESPECIFICACIÓN DE LOS MODELOS	78
5.2	FACTORES DEMOGRÁFICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO.....	82
5.3	FACTORES SOCIOECONÓMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO.....	96
5.4	FACTORES ACADÉMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO	117
6.	DISCUSIÓN.....	131
7.	CONCLUSIONES	140
ANEXO		
ANEXOS 1. RESULTADOS ESTADÍSTICOS ANÁLISIS BIVARIADO.....		147

LISTA DE TABLAS E IMÁGENES

	Pág.
Tabla 1. Algunas distribuciones continuas	36
Tabla 2. Factores demográficos	47
Tabla 3. Factores socioeconómicos	48
Tabla 4. Factores académicos.....	50
Tabla 5. Componente genérico	51
Tabla 6. Resultados módulo Inglés	54
Tabla 7. Resultados de Desempeño en Inglés	55
Tabla 8. Resultados Razonamiento Cuantitativo	55
Tabla 9. Resultados del puntaje en Comunicación Escrita.....	56
Tabla 10. Resultados del Desempeño de Comunicación Escrita	57
Tabla 11. Puntaje en Lectura Crítica	57
Tabla 12. Puntajes 2016.....	60
Tabla 13. Desempeño de Comunicación Escrita 2016.....	60
Tabla 14. Desempeño en Inglés 2016.....	61
Tabla 15. Descripción factores demográficos.	61
Tabla 16. Descripción factores socioeconómicos.	63
Tabla 17. Descripción factores académicas.....	66
Tabla 18. Test no paramétrico de Kruskal Wallis. Puntaje en Razonamiento Cuantitativo	69
Tabla 19. Test no paramétrico de Kruskal Wallis. Puntaje en Inglés	71
Tabla 20. Test no paramétrico de Kruskal Wallis. Puntaje en Lectura Crítica	73
Tabla 21. Análisis de la varianza (ANOVA). Puntaje en Comunicación Escrita.....	75
Tabla 22. Componente sistemático - factor demográfico	82
Tabla 23. Modelo óptimo para el desempeño en Inglés - factor demográfico	95
Tabla 24. Modelo óptimo para el Desempeño en Comunicación Escrita - factor demográfico.....	95
Tabla 25. Componente sistemático - factor socioeconómico	96
Tabla 26. Modelo óptimo para el Desempeño en Inglés - factor socioeconómico.....	116
Tabla 27. Modelo óptimo para el Desempeño en Comunicación Escrita. factor socioeconómico	116

Tabla 28. Componente sistemático - factor académico	117
Tabla 29. Modelo óptimo para el desempeño en Inglés - factor académico.....	129
Tabla 30. Modelo óptimo para el desempeño en Comunicación Escrita -factor académico	130
Tabla 31. Factores demográficos que influyen en los puntajes del componente genérico.	132
Tabla 32. Factores socioeconómicos que influyen en los puntajes del componente genérico.....	134
Tabla 33. Factores académicos que influyen en los puntajes del componente genérico.	137
Imagen 5-1. Modelo óptimo para Razonamiento Cuantitativo.....	83
Imagen 5-2. Modelo óptimo para Inglés.....	87
Imagen 5- 3. Modelo óptimo Lectura Crítica	90
Imagen 5-4. Modelo óptimo para Comunicación Escrita	92
Imagen 5-5. Modelo óptimo para Razonamiento Cuantitativo.....	97
Imagen 5-6. Modelo óptimo para el puntaje en Inglés.....	101
Imagen 5-7. Modelo óptimo para Lectura Crítica	107
Imagen 5-8. Modelo óptimo Comunicación escrita	110
Imagen 5-9. Modelo óptimo para Razonamiento Cuantitativo.....	118
Imagen 5-10. Modelo óptimo para Inglés.....	122
Imagen 5-11. Modelo óptimo Lectura Crítica	124
Imagen 5-12. Modelo óptimo para Comunicación Escrita.....	127

LISTA DE FIGURAS Y GRÁFICAS

	Pág.
Figura 1. Función <code>plot.gamlss()</code>	35
Figura 2. Función <code>wp()</code>	35
Figura 3. Comparación de los puntajes del componente genérico.....	58
Gráfica 5-1. Ajuste distribución en Inglés.....	79
Gráfica 5-2. Ajuste distribución en Razonamiento Cuantitativo.....	80
Gráfica 5-3. Ajuste distribución en Lectura Crítica.....	81
Gráfica 5-4. Ajuste distribución en Comunicación Escrita	82
Gráfica 5-5. Diagnóstico modelo para puntaje en Razonamiento Cuantitativo - factor demográfico.....	86
Gráfica 5-6. Diagnóstico del modelo para puntaje en Razonamiento Cuantitativo - factor demográfico.....	86
Gráfica 5-7. Diagnóstico modelo para puntaje en Inglés - factor demográfico	88
Gráfica 5-8. Ajuste de la distribución de los residuales	89
Gráfica 5-9. Diagnóstico modelo para el puntaje en Lectura Crítica - factor demográfico (<i>plot</i> y el <i>worm plot</i>).....	92
Gráfica 5-10. Diagnóstico modelo para el puntaje en Comunicación Escrita - factor demográfico (<i>plot</i> y el <i>worm plot</i>)	94
Gráfica 5-11. Diagnóstico modelo para el puntaje en Razonamiento Cuantitativo - factor socioeconómico	99
Gráfica 5-12. Diagnóstico modelo para el puntaje en Razonamiento Cuantitativo - factor socioeconómico	100
Gráfica 5-13. Diagnóstico del modelo para el puntaje en Inglés - factor socioeconómico	105
Gráfica 5-14. Diagnóstico del modelo para el puntaje en Inglés - factor socioeconómico	106

Gráfica 5-15. Diagnóstico modelo para el puntaje en Lectura Crítica - factor demográfico (<i>plot</i> y el <i>worm plot</i>).....	110
Gráfica 5-16. Diagnóstico del modelo para el puntaje en Comunicación Escrita - factor socioeconómico	115
Gráfica 5-17. Diagnóstico modelo para puntaje en Razonamiento Cuantitativo - factor académico.....	121
Gráfica 5-18. Diagnóstico del modelo para puntaje en Razonamiento Cuantitativo – factor académico.....	121
Gráfica 5-19. Diagnóstico del modelo para el puntaje en Inglés - factor académico.....	124
Gráfica 5-20. Diagnóstico modelo para el puntaje en Lectura Crítica - factor académico (<i>plot</i> y el <i>worm plot</i>).....	126
Gráfica 5-21. Diagnóstico modelo para el puntaje en Comunicación Escrita - factor académico (<i>plot</i> y el <i>worm plot</i>).....	129

1. PRESENTACIÓN DEL PROYECTO

Este estudio se justifica desde los Lineamientos para Acreditación de Programas establecidos por la Comisión Nacional de Acreditación (CNA), los cuales indican en el factor de Procesos Académicos, en la característica de Integralidad del currículo, que es indispensable tener información detallada sobre “el desempeño de los estudiantes del programa en las Pruebas de Estado de Educación Superior, en los últimos cinco años”.

Para la Escuela de Administración de Empresas Agropecuarias de la Facultad Seccional Duitama de la Universidad Pedagógica y Tecnológica de Colombia, es pertinente realizar, por ejemplo, un análisis estadístico con enfoque descriptivo - cuantitativo para conocer los factores demográficos, socioeconómicos y académicos que influyen sobre los resultados del componente genérico de la prueba Saber Pro, a partir de información desde el segundo periodo de 2011 hasta el segundo semestre de 2015 suministrada por el ICFES, lo que permitirá proporcionar información en procura del potenciamiento de las capacidades de los estudiantes. En consecuencia, este trabajo investigativo pretende hacer un aporte a la concreción de los objetivos institucionales, generando reflexión y discusión alrededor del tema “Educación de Calidad” al interior de la Institución, permitiéndole además generar estrategias que potencien su modelo educativo, basadas en las causas que están afectando el resultado de las pruebas como las Saber Pro en los estudiantes.

El Ministerio de Educación Nacional en “Colombia, la mejor educada en el 2025”¹, afirma que el objetivo de la prueba Saber Pro es contribuir al mejoramiento de la calidad de la educación colombiana, mediante la realización de evaluaciones periódicas (censales y muestrales) en las que se valoran las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros. Los resultados de estas evaluaciones permiten que los establecimientos educativos, las secretarías de educación, el MEN y la sociedad en general conozcan cuáles son las fortalezas y debilidades y, a partir de estas, puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación. Su carácter periódico posibilita, además, valorar cuáles han sido los avances en un determinado lapso y establecer el impacto de programas y acciones específicas de mejoramiento. Teniendo en cuenta lo anterior es por eso que en esta investigación no solo trata de estimar y dar unos resultados, sino también, proporcionar información a la Universidad del rendimiento de sus estudiantes en la prueba Saber Pro y los principales factores que inciden en ella, para que de la misma forma puedan orientar a futuros estudiantes en la mejora de su aprendizaje.

Es de conocimiento de todos, la necesidad que existe tanto nacional como internacionalmente, que los niveles educativos estén en una constante mejora, por lo tanto, las instituciones que brindan formación a la población estudiantil realizan investigaciones para indagar y descubrir cuáles las debilidades o fallas que se presentan internamente, y, de esta manera tomar medidas pertinentes para poder alcanzar altos estándares de calidad.

¹ MINISTERIO DE EDUCACIÓN NACIONAL. Colombia, la mejor educada en el 2025. Líneas estratégicas de la política educativa del ministerio de educación nacional. 2014. Disponible en: <http://www.mineducacion.gov.co>

Rodríguez², llevo a cabo una investigación de tipo analítico descriptiva, reflexiva, empírica y con un enfoque cuantitativo sobre los determinantes del desempeño académico universitario. en el caso de la prueba ECAES -2009 en la región Caribe Colombiana. Los resultados obtenidos demuestran y confirman que existe una brecha de género a favor de los hombres y, por tanto, si el estudiante es hombre el rendimiento promedio en la prueba aumenta entre 1 y 3 puntos. En cuanto al nivel socioeconómico, tuvo una relación directa con el rendimiento académico. cuanto más alto es el nivel socioeconómico mejor es el rendimiento promedio en la prueba.

De igual manera Ramírez y Teichler³, realizaron una investigación de tipo analítico descriptiva, empírica e inferencial y con un enfoque cuantitativo sobre los factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba Saber Pro 2009 y concluyen que, el desempeño académico previo medido en los puntajes obtenidos en las pruebas SABER 11° está asociado fuerte y positivamente con el de desempeño de los estudiantes, sin importar su género o nivel de formación. Por su parte, las variables socioeconómicas y educativas estudiadas resultaron, en comparación con el desempeño previo, asociadas débilmente con el desempeño en la educación superior.

² RODRÍGUEZ, Gustavo. Determinantes del desempeño académico universitario. el caso de la Región Caribe Colombiana. ICFES: Saber Pro. Bogotá, D.C., noviembre de 2014.

³ RAMÍREZ, C. & TEICHLER, U. Factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba Saber Pro 2009. ICFES: Saber Pro. Bogotá, D.C., noviembre de 2014.

Escobar y Orduz⁴, llevaron a cabo un estudio sobre impacto que tienen las condiciones socioeconómicas del estudiante y las características de la institución educativa sobre el rendimiento académico del estudiante de educación superior, en la cual concluyen que los estudiantes con mejores condiciones socioeconómicas, forma de financiación de los programas, nivel educativo e ingreso de los padres y en general, las condiciones de la vivienda, que atienden instituciones con programas acreditados y que disponen de grupos de investigación, aumentan la probabilidad de obtener mejores resultados académicos.

De la misma forma Bahamón y Reyes⁵, realizaron un estudio sobre la caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012 y concluyen que los estudiantes con mejor desempeño tienen unas características sociodemográficas específicas, capacidad intelectual alta, buen desempeño académico, intereses profesionales relacionados con las áreas de la carrera y adecuados hábitos y técnicas de estudio.

Asimismo Osma, Mojica y Rivera⁶, realizaron una investigación sobre los factores asociados al rendimiento en las pruebas Saber Pro en estudiantes de ingeniería

⁴ ESCOBAR, S. & ORDUZ, M. Determinantes de la calidad en la educación superior en Colombia. Facultad de Ciencias Económicas y Administrativas. Pontificia Universidad Javeriana. Bogotá, D.C., 2013.

⁵ BAHAMÓN, M., & REYES, R. L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012. *Avances en Psicología Latinoamericana*, 32(3), 459-476.

⁶ OSMA, W., MOJICA, A., & RIVERA, T. Factores asociados al rendimiento en las pruebas Saber Pro en estudiantes de Ingeniería Civil en universidades colombianas. Unidad de Santander. *Innovaciencia. Facultad de ciencias. Exactas fis. Naturales*. 2014; 2(1): 17 – 24.

civil en universidades colombianas, concluyen mediante la prueba Ji-cuadrado de independencia que ni el título de bachiller obtenido por el estudiante, ni si el individuo trabaja o no, tienen relación con los resultados que obtiene en la pruebas, Por el contrario, para las variables género, tipo de educación del padre, tipo de educación de la madre y estrato influyen en los resultados.

El objetivo principal para este trabajo fue describir los factores demográficos, socioeconómicos y académicos según los resultados de las pruebas Saber Pro de los estudiantes del programa de Administración de Empresas Agropecuarias, presentadas entre el segundo periodo de 2012 y 2015. Los objetivos específicos fueron establecer los factores demográficos, socioeconómicos y académicos que influyen en el puntaje y desempeño de comunicación escrita e inglés, así como en el puntaje de lectura crítica y razonamiento cuantitativo y determinar los aspectos a mantener o mejorar en el programa de Administración de Empresas Agropecuarias tendientes a potenciar las competencias relacionadas con las competencias genéricas de la Prueba Saber Pro.

En esta investigación con enfoque descriptivo - cuantitativo se utilizó como metodología, distribuciones de frecuencia, inferencias, pruebas de hipótesis e independencia entre variables. Posteriormente, se utilizó un modelo logístico ordinal y los modelos aditivos generalizados de localización, escala y forma - GAMLSS, donde se llegó a determinar los factores demográficos, socioeconómicos y académicos que influyen sobre los resultados del componente genérico de la prueba Saber Pro durante el segundo periodo de 2011 a 2015 de los estudiantes del programa de Administración de Empresas Agropecuarias.

En el documento se presentan cinco capítulos, el primero presenta el marco general el cual explica la finalidad y el propósito de la investigación, el segundo es la planeación y construcción de la teoría, el tercero consiste en la metodología con la cual se trabajó, el cuarto trata de la descripción de los datos de estudio, el

quinto se basa en el análisis de la información basados en los modelos Aditivos Generalizados para Localización Escala y Forma, Regresión Logística Ordinal y finalmente las recomendaciones y conclusiones.

2. REFERENTES TEÓRICOS

A continuación, se muestran las teorías que sirvieron de referencia e investigaciones que muestran estudios relacionados con el tema tratado.

2.1 LA EVALUACIÓN EN LA EDUCACIÓN SUPERIOR

La importancia que han cobrado en las últimas décadas los procesos de evaluación está supeditada por la conexión establecida entre el saber y el capital. Debido a que se quiso establecer el conocimiento como generador de grandes dividendos financieros, las empresas empezaron a centrar su atención en los sistemas de educación para influir en los parámetros de formación de los futuros profesionales, esto gracias a la iniciativa y recomendación de organismos multilaterales como el Banco Mundial, la Organización Mundial del Comercio y la Organización de Cooperación para el Desarrollo Económico. Según lo anterior en las universidades el conocimiento se ha centrado en la innovación con un enfoque empresarial y capitalista.⁷

Guzmán y Serna⁸ afirman que la evaluación representa una manera de comprender la educación en general; es decir, el propósito y los objetivos con los cuales son diseñados los instrumentos para evaluar la calidad de la educación, reflejan los ámbitos de interés que han sido asignados a la educación. Según lo anterior, la evaluación de la educación superior es un proceso que debe motivar

⁷ GUZMÁN, C., SERNA, C. & HOYOS, D. (2012). Las pruebas ECAES en Colombia: una evaluación a la evaluación. Bogotá: Panorama.

⁸ Ibid., p. 24

hacia la reflexión de su quehacer y no simplemente calificar y jerarquizar de acuerdo con estándares de productividad.

La educación superior debe autoevaluarse y coevaluarse para examinar hasta qué punto está cumpliendo sus tareas con la máxima calidad posible; pero es indispensable que los instrumentos que se utilicen para esta evaluación correspondan a su naturaleza y a su complejidad esencial, y no establezcan identidades equivocadas y peligrosas entre los procesos de trabajo académico que se adelantan en la universidad y el proceso industrial de producción de mercancías. La caracterización cuidadosa de las tareas de la educación superior debería permitir señalar la legitimidad de los instrumentos de evaluación y hacer explícita su especificidad.⁹

Guzmán, Serna y Hoyos¹⁰, afirman que los sistemas de gestión de la calidad de la educación en Colombia se basan, fundamentalmente, en mediciones apoyadas en evaluaciones estandarizadas que determinan la proximidad o convergencia de las instituciones de educación superior a los ideales de la formación establecidos por los Estados; para nuestro caso, el Estado colombiano desarrolla esta función a través del MEN, el Instituto Colombiano para la Evaluación de la Educación (ICFES) y el Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS). Dentro de las mediciones y evaluaciones que realiza el Estado a través de los organismos mencionados, se encuentran el Examen de Calidad de Educación Media (Saber 11) y el Examen de Calidad de la Educación Superior (ECAES, hoy llamado Saber Pro). Este último se encarga de medir específicamente “el grado de desarrollo de las competencias de los estudiantes

⁹ MISAS, Gabriel. (2004). La educación superior en Colombia. Análisis y estrategias para su desarrollo. Bogotá: Universidad Nacional de Colombia.

¹⁰ GUZMÁN, C., SERNA, C. & HOYOS, D. (2012). Las pruebas ECAES en Colombia: una evaluación a la evaluación. Bogotá: Panorama.

que cursan el último año de los programas académicos de pregrado que ofrecen las Instituciones de Educación Superior”.

2.2 EXAMEN SABER PRO

Según el Decreto 1075 del 26 de mayo de 2015, desarrollado por el MEN, mediante artículo 2.5.3.4.1.1 afirma que el Examen de Estado de Calidad de la Educación Superior, es un instrumento estandarizado para la evaluación externa de la calidad de la Educación Superior. Forma parte, con otros procesos y acciones, de un conjunto de instrumentos que el Gobierno Nacional dispone para evaluar la calidad del servicio público educativo y ejercer su inspección y vigilancia. También expuesto en el decreto 3963 de 2009, artículo 1. Según el MEN los objetivos del Examen de Estado de Calidad de la Educación Superior son¹¹.

- Comprobar el grado de desarrollo de las competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior.
- Producir indicadores de valor agregado de la educación superior en relación con el nivel de competencias de quienes ingresan a este nivel; proporcionar información para la comparación entre programas, instituciones y metodologías, y mostrar su evolución en el tiempo.
- Servir de fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e instituciones de educación superior y del servicio público educativo, que fomenten la cualificación de los procesos institucionales y la formulación de políticas, y soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

¹¹ MINISTERIO DE EDUCACIÓN NACIONAL. Decreto 1075 de 2015. 2015. Disponible en: <http://www.mineducacion.gov.co/1759/w3-article-351080.html>

El Artículo 2.5.3.4.1.2 del Decreto 1075 del 26 de mayo de 2015 citando el Artículo 2° del Decreto 3963 de 2009, afirma que serán objeto de evaluación del Examen de Estado de Calidad de la Educación Superior las competencias de los estudiantes que están próximos a culminar los distintos programas de pregrado, en la medida en que éstas puedan ser valoradas con exámenes externos de carácter masivo, incluyendo aquellas genéricas que son necesarias para el adecuado desempeño profesional o académico independientemente del programa que hayan cursado. Las competencias específicas que se evalúen definidas por Ministerio Educación Nacional, con la participación la comunidad académica, profesional y del sector productivo, mediante mecanismos que defina el mismo Ministerio, teniendo en cuenta elementos disciplinares fundamentales la formación superior que son comunes a grupos de programas en una o más áreas del conocimiento.

En el Artículo 2.5.3.4.1.3 citando el Artículo 3 del Decreto 3963 de 2009, certifica que el examen está compuesto por pruebas que evalúan las competencias genéricas y las específicas. Todos los estudiantes deberán presentar los módulos de competencias genéricas, sin importar el programa de formación que cursen. Los módulos de competencias genéricas son. Escritura, Razonamiento Cuantitativo, Comunicación Escrita, Lectura Crítica, Competencias Ciudadanas e Inglés. Los módulos de competencias específicas corresponden a un saber hacer complejo en contexto y se definen con base en las competencias fundamentales de formación de grupos de programas o áreas de conocimiento.

El número de pruebas y componentes serán determinados por el ICFES mediante acuerdo de su Junta Directiva. La estructura de las pruebas de cada conjunto de competencias se establecerá de forma independiente y su adopción por el ICFES será gradual. Para efectos de la comparabilidad, cada una de ellas se mantendrá por lo menos 12 años a partir de la primera vez que se aplique a la población, sin perjuicio de que puedan introducirse modificaciones y mejoras, siempre que no afecten la comparabilidad de los resultados en el tiempo. El ICFES, con

fundamento en lo dispuesto en esta y en otras normas que la complementen, dirigirá y coordinará el diseño, la aplicación, la obtención y análisis de los resultados del Examen de Estado de Calidad de la Educación Superior, para lo cual podrá apoyarse en las comunidades académicas, profesionales y el sector productivo del orden nacional o internacional.¹²

Aspectos evaluados en los Saber Pro para administración

Teniendo en cuenta la Resolución 2767 de 2003 Orozco y Jiménez¹³, afirman que los profesionales en Administración deben adquirir y desarrollar competencias para.

- La comprensión de las organizaciones, su gerencia y el manejo de sus relaciones con entornos de gran complejidad dentro del tejido empresarial.
- La innovación, el liderazgo y el espíritu empresarial en la gestión de negocios de diversa naturaleza.
- La formación para el aprendizaje autónomo y para el desarrollo de habilidades de pensamiento, de interpretación, uso de información y de interrelación en procesos de trabajo con equipos interdisciplinarios.
- El desarrollo de competencias cognitivas y comunicativas en lengua materna y en una segunda lengua, así como las competencias socioafectivas necesarias para el ejercicio profesional.

¹² MINISTERIO DE EDUCACIÓN NACIONAL. Decreto 4216 del 30 de octubre de 2009. 2009. Disponible en: <http://www.icfes.gov.co/examenes/component/docman/doc download/71-decreto-4216-de-2009-modificacion-decreto-3963-reglamentacion-examen-saber-pro> [consultado 1 Jul 2012].

¹³ OROZCO, T. y JIMÉNEZ, R. (2009). Los ECAES y la cultura de evaluación en la educación superior. Universidad de Santo Tomas: RIIEP.

A continuación, se abordarán los referentes que tienen que ver con la parte estadística.

2.3 MODELOS LINEALES

El empleo de modelos es un proceso consustancial al ser humano. Para comprender lo que sucede a nuestro alrededor, a partir de la observación detallada de los acontecimientos se suele elaborar modelos mentales sobre cómo funcionan los fenómenos, pudiendo realizar, incluso, predicciones sobre ellos. En el ámbito científico, un modelo que explica un fenómeno suele expresarse de forma matemática (un modelo que, igualmente, ha sido derivado de descripciones y que probablemente será útil para predecir). McCullagh y Nelder citados por Díaz¹⁴, señalan que la construcción de modelos requiere de una mezcla de arte y conocimientos por parte del investigador.

Un modelo pretende explicar la variación de una respuesta a partir de la relación conjunta de dos fuentes de variabilidad, una de carácter determinista y otra aleatorio.

Los modelos estadísticos más sencillos son los modelos lineales, son un tipo de modelos de regresión univariados. Se considera que la variable de respuesta sigue una distribución normal y, además, que la relación entre las variables explicativas y la variable de respuesta tienen una forma lineal. Si las derivadas de la función respecto a cada uno de los parámetros β_j (efecto variables explicativas con $j = 1, \dots, r$)¹⁵.

¹⁴ DÍAZ, L. & MORALES, M. (2012). Análisis estadístico de datos categóricos. Universidad Nacional de Colombia. Primera edición, 2009. Bogotá, Colombia: ISBN 978-958-719-186-8.

¹⁵ PÉREZ, M. (2016). Modelos Aditivo Generalizados para Localización, Escala y Forma (GAMLSS). Trabajo para optar al título de máster. Universidad de Vigo. Pontevedra, España. 2016. p. 2

El Modelo Lineal - LM - se puede expresar de la siguiente manera.

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_r x_{ri} + \varepsilon_i$$

donde y_i representa la variable de respuesta del individuo i ($i = 1, \dots, n$), β_j representa el efecto sobre la variable explicativa j ($j = 1, \dots, r$), β_0 es el intercepto y ε_i es el error del individuo i .

Utilizando notación matricial, el modelo se puede reescribir de la siguiente manera.

$$Y = X\beta + \varepsilon$$

donde,

$$E[Y] = \mu = X\beta$$

donde X es una matriz $n \times p$ ($p = r + 1$) que contiene todas las variables explicativas (además de una columna de unos si se necesita la constante) el β es un vector desconocido de tamaño p que se estimará a partir de las variables explicativas.

2.4 MODELO LINEAL GENERALIZADO

Según Pérez¹⁶, el modelo lineal generalizado - GLM - fue introducido por Nelder y Wedderburn 1972 y desarrollado por McCullagh y Nelder 1989, los cuales definieron dos innovaciones principales en su enfoque.

- La distribución normal para la variable de respuesta Y es reemplazada por una distribución que pertenezca a la familia exponencial de densidades.

¹⁶ *Ibíd.*, p. 2

- La función de enlace monótona $g(\cdot)$, que se usa para modelar la relación entre μ_i (valor esperado de la variable de respuesta) y las variables explicativas.

De acuerdo con estas consideraciones, se puede escribir, usando una notación matricial, el modelo GLM de la siguiente manera.

$$y \sim \text{ExpF}(\mu, \phi)$$

$$g(\mu) = X\beta$$

donde la denominación *ExpF* se refiere a la familia exponencial, μ se refiere al vector esperado de y y ϕ es el parámetro de escala.

De acuerdo a Pérez¹⁷, para cada observación, la función de densidad de probabilidad de la familia exponencial puede escribirse de la siguiente manera.

$$f_Y(y, \mu, \sigma) = \exp \left[\frac{y\theta - b(\theta)}{\phi} + c(y \cdot \phi) \right]$$

Donde $E(y) = \mu = b'(\theta)$ y $Var(y) = \phi v(\mu)$ siendo $v(\mu) = b''[\theta(\mu)]$. θ , es el vector de todos los parámetros de la distribución y ϕ es el parámetro de escala. Esta función incluye muchas distribuciones importantes, incluidas la Normal, Poisson, Gamma, Inversa Gausiana y Tweedie, (Tweedie, 1984), distribuciones que tienen funciones de varianza $V(\mu) = 1, \mu, \mu^2, \mu^3$ $p > 0$ o $p > 1$.

Con la introducción de los modelos GLM, se consiguió flexibilizar considerablemente la restricción de que la distribución de la variable de respuesta tenía que ser necesariamente Normal.

¹⁷ *Ibíd.*, p. 3

2.5 MODELOS ADITIVOS GENERALIZADOS

Stasinopoulos y Rigby¹⁸, afirman que los modelos aditivos generalizados – GAM - se hicieron populares en la década de los 80. Los primeros en presentar estos modelos fueron Hastie y Tibshirani (1990) y, más tarde, fueron extendidos por Wood (2006). Aquellos modelos se consideraron una extensión no paramétrica de GLM; la idea de los modelos GAM es permitir que los propios datos sean los que determinen el predictor lineal η . Según Pérez¹⁹, el modelo GAM se puede escribir de la siguiente manera.

$$y \sim \text{ExpF}(\mu, \phi)$$
$$g(\mu) = X\beta + \sum_{j=1}^J h_j(x_j),$$

donde h_j son funciones de suavizado no paramétrico que se estiman a partir de las variables explicativas continuas.

Con la introducción de los modelos de GAM quedan flexibilizados a unas condiciones más restrictivas los LM que eran imposibles de considerar una familia distinta a una normal y de considerar una relación no lineal entre la variable de respuesta y las diferentes variables explicativas. Aun así, todo es mejor y con la idea de introducir nuevas distribuciones de la familia exponencial y de poder modelar el parámetro de escala y los parámetros de forma directamente de las variables explicativas, sin la necesidad de utilizar la relación existente entre el parámetro de localización y los parámetros restantes, aparecen los modelos GAMLSS.²⁰

¹⁸ STASINOPOULOS, M., RIGBY, B. y AKANTZILIOTOU, C. (2006). Instrucciones de cómo usar el paquete gamlss en R. Segunda edición. Centro de investigación de la Universidad Metropolitana de Londres. Londres, Inglaterra.

¹⁹ PÉREZ. Op. cit., p. 3.

²⁰ PÉREZ. Op. cit., p. 3

2.6 MODELOS ADITIVOS GENERALIZADOS DE LOCALIZACIÓN, ESCALA Y FORMA

Los modelos de localización escala y forma – GAMLSS - fueron introducidos por Rigby, Stasinopoulos y Akantziliotou²¹, como una forma de superar algunas de las limitaciones asociadas con los GLM y los GAM. Según Pérez²², los modelos GAMLSS son una clase de modelos univariantes. Además, los GAMLSS se pueden considerar como modelos de regresión semi-paramétricos. Son paramétricos en el sentido de que necesitan una distribución paramétrica para la variable de respuesta y semi en el sentido en el cual el modelado de los parámetros de la distribución en función de las variables explicativas puede implicar el uso de funciones suavizadas no paramétricas. Además, este autor también afirma que, en los GAMLSS, el supuesto de la distribución de la familia exponencial para la variable de respuesta y es reemplazado por una familia de distribución más general. La parte sistemática del modelo se amplía para permitir el modelado no solo para la media (o ubicación) sino la varianza (parámetro de escala), la asimetría y la kurtosis (parámetros de forma).

Stasinopoulos y Rigby²³, afirman que el modelo GAMLSS supone observaciones independientes y_i para $i = 1, 2, 3, \dots, n$ con función de densidad de probabilidad $f(y_i|\theta^i)$, donde $\theta^i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})$, es el vector de todos los parámetros de la distribución. Teniendo en cuenta que el GAMLSS se modela a partir de la media (parámetro de localización), la varianza (parámetro de escala), la asimetría y la kurtosis (parámetros de forma), Estos cuatro (4) parámetros denotados como μ

²¹ STASINOPOULOS Y RIGBY. Op. cit., p. 13

²² PÉREZ. Op. cit., p. 7

²³ STASINOPOULOS, M., RIGBY, B. & AKANTZILIOTOU, C. (2006). Instrucciones de cómo usar el paquete gamlss en R. Segunda edición. Centro de investigación de la universidad metropolitana de Londres. Londres, Inglaterra.

(media), σ (varianza) y ν (asimetría) y τ (kurtosis), la función de enlace monótona $g_k(\cdot)$ está dada de la siguiente forma.

$$y \sim D(\mu, \sigma, \nu, \tau)$$

$$g_1(\mu) = \eta_1 = X_1\beta_1 + \sum_{j=1}^{J_1} h_{j1}(x_{j1})$$

$$g_2(\sigma) = \eta_2 = X_2\beta_2 + \sum_{j=1}^{J_2} h_{j2}(x_{j2})$$

$$g_3(\nu) = \eta_3 = X_3\beta_3 + \sum_{j=1}^{J_3} h_{j3}(x_{j3})$$

$$g_4(\tau) = \eta_4 = X_4\beta_4 + \sum_{j=1}^{J_4} h_{j4}(x_{j4})$$

donde μ, σ, ν, τ son parámetros de la distribución, η_k el predictor para el k –ésimo parámetro de la distribución y x_{jk} es el vector de la variable explicativa, para $j = 1, 2, \dots, J_k$ y $k = 1, 2, 3, 4$ son vectores de tamaño n , β_k es el vector de parámetros.

La función h_{jk} es una función aditiva no paramétrica de la variable explicativa X_{jk} evaluado en x_{jk} . Los vectores explicativos x_{jk} se suponen fijos y conocidos.

Según Pérez²⁴, si queremos generalizar un poco la notación utilizada se puede escribir de la siguiente forma.

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk}),$$

donde θ_k es el vector de parámetros de la distribución.

²⁴ PÉREZ, M. (2016). Modelos Aditivo Generalizados para Localización, Escala y Forma (GAMLSS). Trabajo para optar al título de máster. Universidad de Vigo. Pontevedra, España. 2016. P. 6

En GAMLSS la estimación se llevará a cabo maximizando la siguiente función de probabilidad penalizada²⁵.

$$l_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} G_{jk} \gamma_{jk},$$

donde $l = \sum_{i=1}^n \log f(y_{i1}|\theta_i)$ es el logaritmo de la función de verosimilitud, λ_{jk} son los parámetros de penalización y G_{jk} es una matriz simétrica que depende de los parámetros λ_{jk} . Por lo tanto, según Pérez²⁶, se debe estimar β, λ y γ

Sea $M = \{D, G, T, \Lambda\}$ un modelo GAMLSS donde,

- D especifica la distribución de la variable de respuesta
- G especifica el conjunto de funciones enlace (g_1, \dots, g_p) para los parámetros $(1, \dots, p)$
- T engloba a las variables utilizadas en los predictores correspondientes
- Λ engloba los distintos parámetros de suavizado para las funciones de suavizado

Según Pérez²⁷, para un conjunto de datos en concreto, el proceso de selección consiste en comparar muchos modelos construidos combinando los distintos componentes de M . A continuación, se explicará cada uno de los componentes, expuestos por Pérez²⁸.

²⁵ Ibid., p. 6

²⁶ Ibid., p. 7

²⁷ Ibid., p. 7

²⁸ Ibid., p. 7,8

Componente D. La selección de una distribución apropiada se hace ajustando diferentes distribuciones y quedándonos con aquella que propicie un modelo con menor GAICG.

Componente G. La selección de la función de enlace está determinada generalmente por el rango de valores de la variable explicativa. Una buena elección de la función de enlace puede mejorar el ajuste en el modelo considerablemente; la elección de esta función enlace aparece usando el criterio deviance (se selecciona la que tenga menor deviance).

Componente T. La selección de los términos aditivos del modelo puede llevarse a cabo a través de procesos (forward, backward, o stepwise), además, estos procesos pueden aplicarse sobre cada parámetro por separado o sobre todos los parámetros a la vez.

Componente Λ . para cada término suavizado se necesita de su correspondiente parámetro de suavizado. Este puede ser previamente fijado o estimado a partir de los datos.

Según Pérez²⁹, para el ajuste del modelo lo primero que se debe hacer es ajustar el modelo sin variables explicativas, es decir buscar, la distribución que mejor se ajusta a la variable respuesta. Para este ajuste se puede utilizar el criterio de información Aikaike (AICG), el cual evalúa el ajuste de la distribución a los datos. Cuanto más pequeño es el AICG mejor es el ajuste.

Para Selección del modelo se debe tener en claro cuál es la variable de respuesta, para poder identificar las posibles variables explicativas. Después de tener en claro el modelo y su ajuste, de acuerdo a Pérez³⁰, se deben tener en cuenta las siguientes funciones.

²⁹ Ibid., p. 18

³⁰ Ibid., p. 22

- `stepGAICG()`. A través del criterio AICG podemos elegir el modelo más adecuado. Estas funciones solo realizan la selección del modelo teniendo en cuenta solo un parámetro de la distribución.

Si se pretende realizar una selección teniendo en cuenta todos los parámetros, las funciones que se deben usar son las que están expuestas a continuación.

- `stepGAICGAll.A()`. La estrategia utilizada es la siguiente.
 1. Utilizar los criterios de GAICG hacia adelante para seleccionar un modelo adecuado para μ (media, parámetro de localización), fijando σ (la varianza, parámetro de escala), ν (asimetría) y τ (kurtosis) parámetros de forma.
 2. Dado el modelo para μ obtenido en 1 y para ν y τ use un proceso de selección hacia adelante para elegir un modelo adecuado para σ .
 3. Dado los modelos para μ y σ obtenidos en 1 y 2 respectivamente y con τ , usamos un proceso de selección hacia adelante para elegir un modelo adecuado para ν .
 4. Dado los modelos para μ , σ y ν obtenidos en 1, 2 y 3 respectivamente, utilizamos un proceso de selección hacia adelante para elegir un modelo adecuado para τ .
 5. Dado los modelos para μ , σ y τ obtenidos en 1, 2 y 4 respectivamente, utilizamos un proceso de selección hacia atrás para elegir un modelo adecuado para ν .
 6. Dado los modelos para μ , ν y τ obtenidos en 1, 5 y 4 respectivamente, utilizamos un proceso de selección hacia atrás para elegir un modelo adecuado para σ .
 7. Dado los modelos para σ , ν y τ obtenidos en 6, 5 y 4 respectivamente, utilizamos un proceso de selección hacia atrás para elegir un modelo adecuado para μ y finalizamos el proceso.

- stepGAICGAll.B(). La estrategia B, a diferencia de la Estrategia A, obliga a que sean elegidos los mismos términos para todos los parámetros de la distribución, esto según Pérez³¹.

Una vez ajustado el modelo GAMLSS, para el diagnóstico como parte final, se utilizan los residuales cuantiles aleatorizados descritos por Dunn y Smyth (1996) y mencionados por Pérez, para verificar ajuste del modelo y, especialmente, la distribución escogida para la variable de respuesta y .

Según Pérez³², los residuos cuantílicos aleatorizados se expresan mediante expresión $r_i = \Phi^{-1}(\mu_i)$ donde Φ^{-1} es la función inversa de la distribución acumulativa de la variable normal estándar y μ_i se define como.

- $F(y_i|\hat{\theta}_i)$ si y_i es continua.
- Un valor aleatorio de la distribución uniforme en intervalo $\left[F\left((y_i - 1)|\hat{\theta}_i\right), F(y_i|\hat{\theta}_i) \right]$ si y_i es discreta.

La principal ventaja de estos nuevos residuos es que sea cual sea la distribución de la variable de respuesta, estos siempre tienen a una distribución normal estándar (media igual a 0 y varianza estimada menor a 1), siempre y cuando la distribución asignada la variable de respuesta sea apropiada³³.

De acuerdo a Pérez³⁴, la función `plot.gamlss()` se utiliza para identificar el comportamiento de los residuales de cuantiles normalizados (aleatorizados) y definir si el modelo tiene un buen ajuste o si es adecuado, la función `plot()` tiene como salida cuatro (4) gráficos, descritos a continuación.

³¹ Ibid., p. 39

³² Ibid., p. 8

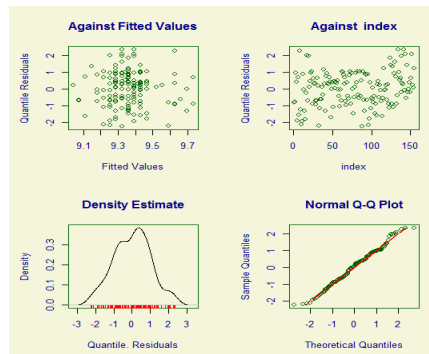
³³ Ibid., p. 9

³⁴ Ibid., p. 25-28

- Residuos contra los valores ajustados del parámetro μ
- Residuos contra un índice o una variable x especificada.
- Estimación de densidad de Kernel de los residuos.
- *QQ – plot* de los residuos.

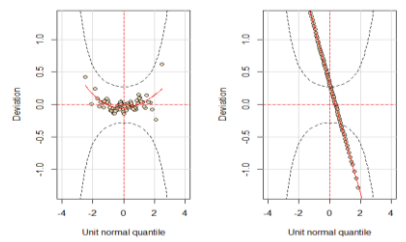
La salida de la función `plot()`, se puede ver a continuación.

Figura 1. Función `plot.gamlss()`



Según Stasinopoulos y Rigby³⁵, el worm plot, figura 2, de los residuos fue introducida por van Buuren y Fredriks [2001] con el fin de identificar regiones (intervalos) de una variable explicativa dentro de la cual el modelo no se ajusta adecuadamente a los datos. Para graficar la parcela de gusanos de los residuos se utiliza, en R, la función `wp()`. Ejemplo.

Figura 2. Función `wp()`



³⁵ STASINOPOULOS y RIGBY. Op. cit., p. 127

Según Pérez³⁶.

- Los puntos representan los residuos y la línea roja discontinua (horizontal) sus valores esperados, por lo tanto, podemos ver que tan lejos están unos de los otros.
- Las líneas negras discontinuas marcan el intervalo del 95%, para que el modelo sea correcto o adecuado, aproximadamente el 5% de los puntos pueden quedar fuera de este intervalo. Si esto no se debe considerar que el modelo es inadecuado para explicar la variable de respuesta.

Para el desarrollo de este trabajo se revisó primero el ajuste de los datos de las variables respuesta a una determinada distribución, a continuación, se describen las distribuciones objeto de estudio en este proyecto.

Tabla 1 . Algunas distribuciones continuas

DISTRIBUCIÓN	VARIABLE DE RESPUESTA	PARÁMETROS			
		μ	σ	ν	τ
Distribución Normal o Gaussiana (NO)	Puntaje Comunicación Escrita	<i>Identity</i>	<i>log</i>		
Distribución Logística (LO)	Puntaje Lectura Crítica	<i>Identity</i>	<i>log</i>		
Distribución Johnson Su (JSU)	Puntaje Inglés	<i>Identity</i>	<i>log</i>	<i>Identity</i>	<i>log</i>
Distribución Skew <i>t</i> Type 2 (ST2)	Puntaje Razonamiento Cuantitativo	<i>Identity</i>	<i>log</i>	<i>Identity</i>	<i>log</i>

Fuente. PÉREZ, M. (2016). Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS). Trabajo para optar el título de máster. Universidad de Vigo. Pontevedra, España. 2016.

³⁶ PÉREZ. Op. cit., p. 28

Distribución Normal o Gaussiana (NO)

Según Stasinopoulos y Rigby³⁷, la distribución normal es el valor por defecto del argumento *family* de la función `gamlss()`. La parametrización utilizada para la función de densidad de probabilidad normal (o gaussiana), denotada por $NO(\mu, \sigma)$, es.

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right]$$

Para $-\infty < y < \infty$ donde $-\infty < \mu < \infty$ y $\sigma > 0$. Entonces la media de Y está dada por $E(Y) = \mu$ y la varianza de Y por $Var(Y) = \sigma^2$, así μ es la media y σ es la desviación estándar de Y .

Distribución Logística (LO)

Según Stasinopoulos y Rigby³⁸, la distribución logística es apropiada para datos moderadamente kurtóticos. La parametrización de la distribución logística, indicada aquí como $LO(\mu, \sigma)$, viene dada por.

$$f_Y(y|\mu, \sigma) = \frac{1}{\sigma} \left\{ \exp\left[-\left(\frac{y - \mu}{\sigma}\right)\right] \right\} \left\{ 1 + \exp\left[-\left(\frac{y - \mu}{\sigma}\right)\right] \right\}^2$$

Para $-\infty < y < \infty$ donde $-\infty < \mu < \infty$ y $\sigma > 0$, con $E(Y) = \mu$ y $Var(Y) = \frac{\pi^2 \sigma^2}{3}$, Johnson et al. (1995) p. 116.

Distribución Johnson Su (JSU)

³⁷ STASINOPOULOS y RIGBY. Op. cit., p. 181

³⁸ STASINOPOULOS y RIGBY. Op. cit., p. 182

Según Stasinopoulos y Rigby³⁹, esta es una reparametrización de la distribución original de Johnson Su, Johnson (1949), de modo que los parámetros μ y σ son la media y la desviación estándar de la distribución. El parámetro ν determina la asimetría de la distribución con $\nu > 0$ que indica un sesgo positivo y $\nu < 0$ negativo. El parámetro τ determina la kurtosis de la distribución. τ , Debería ser positivo y muy probablemente en la región superior a 1. Como $\tau \rightarrow \infty$, la distribución se aproxima a la función de densidad normal. La distribución es apropiada para datos leptocúrticos. La función de densidad de probabilidad de Johnson, denotado aquí como $JSU(\mu, \sigma, \nu, \tau)$, está definido por.

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{\tau}{c\sigma} \frac{1}{(r^2 + 1)^{\frac{1}{2}}} \frac{1}{2\pi} \exp\left[-\frac{1}{2}z^2\right]$$

Para $-\infty < y < \infty$ donde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$ y donde

$$z = -\nu + \tau \sinh^{-1}(r) = -\nu + \log\left[r + (r^2 + 1)^{\frac{1}{2}}\right], \text{ siendo}$$

$$r = \frac{y - \left(\mu + c\sigma\omega^{\frac{1}{2}} \sinh(\Omega)\right)}{c\sigma}, \quad c = \left\{\frac{1}{2}(\omega - 1)[\omega \cosh(2\Omega) + 1]\right\}^{-\frac{1}{2}}$$

$$\omega = \exp(1/\tau^2), \quad Z \sim NO(0,1), \quad \Omega = -\nu/\tau, \quad E(Y) = \mu \text{ y } Var(Y) = \sigma^2.$$

Distribución Skew t Type 2 (ST2)

Denotada por $ST2(\mu, \sigma, \nu, \tau)$, según Stasinopoulos y Rigby⁴⁰ esta distribución está definida por.

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{2}{\sigma} f_{Z_1}(Z) f_{Z_2}(\omega)$$

³⁹ STASINOPOULOS y RIGBY. Op. cit., p. 182

⁴⁰ STASINOPOULOS y RIGBY. Op. cit., p. 188

Para $-\infty < y < \infty$ donde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < v < \infty$, $\tau > 0$ y donde $Z = (y - \mu)/\sigma$, $\omega = v\lambda^{\frac{1}{2}}Z$ y $\lambda = (\tau + 1)/(\tau + Z^2)$ y f_{Z_1} es la función de densidad de probabilidad de $Z_1 \sim TF(0,1,\tau)$ y F_{Z_1} es el cdf de $Z_2 \sim TF(0,1,\tau + 1)$ $Z_2 \square TF(0,1,\tau + 1)$. Esta distribución es el caso univariante de la skew distribución t multivariada introducida por Azzalini y Capitanio (2003).

Aquí la media y la varianza de Y están dadas por $E(Y) = \mu + \sigma E(Z)$ y $Var(Y) = \sigma^2 V(Z)$, donde $E(Z) = v\tau^{\frac{1}{2}}\Gamma\left(\frac{\tau-1}{2}\right)/(\pi^{\frac{1}{2}}(1+v^2)^{\frac{1}{2}}\Gamma(\tau/2))$, para $\tau > 1$ y $E(Z^2) = \tau/(\tau - 2)$, para $\tau > 2$.

2.7 REGRESIÓN LOGÍSTICA ORDINAL

Según Agresti⁴¹, para una variable de respuesta binaria Y y una variable explicativa X , el modelo de regresión logística está dado por.

$$P[Y = 1] = \frac{\exp(\beta_0 + \beta x)}{1 + \exp(\beta_0 + \beta x)}$$

De igual forma, para este autor, para la regresión logística, con una respuesta binaria, se propone que el logaritmo de la “razón de probabilidad” (odds), entendido como el cociente entre la probabilidad de éxito y la de fracaso en un ensayo de Bernoulli, es igual a una función lineal en los parámetros, denominada usualmente predictor lineal. En el caso particular de la regresión logística ordinal – RLO - se utiliza la función de enlace para relacionar de forma lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la

⁴¹ AGRESTI, A. (2002). Categorical Data Analysis. University of Florida, Gainesville, Florida. Second edition. Published simultaneously in Canada. ISBN 0-471-36093-7

categoría i de la variable de respuesta, y la probabilidad que la variable tome un valor mayor que la categoría i ⁴².

La regresión logística ordinal se aplica cuando se puede establecer un orden natural en las categorías de la variable de respuesta. La finalidad analítica de la regresión logística se enmarca en la predicción y el análisis de un problema contextualizado en una situación real.

En la RLO las funciones más utilizadas son la Logit y la Cloglog. Según McCullagh (1980) citado por Agresti⁴³, plantea que el enlace Logit es el más adecuado para analizar datos ordinales, cuya distribución de frecuencia es uniforme a lo largo de todas las categorías, mientras que el enlace Cloglog es preferible para analizar datos categóricos, cuyas categorías de mayor valor son las más probables.

Siguiendo con Agresti⁴⁴, la expresión de la función Logit para la RLO es la siguiente.

$$\ln(O_i) = \alpha_i + \beta x \quad (1)$$

En esta ecuación, O_i es la “razón de probabilidad” (odds) asociada a la categoría i de la variable dependiente, siendo la expresión de esta razón.

$$O_i = \frac{P(\text{valor sea} \leq \text{categoría } i \mid \text{valores de } X)}{P(\text{valor sea} > \text{categoría } i \mid \text{valores de } X)} \quad (2)$$

Lo que es lo mismo que.

⁴² *Ibíd.*, p. 166

⁴³ AGRESTI. *Op. cit.*, p. 275

⁴⁴ AGRESTI. *Op. cit.*, p. 275

$$O_i = \frac{P(\text{valor sea} \leq \text{categoria } i \mid \text{valores de } X)}{1 - P(\text{valor sea} \leq \text{categoria } i \mid \text{valores de } X)} \quad (3)$$

Según Agresti⁴⁵, con el término “valor” en las expresiones (2) y (3) se hace referencia a cualquier valor de la variable dependiente. Como se observa, las probabilidades de las expresiones (2) y (3) son condicionales, es decir, dados los valores de la (s) variable(s) independiente(s). En la ecuación (1), α_i es el intercepto asociado a la ecuación que modela la razón de probabilidad de la categoría i , y β es el coeficiente de la ecuación de regresión. Si existen p variables independientes, existen p coeficientes, y βX se reemplaza por la combinación lineal $\beta_1 X_1 + \dots + \beta_p X_p$. Estos coeficientes cuantifican el efecto de las variables independientes sobre el logaritmo de la razón de probabilidad.

Según Hosmer y Lemeshow⁴⁶, si la variable dependiente tiene k categorías, existen $k - 1$ ecuaciones ya que a la categoría mayor no se asocia odds, al ser la probabilidad acumulada hasta ésta igual a uno.

El modelo de regresión ordinal descrito anteriormente es denominado “modelo logit acumulado”, ya que es construido basándose en las probabilidades acumuladas de la variable de respuesta, dados los valores de las variables explicativas. También es denominado modelo de razón de probabilidad proporcional, ya que los coeficientes de regresión son independientes de las categorías de la variable dependiente, siendo los mismos en las $k-1$, ecuaciones que se forman para las categorías. Esto implica asumir que la relación entre las variables explicativas y la variable dependiente ordinal es independiente de las categorías de esta última, y, por tanto, que los cambios en las variables explicativas provocan el mismo cambio en la razón de probabilidad acumulada de todas las categorías. por lo tanto,

⁴⁵ AGRESTI. Op. cit., p. 276

⁴⁶ HOSMER, D. y LESMESHOW, S. (2000). Applied logistic regression. Segunda edición, 2000. Canadá: ISBN 0-471-35632-8

existen ecuaciones $k-1$ con los mismos coeficientes acompañando a las variables explicativas y que sólo se diferencian en el valor del intercepto⁴⁷.

Según Hosmer y Lesmeshow⁴⁸, para estimar los coeficientes de la ecuación de regresión se utilizan varios procedimientos, siendo la estimación de máxima verosimilitud el más empleado. Una vez obtenida la ecuación, el primer paso es comprobar la significancia global del modelo, lo que significa evaluar si en conjunto las variables regresoras tienen un efecto estadísticamente significativo sobre la variable dependiente. Con este fin se emplea el estadístico G , calculado según la fórmula.

$$G = -2 \ln \left[\frac{\text{verosimilitud de la muestra sin variables explicativas}}{\text{verosimilitud con la ecuación que incluye las variables}} \right]$$

McCullagh citado por Hosmer y Lesmeshow, afirma que, este estadístico sigue distribución *chi – cuadrado*, con tantos grados de libertad como variables explicativas existan. Cuando las predicciones de la variable dependiente que se hacen con el modelo que incluye todas las variables independientes superan las predicciones que se realizan sin considerar éstas, el valor de G tiende a ser grande y conlleva a concluir que al menos una de las variables regresoras tiene efecto sobre la variable dependiente, y, por lo tanto, que la probabilidad de ocurrencia de los valores que representan esta variable varía para alguna de las combinaciones de valores de las variables independientes. La significancia individual de cada variable independiente, generalmente, se analiza a través de la prueba de Wald, la cual se basa en la significancia del coeficiente estimado para cada variable⁴⁹. El estadístico utilizado es el siguiente.

⁴⁷ AGRESTI. Op. cit., p. 276

⁴⁸ HOSMER, D. y LESMESHOW. Op. cit., p. 13

⁴⁹ HOSMER, D. y LESMESHOW. Op. cit., p. 13

$$Z_{Wald} = \frac{b_j}{DES(b_j)}$$

Para Hosmer y Lesmeshow⁵⁰, b_j es el coeficiente de regresión estimado para la variable independiente j . Bajo la hipótesis de que el coeficiente poblacional $\beta_j = 0$ para la variable j , la razón entre la estimación de este coeficiente b_j y el error estándar de esta estimación $DES(b_j)$, sigue una distribución normal estándar. Mayores valores de este estadístico indican que el coeficiente β_j es distinto de cero, y, por ende, que la variable independiente tiene efecto sobre la probabilidad de ocurrencia de los valores de la variable dependiente. También es necesario conocer si el modelo que se obtiene presenta buen ajuste. El empleo de una prueba de bondad de ajuste permite saber si la frecuencia predicha para las combinaciones según el modelo difiere significativamente de la frecuencia con la cual ocurren realmente los valores en estas combinaciones, lo cual evidencia falta de ajuste. Según Hosmer y Lesmeshow⁵¹, para comparar estas frecuencias generalmente se calcula el estadístico chi-cuadrado de Pearson según la fórmula.

$$\chi^2 = \sum_{i=1}^k \sum_{l=1}^m \frac{(y_{il} - m_l p_{il})^2}{m_l p_{il} (1 - p_{il})}$$

- y_{il} es la frecuencia observada de la i –ésima categoría de la variable dependiente en la l –ésima combinación de valores de las variables explicativas.
- p_{il} es la probabilidad estimada con el modelo para la i –ésima categoría de la variable dependiente en la l –ésima combinación de valores de las variables independientes.

⁵⁰ HOSMER, D. y LESMESHOW. Op. cit., p. 16

⁵¹ HOSMER, D. y LESMESHOW. Op. cit., p. 17

- m_l es la cantidad de elementos en la l – ésima combinación de valores de las variables explicativas.

Mientras mayor es el valor del estadístico χ^2 mayor sospecha de falta de ajuste.

Si finalmente se concluye la existencia de relación entre las variables explicativas y la dependiente, y si la ecuación lograda presenta buen ajuste, entonces se pueden hacer otros análisis, por ejemplo, para obtener la razón de probabilidad acumulada de la categoría i de la variable dependiente para determinados valores de las independientes, se despeja esta razón de la función logarítmica de forma que⁵².

$$\frac{P(\text{valor sea } \leq \text{categoría } i \mid \text{valores de } X)}{P(\text{valor sea } > \text{categoría } i \mid \text{valores de } X)} = e^{\alpha_i + \beta X}$$

Con el término “valor” en la anterior expresión se hace referencia a cualquier valor de la variable dependiente. Por lo tanto, se deriva que.

$$P(\text{valor sea } \leq \text{categoría } i \mid \text{valores de } X) = (e^{\alpha_i + \beta X}) / (1 + e^{\alpha_i + \beta X})$$

Finalmente se deduce que.

$$P(\text{valor sea } = \text{categoría } i \mid \text{valores de } X) \\ = P(\text{valor sea } \leq \text{categoría } i) - P(\text{valor sea } \leq \text{categoría } i - 1)$$

Esta expresión posibilita estimar a través de la ecuación obtenida y dado un conjunto de valores de las variables regresoras, la probabilidad que la dependiente tome cada uno de sus valores. También suele calcularse la ratio de la razón de probabilidad (odds ratio) que provoca el cambio en cada una de las variables independientes. El odds ratio de la variable independiente x evalúa la relación

⁵² AGRESTI. Op. cit., p. 276

entre la razón de probabilidad asociada a la categoría i cuando $x = x_2$, y la razón de probabilidad asociada a la categoría i cuando $x = x_1$. Matemáticamente sería⁵³.

$$\text{odds ratio} = \frac{P(Y \leq i|X = X_2) / P(Y > i|X = X_2)}{P(Y \leq i|X = X_1) / P(Y > i|X = X_1)}$$

Como el efecto que tiene una determinada variable predictora es el mismo para todas las categorías de la variable dependiente, para cada variable independiente se determina un solo odds ratio. El odds ratio es utilizado para interpretar el efecto de las variables explicativas sobre la variable objeto de estudio. Si éste es igual a uno, indica que la variable predictora no tiene efecto. Si es menor que uno, lo cual sucede cuando el coeficiente de la variable regresora es negativo, indica que, si las otras variables explicativas permanecen constantes, los cambios en la variable explicativa analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable objeto de estudio. Valores de odds ratio mayores que uno muestran que las variaciones en la variable independiente disminuyen la probabilidad de obtener categorías de mayor valor de la dependiente⁵⁴.

⁵³ AGRESTI. Op. cit., p. 45

⁵⁴ HOSMER, D. y LESMESHOW. Op. cit., p. 50,53

3. DISEÑO METODOLÓGICO

En este trabajo de investigación se realizó un análisis descriptivo con enfoque cuantitativo sobre los resultados del componente genérico de la prueba Saber Pro de Administración de Empresas Agropecuarias desde el segundo semestre de 2011 hasta el segundo semestre de 2015, utilizando información suministrada por el ICFES. El análisis es de tipo descriptivo ya que se utilizaron distribuciones de frecuencia, inferencias, pruebas de hipótesis e independencia entre variables. Posteriormente, se utilizó el modelo logístico ordinal y el modelo GAMLSS para determinar los factores demográficos, socioeconómicos y académicos que más influyen sobre los resultados del componente genérico de la prueba Saber Pro. El proceso para llevar a cabo esta investigación fue la planeación y delimitación del problema de investigación; revisión y elaboración del marco teórico; aplicación de técnicas estadísticas a la base de datos y finalmente la elaboración del informe.

Para la población objeto de estudio se tomó en cuenta a los estudiantes de últimos semestres de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Seccional Duitama que presentaron las pruebas Saber Pro, desde el segundo semestre de 2011 hasta el segundo semestre de 2015. Se dispuso de la base de datos suministrada por el ICFES, la cual está conformada por 222 observaciones y 31 variables. Dentro de las variables independientes se incluyeron aquellas que describen el entorno socioeconómico, demográfico y académico del estudiante y como variable dependiente se tomó en cuenta el componente genérico. Se empleó el software estadístico R 3.1.1 y el programa Microsoft Excel para el procesamiento de los datos. Además, se revisaron varias investigaciones que sirvieron como objeto de referencia y que muestran estudios relacionados con el tema tratado, tales como; teoría sobre evaluación de la educación superior, teoría sobre modelos

estadísticos, teoría sobre los GAMLSS, Modelo logístico, Regresión Logística Ordinal, entre otros.

3.1 VARIABLES DE ESTUDIO

La variable dependiente es el componente genérico de la prueba Saber Pro de Administración de Empresas Agropecuarias durante el periodo de 2012 a 2015. Las variables independientes o explicativas son los factores demográficos, socioeconómicos y académicos de los estudiantes.

3.1.1 Factores demográficos.

Los factores demográficos son los relativos a los aspectos y modelos culturales, creencias, actitudes, y características que identifican una población, ejemplo. edad, sexo, educación, ingresos, estado civil, trabajo, religión, tasa de natalidad, tasa de mortalidad y tamaño de la familia.

Tabla 2. Factores demográficos

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Género Cod. ESTU_GENERO	"F" Femenino, "M" Masculino	Cualitativa nominal
Edad Cod. ESTU_EDAD	Edad en años	Cuantitativa
Estado civil Cod. ESTU_ESTADO_CIVIL	Estado civil; "C" Casado; "SLT" Soltero; "SP" Separado; "UL" Unión libre.	Cualitativa nominal

Continuación tabla anterior

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Número de personas en el hogar Cod. FAMI_NUM_PERS_GRUP_FAM	Número de personas que conforman el hogar	Cuantitativa
Es cabeza de familia Cod. ESTU_SN_CABEZA_FMLIA	Es cabeza de familia."0" No, "1" Si	Cualitativa nominal
Número de personas a cargo Cod. FAMI_NUM_PERS_CARGO	Número de personas de las que usted se encuentra a cargo	Cuantitativa
Situación del hogar Cod. ESTU_HOGAR_ACTUAL	Situación de su hogar actual. "H" Es habitual o permanente; "T" Temporal por razones de estudio u otra razón.	Cualitativa nominal

3.1.2 Factores socioeconómicos.

Los factores socioeconómicos son las experiencias sociales y económicas y las realidades que ayudan a moldear la personalidad, las actitudes y la forma de vida.

Tabla 3. Factores socioeconómicos

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Forma de pago de la matrícula Cod. FORMA_PAGO_MATRICULA	Los recursos con los que canceló la matrícula el año pasado. Provinieron de beca "B"; Ayuda de los padres "PD"; Recursos propios "PP"; "VFP" Varias formas de pago.	Cualitativa nominal
Ocupación del padre Cod.FAMI_COD_OCUP_PADRE	Ocupación del padre; "E" Empleado; "T" Trabajador independiente; "H" Hogar; "P" Pensionado.	Cualitativa nominal
Ocupación de la madre Cod. FAMI_COD_OCUP_MADRE	Ocupación de la madre; "E" Empleada; "T" Trabajadora independiente; "H" Hogar; "P" Pensionada.	Cualitativa nominal
Valor anual de la matrícula del año anterior Cod. INST_VLR_MATRICULA_ANTI OR	"0" No pagó matrícula; "1" Menos de 500 mil; "2" Entre 500 mil y Menos de 1 millón; "3" Entre 1 millón y Menos de 3 millón.	Cualitativa ordinal

Continuación tabla anterior

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Estrato Cod. ESTU_ESTRATO	Estrato socioeconómico de la residencia del estudiante según factura de energía; "1" Estrato 1; "2" Estrato 2; "3" Estrato 3; "4" Estrato 4.	Cualitativa ordinal
El hogar cuenta con celular Cod. ECON_SN_CELULAR	El hogar cuenta con celular. "0" No; "1" Si	Cualitativa nominal
El hogar cuenta con conexión a internet Cod. ECON_SN_INTERNET	El hogar cuenta con conexión a internet. "0" No; "1" Si	Cualitativa nominal
El hogar cuenta con Servicio cerrado de televisión Cod. ECON_SN_SERVICIO_TV	El hogar cuenta con Servicio cerrado de televisión. "0" No; "1" Si	Cualitativa nominal
El hogar cuenta con servicio de teléfono fijo Cod. ECON_SN_TELEFONIA	El hogar cuenta con servicio de teléfono fijo. "0" No; "1" Si	Cualitativa nominal
El hogar cuenta con lavadora Cod. ECON_SN_LAVADORA	El hogar cuenta con lavadora. "0" No; "1" Si	Cualitativa nominal
El hogar cuenta con computador Cod. ECON_SN_COMPUTADOR	Tiene computador en su hogar. "0" No; "1" Si	Cualitativa nominal
Ingresos familiares mensuales Cod. FAMI_ING_FMLIAR_MENSUAL	Ingresos familiares mensuales representado en salarios mínimos mensuales, "1" Menos de 1 SM; "2" Entre 1 y Menos de 2 SM; "3" Entre 2 y Menos de 3 SM; "4" Entre 3 y Menos de 7 SM.	Cualitativa ordinal
Situación actual del trabajo Cod. ESTU_TRABAJO	Trabaja actualmente "0" No; "1" Si, como ayudante sin remuneración, "2" Si, por ser práctica obligatoria del plan de estudios; "3" Si, para adquirir experiencia y/o recursos para sus gastos personales	Cualitativa nominal

3.1.3 Factores académicos.

Los factores académicos son aquellos que identifican las características de un estudiante en formación. A continuación, se detallan.

Tabla 4. Factores académicos

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
El estudiante tomó algún curso de preparación Cod. ESTU_TOMOCURSO	Tomó algún curso de preparación para el examen. "0" No; "1" Si	Cualitativa Nominal
Tipo de bachillerato Cod. ESTU_TITULO_BTO	Tipo de bachillerato del que se graduó el evaluado. "A" Académico; "N" Técnico; "T" Normalista Superior; Desconocido	Cualitativa Nominal
Nivel educativo alcanzado por el padre Cod. FAMI_COD_EDUCA_PADRE	Máximo nivel educativo alcanzado por el padre. "0" Ninguno; "1" Básica primaria; "2" Básica secundaria; "3" Educación superior.	Cualitativa Ordinal
Nivel educativo alcanzado por la madre Cod. FAMI_COD_EDUCA_PADRE	Máximo nivel educativo alcanzado por el padre. "0" Ninguno; "1" Básica primaria; "2" Básica secundaria; "3" Educación superior.	Cualitativa Ordinal

3.1.4 Variables del componente genérico.

Las competencias genéricas que se tuvieron en cuenta para esta investigación fueron Puntaje y Desempeño en Comunicación Escrita e Inglés, Puntaje en Razonamiento Cuantitativo, Puntaje en Lectura Crítica.

Tabla 5. Componente genérico

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Puntaje Comunicación Escrita Cod. MOD_COMUNICA_ESCRITA_PUNT	Este resultado se presenta en una escala con media 10 y desviación estándar 1. Está representado en sixtiles. >9 "Muy bajo" 9-9.5 "Bajo" 9.5-10.2 "Regular" 10.3-10.6 "Bueno" 10.7-11 "Muy bueno" >11 "Excelente"	Cuantitativa
Desempeño Comunicación Escrita Cod. MOD_COMUNICA_ESCRITA_DESEM	Toma ocho niveles de valoración. 1, 2, 3, 4, 5, 6, 7, 8. Con uno el más bajo y ocho el más alto.	Cualitativa ordinal
Puntaje Inglés Cod. MOD_INGLÉS_PUNT	Este resultado se presenta en una escala con media 10 y desviación estándar 1. Está representado en sixtiles. >9 "Muy bajo" 9-9.5 "Bajo" 9.5-10.2 "Regular" 10.3-10.6 "Bueno" 10.7-11 "Muy bueno" >11 "Excelente"	Cuantitativa
Desempeño Inglés Cod. MOD_INGLÉS_DESEM CAPR_DESEMPEÑO_INGL	Toma cinco niveles de valoración. A-, A1, A2; B+, B1; B2. Como A- el más bajo y B2 el más alto	Cualitativa ordinal
Puntaje Lectura Crítica Cod. MOD_LECTURA_CRITICA	Este resultado se presenta en una escala con media 10 y desviación estándar 1. Está representado en sixtiles >9 "Muy bajo" 9-9.5 "Bajo" 9.5-10.2 "Regular" 10.3-10.6 "Bueno" 10.7-11 "Muy bueno" >11 "Excelente"	Cuantitativa
Puntaje Razonamiento Cuantitativo Cod. MOD_RAZONAMIENTO_CUANTITATIVO	Este resultado se presenta en una escala con media 10 y desviación estándar 1. Está representado en sixtiles. >9 "Muy bajo" 9-9.5 "Bajo" 9.5-10.2 "Regular" 10.3-10.6 "Bueno" 10.7-11 "Muy bueno" >11 "Excelente"	Cuantitativa

3.2 MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS

Para este trabajo de investigación se utilizó técnicas de tipo descriptivo, prueba de diferencia de medias y medianas, pruebas de correlación, pruebas de independencia, pruebas de bondad de ajuste, técnicas de modelamiento, tales, como. `gamlss`, modelo logístico ordinal. Para el procesamiento de datos se utilizó el Excel, el software de R versión 3.1.1 con sus paquetes. `gamss`, `gamlss.dist`, `gamlss.data`, `MASS`, `nlme`, `parallel`, `splines` y `R commander`.

4. DESCRIPCIÓN DE LOS DATOS

A continuación se describe el desempeño de los estudiantes del Programa de Administración de Empresas Agropecuarias de la Uptc Seccional Duitama en las Pruebas Saber Pro. En primer lugar, se describen los resultados, entre el segundo semestre de 2011 hasta el segundo semestre de 2015. La información tomada para desarrollar las secciones 4.1 y 4.2 fue tomada de la página <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados>. La información que sustenta la sección 4.3 se extrajo de las bases de datos que proporciona el ICFES para desarrollar investigaciones.

4.1 DESCRIPCIÓN UNIVARIADA POR COMPONENTE

Las competencias genéricas son las que se consideran que debe desarrollar todo estudiante, independiente de su formación profesional, se evalúa Inglés, Razonamiento Cuantitativo, Lectura Crítica, desempeño de Comunicación Escrita y puntaje en Comunicación Escrita. Los resultados de cada módulo correspondiente al Programa de Administración de Empresas Agropecuarias de la UPTC Seccional Duitama se presentan en una tabla. La cual incluye de los resultados correspondientes al Programa, n el número de estudiantes que presentaron la prueba, \bar{x} , puntaje promedio por año, Sd desviación estándar CA , coeficiente de asimetría, CC , coeficiente de kurtosis, CV , coeficiente de variación, Q_1 , Q_2 y Q_3 representan los cuantiles uno, dos y tres respectivamente y NA se refiere al número de datos faltantes.

A continuación, se presenta una descripción de los resultados del módulo de Inglés.

Tabla 6. Resultados módulo Inglés

<i>Año</i>	<i>n</i>	\bar{x}	<i>Sd</i>	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	<i>NA</i>
2011	39	9.7	0.64	1.33	3.33	0.07	9.4	9.5	10.0	0
2012	38	9.6	0.51	0.13	0.50	0.05	9.2	9.6	9.9	0
2013	70	9.7	0.55	0.54	1.13	0.06	9.4	9.6	10.0	0
2014	42	9.7	0.59	2.27	8.40	0.06	9.2	9.6	9.9	0
2015	33	9.9	0.64	0.90	1.74	0.06	9.5	9.9	10.3	0

En la Tabla 6 se presentan los resultados obtenidos en **INGLÉS**, se observa que, para cada uno de los años que han presentado la prueba, los resultados de los estudiantes de Administración de Empresas Agropecuarias son muy homogéneos respecto a los puntajes (la variabilidad relativa, es decir los coeficientes de variación, son inferiores al 7%). Los coeficientes de kurtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba entre 2011 y 2015, excepto del 2014 donde la kurtosis es 8.40, la cual indica que los resultados de los estudiantes en este año son heterogéneos a causa de que se encuentra un puntaje de 12.4. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 9.7 y la desviación estándar es de 0.59 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 8.52 y 10.88 puntos.

Los resultados anteriores se pueden complementar con los resultados cualitativos en Inglés, tabla 7, de la cual se afirma que, sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes (43.2%) se ubican en el nivel de desempeño A1, indicando que son capaces de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. De otro lado se tiene que el

18% de los estudiantes se ubican en el nivel A2 y el 2% en el nivel más alto de desempeño, B1.

Tabla 7. Resultados de Desempeño en Inglés

AÑO	NIVELES						N	NA
	A-	A1	A2	B1	B+			
2011	33%	44%	18%	5%	0%	39	0	
2012	39%	47%	11%	3%	0%	38	0	
2013	36%	40%	17%	7%	0%	70	0	
2014	36%	45%	15%	2%	2%	42	0	
2015	21%	40%	30%	9%	0%	33	0	

Los resultados de **RAZONAMIENTO CUANTITATIVO** se presentan en la tabla 8. Se observa que, para cada uno de los años que han presentado la prueba, los resultados de los estudiantes de Administración de Empresas Agropecuarias son muy homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 9%). Los coeficientes de kurtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 9.7 y la desviación estándar es de 0.70 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 8.3 y 11.1 puntos.

Tabla 8. Resultados Razonamiento Cuantitativo

Año	<i>n</i>	\bar{x}	<i>Sd</i>	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	NA
2011	39	9.74	0.89	0.17	0.58	0.09	9.15	9.7	10.2	0
2012	38	9.7	0.78	0.76	0.87	0.08	9.1	9.6	10.2	0
2013	70	9.7	0.55	0.74	2.23	0.06	9.3	9.8	10.0	0
2014	42	9.5	0.56	0.18	-0.65	0.06	9.2	9.4	10.0	0
2015	33	10	0.75	0.50	-0.49	0.07	9.6	9.9	10.8	1

Los resultados del **PUNTAJE EN COMUNICACIÓN ESCRITA** se presentan en la tabla 9. Se observa que, para cada uno de los años que han presentado la prueba, los resultados de los estudiantes de Administración de Empresas Agropecuarias son homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 10%). Los coeficientes de kurtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 10.2 y la desviación estándar es de 0.84 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 8.52 y 11.88 puntos.

Tabla 9. Resultados del puntaje en Comunicación Escrita

Año	<i>n</i>	\bar{x}	<i>Sd</i>	<i>CA</i>	<i>CC</i>	<i>CV</i>	<i>Q</i> ₁	<i>Q</i> ₂	<i>Q</i> ₃	<i>NA</i>
2011	39	10.2	0.92	0.25	-0.11	0.09	9.7	10.2	10.9	1
2012	38	10.3	0.83	0.20	0.71	0.08	9.8	10.1	10.7	0
2013	70	10.3	0.67	-0.47	1.55	0.06	9.9	10.4	10.6	0
2014	42	10.2	0.98	-0.04	-0.55	0.10	9.4	10.3	10.7	1
2015	33	10.2	0.82	0.55	0.97	0.08	9.5	10.2	10.5	0

Los resultados anteriores se pueden complementar con los resultados cualitativos en comunicación escrita, tabla 10. Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de los estudiantes (39.4%) se ubican en el nivel 4 de desempeño, indicando que el estudiante es capaz, en cualquier texto de encontrar una idea central que se desarrolla de acuerdo con una intención comunicativa, le da una estructura básica, en otras palabras, es posible que pueda dar una introducción al tema que se abordará, un desarrollo y una conclusión. Sin embargo, el texto no incluye toda la información necesaria (progresión temática), su organización no es completamente efectiva, o rompe la unidad al incluir temas que no se relacionan con el marco semántico que desarrolla. Se aprecia un uso aceptable del lenguaje (se aplican las reglas gramaticales más importantes). De otro lado se tiene que el 33.6% de los estudiantes se ubican en el nivel 5 y muy

pocos en el nivel 7 con un porcentaje de 2.6%. No se tienen estudiantes en el nivel 8.

Tabla 10. Resultados del Desempeño de Comunicación Escrita

AÑO	NIVELES								N	NA
	1	2	3	4	5	6	7	8		
2011	0%	0%	15%	38%	26%	13%	5%	0%	39	1
2012	0%	3%	8%	47%	23%	16%	3%	0%	38	0
2013	0%	1%	3%	42%	44%	10%	0%	0%	70	0
2014	0%	0%	17%	24%	36%	19%	2%	0%	42	1
2015	0%	0%	6%	46%	39%	6%	3%	0%	33	0

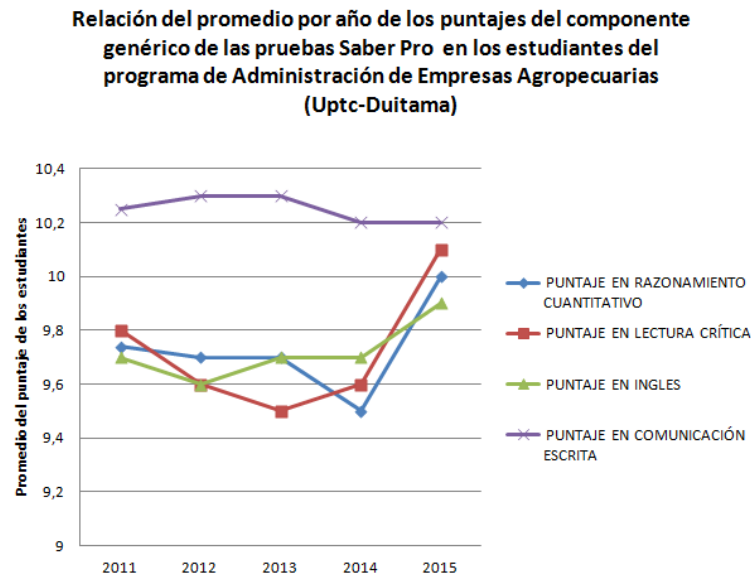
Los resultados de **LECTURA CRÍTICA** se presentan en la tabla 11. Se observa que, para cada uno de los años que han presentado la prueba, los resultados de los estudiantes de Administración de Empresas Agropecuarias son muy homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 10%). Los coeficientes de asimetría indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 9.7 y la desviación estándar es de 0.79 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 8,12 y 11.28 puntos.

Tabla 11. Resultados en Lectura Crítica

Año	n	\bar{x}	Sd	CA	CC	CV	Q ₁	Q ₂	Q ₃	NA
2011	39	9.8	0.96	1.33	2.62	0.1	9.3	9.6	10.2	0
2012	38	9.6	0.78	-0.21	0.002	0.08	9.2	9.6	10.2	0
2013	70	9.5	0.85	-0.59	0.45	0.09	9.0	9.5	10.1	0
2014	42	9.6	0.73	-0.45	0.55	0.07	9.0	9.7	10.1	0
2015	33	10.1	0.62	0.01	-0.38	0.06	9.8	10.2	10.5	0

A continuación se presenta una gráfica comparativa de los puntajes del componente genérico.

Figura 3. Comparación de los puntajes del componente genérico.



En la anterior gráfica se puede observar que en Comunicación Escrita fue donde se obtuvo mejor rendimiento, ya que en relación con los puntajes promedio por año estos son más altos que en Inglés, Razonamiento Cuantitativo y Lectura Crítica.

4.2 RESULTADOS EN COMPETENCIAS GENÉRICAS DE LA PRUEBA SABER PRO AÑO 2016 DEL PROGRAMA DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS

Para el año 2016, los resultados de los puntajes del componente genérico en las pruebas Saber Pro de los estudiantes del programa de Administración de Empresas Agropecuarias de la Uptc Seccional Duitama, fueron.

Resultados en competencias genéricas

Los resultados del puntaje en Comunicación Escrita se presentan en la tabla 12. Se observa que, a nivel del programa el puntaje promedio fue de 144 puntos con

una desviación estándar de 28.27 y un coeficiente de variación del 20%, lo cual indica que los resultados de los estudiantes de Administración de Empresas Agropecuarias son homogéneos. Teniendo en cuenta que el promedio es de 144 y la desviación estándar de 28.27 puntos, al menos el 75% de los estudiantes obtuvieron resultados entre 87.46 y 200.54 puntos.

Con los resultados del puntaje en Inglés se afirma que, a nivel del programa el puntaje promedio fue de 140 puntos con una desviación estándar de 22.97 y un coeficiente de variación del 16%, lo cual indica que los resultados de los estudiantes de Administración de Empresas Agropecuarias son homogéneos. Teniendo en cuenta que el promedio es de 140 y la desviación estándar de 22.9 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 94.2 y 185.8 puntos.

Según los resultados del puntaje en Lectura Crítica mostrados en la tabla 12, se observa que, a nivel del programa el puntaje promedio fue de 147 puntos con una desviación estándar de 21.3 y un coeficiente de variación del 19%, lo cual indica que los resultados de los estudiantes de Administración de Empresas Agropecuarias son homogéneos. Teniendo en cuenta que el promedio es de 143 y la desviación estándar de 27.01 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 88.98 y 197.02 puntos.

Según los resultados del puntaje en Razonamiento Cuantitativo descritos en la tabla 12, se observa que, a nivel del programa el puntaje promedio fue de 143 puntos con una desviación estándar de 27.01 y un coeficiente de variación del 14%, lo cual indica que los resultados de los estudiantes de Administración de Empresas Agropecuarias son homogéneos. Teniendo en cuenta que el promedio es de 147 y la desviación estándar de 21.3 puntos, aproximadamente el 75% de los estudiantes obtuvieron resultados entre 104.4 y 189.6 puntos.

Tabla 12. Puntajes 2016

Año 2016	n	\bar{x}	Sd	CA	CC	CV	Q ₁	Q ₂	Q ₃	NA
<i>Puntaje Comunicación Escrita</i>	33	144	28.27	-0.36	-0.21	0.2	126	157	162	0
<i>Puntaje Inglés</i>	33	140	22.97	0.88	0.91	0.16	123	141	152	0
<i>Puntaje lectura crítica</i>	33	143	27.01	0.80	0.83	0.19	126	137	156	0
<i>Puntaje Razonamiento Cuantitativo</i>	33	147	21.3	0.99	3.25	0.14	129	146	157	0

La anterior tabla muestra que, en Razonamiento Cuantitativo se tuvo mejor rendimiento que Inglés, Comunicación Escrita y Lectura Crítica. También se puede concluir que en este año se mejoró el rendimiento del puntaje en Razonamiento Cuantitativo en relación con los años anteriores (2011-2015).

En desempeño de Comunicación Escrita la mayoría de los estudiantes del Programa (43%) se ubica en nivel 3 de desempeño, ubicado en el rango de puntajes desde 156 a 190 puntos, es decir, los estudiantes tan sólo son capaces en un texto de darle una estructura básica, en la que es posible identificar una introducción, un desarrollo, un propósito y un cierre.

Tabla 13. Desempeño de Comunicación Escrita 2016

AÑO 2016	NIVELES								n	NA
	1	2	3	4	5	6	7	8		
Desempeño Comunicación Escrita	12%	36%	43%	9%	0%	0%	0%	0%	39	1

En desempeño de Inglés la mayoría de los estudiantes del Programa (46%) se ubica en nivel de desempeño A1, es decir, los estudiantes tan sólo son capaces de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. Puede presentarse él mismo y ante otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce. Puede relacionarse de

forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar.

Tabla 14. Desempeño en Inglés 2016

	NIVELES							
AÑO 2016	A-	A1	A2	B1	B+	B2	N	NA
<i>Desempeño Inglés</i>	0%	46 %	21%	6%	0%	3%	33	8

4.3 DESCRIPCIÓN UNIVARIADA DE ALGUNAS CARÁCTERÍSTICAS DE LOS ESTUDIANTES DE ADMINISTRACIÓN DE EMPRESAS AGROPECUARIAS

En la tabla 15 se presenta un resumen descriptivo respecto a las características demográficas de los estudiantes que se tuvieron en cuenta para esta investigación y su respectiva clasificación.

Tabla 15. Descripción factores demográficos.

FACTORES DEMOGRÁFICOS		
VARIABLE	UNIDADES DE MEDIDA/CATEGORÍAS	RESUMEN NUMÉRICO
Género	Femenino	F. 63%
	Masculino	M.37%
Edad	Edad en años	Promedio=25
		Coficiente de variación=0.12
		Desviación estándar=3.1
Estado civil	Casado	C. 3%
	Soltero	SLT. 89%
	Separado	SP. 1%
	Unión libre	UL. 6%
	NA's	NA's. 1%

Continuación tabla anterior

VARIABLE	UNIDADES DE MEDIDA/CATEGORÍAS	RESUMEN NUMÉRICO
Número de personas en el hogar	Número de personas que conforman el hogar	Promedio=4
		Coeficiente de variación=0.40
		Desviación estándar=1.73
Es cabeza de familia	SI	12%
	NO	87%
	NA's	1%
Número de personas a cargo	Número de personas de las que el estudiante se encuentra a cargo	Promedio=0.2941
		Coeficiente de variación= 2.36
		Desviación estándar=1.73
Situación del hogar	Es habitual o permanente	H. 73%
	Temporal	T. 26%
	NA's	NA's. 1%

De acuerdo con los resultados estadísticos de la tabla anterior, los estudiantes del programa de Administración de Empresas Agropecuarias que presentaron la prueba Saber Pro desde el segundo semestre de 2011 hasta el segundo semestre de 2015 se identifican por tener una edad promedio de 25 años, el 63% son mujeres y el 37% son hombres, 87% son solteros, además la gran mayoría no tienen personas a cargo y tan solo el 12% son cabeza de hogar.

En Tabla 16, se presenta un resumen descriptivo respecto a las características socioeconómicas de los 222 estudiantes del programa de Administración de Empresas Agropecuarias.

Tabla 17. Descripción factores socioeconómicos.

FACTORES SOCIOECONÓMICOS		
VARIABLES	DESCRIPCIÓN	TIPO DE VARIABLE
Forma de pago de la matrícula	Provinieron de beca	B. 2%
	Ayuda de los padres	PD. 10%
	Recursos propios	PP. 30%
	Varias formas de pago	VFP. 55%
	NA's	NA's. 4%
Valor de la matrícula del año anterior	No pagó matrícula	4%
	Menos de 500 mil	70%
	Entre 500 mil y Menos de 1 millón	22%
	Entre 1 millón y Menos de 3 millón	3%
	NA's	1%
Estrato	1	12%
	2	71%
	3	14%
	4	2%
	NA's	1%
El hogar cuenta con celular	SI	90%
	NO	9%
	NA's	1%

De los 222 estudiantes del programa de Administración de Empresas Agropecuarias que presentaron la prueba Saber Pro, el 70% tienen que pagar una matrícula de menos de 500 mil pesos, el 55% tienen varias formas de pago, dentro de estas incluyendo la ayuda de los padres, crédito, recursos propios y demás.

Continuación tabla anterior

FACTORES SOCIOECONÓMICOS		
VARIABLES	DESCRIPCIÓN	TIPO DE VARIABLE
El hogar cuenta con conexión a internet	SI	29%
	NO	70%
	NA's	1%
El hogar cuenta con Servicio cerrado de televisión	SI	46%
	NO	53%
	NA's	1%
El hogar cuenta con servicio de teléfono fijo.	SI	14%
	NO	85%
	NA's	1%
Ocupación del padre	Empleado	E. 38%
	Trabajador independiente	TI. 38%
	Hogar	H. 2%
	Pensionado	P. 21%
	NA's	NA's. 1%
Ocupación de la madre	Empleada	E. 17%
	Trabajadora independiente	TI. 12%
	Hogar	H. 65%
	Pensionado	P. 5%
	NA's	NA's. 1%
El hogar cuenta con computador	SI	24%
	NO	75%
	NA's	1%

Para aquellos estudiantes que presentaron la prueba Saber Pro entre 2011 y 2015 la mayoría de los padres son empleados y trabajadores independientes, además mayoritariamente (65%) de las madres son amas de casa, gran parte del grupo de estudiantes no cuenta en su hogar con el servicio de internet, teléfono fijo y computador.

Continuación factores socioeconómicos

FACTORES SOCIOECONÓMICOS		
VARIABLES	DESCRIPCIÓN	TIPO DE VARIABLE
Ingresos mensuales familiares	Menos de un (1) SM	27%
	Entre un (1) y menos de dos (2) SM	56%
	Entre dos (2) y menos de tres (3) SM	11%
	Entre tres (3) y menos de siete (7) SM.	5%
	NA's	1%
Situación actual del trabajo	No	No. 44%
	Si, como ayudante sin remuneración	ASR. 43%
	Si, por ser práctica obligatoria del plan de estudios	PO. 1%
	Si, para adquirir experiencia y/o recursos para sus gastos personales	EGP.11%
	NA's	NA's . 1%

De los 222 estudiantes que presentaron la prueba el 27% tienen ingresos familiares mensuales de menos uno (1) SM y el 56% tienen ingresos familiares mensuales entre uno (1) SM y menos de dos (2) SM, además el 87% de los estudiantes trabajan sin remuneración o no trabajan.

En la tabla 17, se presenta un resumen descriptivo respecto a las características académicas de los 222 estudiantes del programa de Administración de Empresas Agropecuarias que presentaron la prueba Saber Pro entre el segundo periodo de 2011 hasta el segundo periodo de 2015.

Tabla 18. Descripción factores académicas.

FACTORES ACADÉMICOS		
VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
El estudiante tomó algún curso de preparación.	SI	1%
	NO	73%
	NA's	26%
Tipo de bachillerato	Académico	A. 51%
	Normalista Superior	NS. 2%
	Técnico	T. 46%
	NA's	NA's. 1%
Nivel educativo alcanzado Por el padre	Ninguno	N. 4%
	Básica primaria	BP. 49%
	Básica secundaria	BS. 28%
	Educación superior	ES. 18%
	NA's	NA's . 1%
Nivel educativo alcanzado por la madre	Ninguno	N. 2%
	Básica primaria	BP. 46%
	Básica secundaria	BS. 38%
	Educación superior	ES. 13%
	NA's	NA's . 1%

Del anterior resumen descriptivo se puede afirmar que la mayoría de los estudiantes del programa de Administración de Empresas Agropecuarias que presentaron las pruebas Saber Pro entre el segundo semestre de 2011 hasta el segundo semestre de 2015 no tomaron ningún curso de preparación para esta prueba (73%). Además, también se puede afirmar que el nivel educativo alcanzado por el padre y la madre del estudiante se encuentra mayoritariamente en básica primaria con un 49% y un 46%.

4.4 DESCRIPCIÓN BIVARIADA DE LAS CARACTERÍSTICAS DE LOS ESTUDIANTES

A continuación se presentan las conclusiones al relacionar los puntajes del componente genérico y algunas características que describen el entorno académico, demográfico, socioeconómico de los estudiantes del programa de Administración de Empresas Agropecuarias que presentaron las pruebas Saber Pro durante el segundo periodo de 2011 hasta el 2015. Los soportes de los resultados presentados se muestran en el Anexo 1 “Resultados estadísticos análisis bivariado”.

Para esta parte de la investigación una de las técnicas que se utilizó fue la prueba de correlación de Spearman, la cual mide la correlación (la asociación o interdependencia) entre dos variables continuas⁵⁵.

Los resultados del coeficiente de correlación de Spearman obtenidos a través de R, muestran si existe o no correlación entre los puntajes del componente genérico y algunas variables que identifican al estudiante. A un nivel de significancia del 5% se afirma que, la edad del estudiante, el número de personas de las que se encuentra a cargo el estudiante, el número de personas que conforman el hogar, el número de dormitorios de la residencia, tienen una correlación negativa con el puntaje en Razonamiento Cuantitativo. Esto quiere decir, a valores altos de una de ellas le suelen corresponder valores bajos de la otra o viceversa.

De la misma manera se puede afirmar que, el número de personas que conforman el hogar y el número de dormitorios de la residencia tienen una correlación positiva con el puntaje en Inglés. Esto quiere decir, que las dos variables se correlacionan en sentido directo. A valores altos de una de ellas le suelen corresponder valores

⁵⁵ CAMACHO, J. (2008). Asociación entre variables cuantitativas: análisis de correlación. Acta Médica Costarricense 2008; 50: 94-96. ISSN 0001-6002/2008/50/3/144-146.

altos de la otra e igualmente que con los valores bajos. Además, también se encontró que, solo el número de dormitorios de la residencia tienen una correlación positiva con el puntaje en Lectura Crítica y con el puntaje en Comunicación Escrita, las demás variables tienen una correlación negativa con estos puntajes.

Se encontró que, con un nivel de significancia del 5%, el desempeño en Inglés está asociado a la ocupación de la madre, a la ocupación del padre del estudiante, al nivel educativo alcanzado por la madre y al valor anual de la matrícula del año anterior del estudiante. Además, se encontró que con un nivel de significancia del 5% el desempeño Comunicación Escrita está asociado a la ocupación del padre del estudiante y de sí o no trabaja actualmente.

Se llevó a cabo la prueba de normalidad de Shapiro Wilk, según estos resultados, el puntaje en Comunicación Escrita es la única variable de respuesta que sigue una distribución normal, con un nivel de significancia del 5%.

Como ninguno de los puntajes a excepción del puntaje en Comunicación Escrita, siguen una distribución normal se realizó el test no paramétrico de Kruskal Wallis, con el fin de ver si dos variables (una independiente y otra dependiente) están relacionadas o tienen alguna asociación, con base a si las medianas de la variable dependiente (puntajes componente genérico) son diferentes o iguales en las categorías o grupos de la variable independiente (variable cualitativa del estudiante). Es decir, señala si las medianas entre dos o más grupos son similares o diferentes.

En la tabla 18, se muestra un resumen del test no paramétrico de Kruskal Wallis respecto al puntaje en Razonamiento Cuantitativo vs cada variable cualitativa.

Tabla 19. Test no paramétrico de Kruskal Wallis. Puntaje en Razonamiento Cuantitativo

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
El hogar cuenta con celular	No	9.2	2.5379	1	0.1111
	Si	9.7			
Tiene computador en su hogar	No	9.7	0.1434	1	0.7049
	Si	9.7			
El hogar cuenta con conexión a internet	No	9.7	0.1718	1	0.6785
	Si	9.8			
Situación del hogar actual	Habitual	9.7	0.0061	1	0.9376
	Temporal	9.6			
Es cabeza de familia	Si	9.65	0.0005	1	0.983
	No	9.70			
Género	Femenino	9.8	4.3141	1	0.0378
	Masculino	9.7			
Titulo	Académico	9.65	0.6527	1	0.7215
	Técnico	9.70			
	Normalista	9.80			
Trabaja actualmente	Si, como ayudante sin remuneración	9.7	1.5514	1	0.6705
	Si, para adquirir experiencia y/o recursos para sus gastos personales	9.9			
	Si, por ser práctica obligatoria del plan de estudios	9.7			
El hogar cuenta con lavadora	Si	9.7	0.433	1	0.5105
	No	9.7			
El hogar cuenta con Servicio cerrado de televisión	Si	9.75	1.1214	1	0.2896
	No	9.70			
El hogar cuenta con servicio de teléfono fijo	No	9.7	0.0887	1	0.7658
	Si	9.8			
Estado civil	Soltero	9.7	3.0014	1	0.3914
	Separado	9.8			
	Unión libre	9.6			
	Casado	9.2			

De acuerdo al resumen del test no paramétrico de Kruskal Wallis, se encontró que con cualquier nivel de significancia considerado, el valor de la mediana de los puntajes de Razonamiento Cuantitativo de aquellos estudiantes que tienen computador en su hogar respecto al valor de la mediana de los puntajes del mismo componente para aquellos estudiantes que no cuentan con este medio, son iguales, lo cual quiere decir que el puntaje en Razonamiento Cuantitativo no asocia con el tener o no computador en el hogar.

Los valores de la mediana de los puntajes de Razonamiento Cuantitativo correspondientes a los estudiantes solteros, casados, que viven en unión libre, y que son separados son iguales, lo cual explica que el estado civil del estudiante no se asocia con el puntaje en Razonamiento Cuantitativo.

El valor de la mediana de los puntajes de Razonamiento Cuantitativo de los hombres y el valor de la mediana de los puntajes de este mismo componente correspondiente a las mujeres son diferentes, lo cual que indica que el género se asocia con los puntajes de Razonamiento Cuantitativo.

Continuación test no paramétrico de Kruskal Wallis

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
Ocupación de la madre	Trabajadora independiente	9.7	6.1148	1	0.1062
	Hogar	9.6			
	Pensionada	9.4			
	Empleada	9.9			
Ocupación del padre	Trabajador independiente	9.70	1.6247	1	0.6538
	Hogar	9.70			
	Pensionado	9.75			
	Empleado	9.65			
Forma de pago de la matrícula	Beca	10.5	1.755	1	0.6247
	Padres	9.7			
	Recursos propios	9.7			
	Varias formas de pago	9.7			

Según los resultados anteriores, los valores de la mediana de los puntajes de Razonamiento Cuantitativo correspondientes a los estudiantes cuyos padres son empleados, pensionados, trabajadores independientes y aquellos que se dedican al hogar son similares, lo cual significa que la ocupación del padre no está asociada con el puntaje en Razonamiento Cuantitativo y éste tampoco se asocia con la ocupación de la madre.

A continuación se muestra un resumen del test no paramétrico de Kruskal Wallis respecto al puntaje en Inglés vs cada variable cualitativa.

Tabla 20. Test no paramétrico de Kruskal Wallis. Puntaje en Inglés

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
El hogar cuenta con celular	No	9.8	0.4305	1	0.5118
	Si	9.6			
Tiene computador en su hogar	No	9.6	0.9971	1	0.318
	Si	9.6			
El hogar cuenta con conexión a internet	No	9.6	0.3638	1	0.5464
	Si	9.6			
Situación del hogar actual	Habitual	9.6	0.78	1	0.3771
	Temporal	9.6			
Es cabeza de familia	Si	9.5	2.5626	1	0.1323
	No	9.60			
Género	F	9.6	0.2435	1	0.6217
	M	9,7			

De acuerdo al resumen del test no paramétrico de Kruskal Wallis, se encontró que el valor de la mediana de los puntajes de Inglés correspondiente a los hombres y el valor de la mediana de los puntajes de este mismo componente correspondiente a las mujeres son similares, lo cual que indica que el género no se asocia con los puntajes de Inglés. Además el valor de la mediana de los puntajes de Inglés correspondientes a los estudiantes que son cabeza de familia y el valor de la mediana de los estudiantes que no son cabeza de familia son similares, lo cual indica que el ser o no ser cabeza de familia tampoco está asociado con el puntaje en Inglés.

Continuación test no paramétrico de Kruskal Wallis

VARIABLES	RANGO	MEDIAN A	CHI-CUADRADO	GL	P-VALOR
Titulo	Académico	9.6	0.7945	1	0.6722
	Técnico	9.6			
	Normalista	9.73			
Tomó algún curso de preparación para el examen	No	9.61	2.4123	1	0.1204
	Si	9.03			
Trabaja actualmente	No	9.7	8.2005	1	0.04204
	Si, como ayudante sin remuneración	9.5			
	Si, para adquirir experiencia y/o recursos para sus gastos personales	9.6			
	Si, por ser práctica obligatoria del plan de estudios	9.1			
El hogar cuenta con lavadora	Si	9.6	4.2931	1	0.03813
	No	9.7			
El hogar cuenta con Servicio cerrado de televisión	Si	9.6	0.7362	1	0.3909
	No	9.6			
El hogar cuenta con servicio de teléfono fijo	No	9.60	0.3939	1	0.5302
	Si	9.65			
Ocupación de la madre	Trabajadora independiente	9.7	2.9797	1	0.3948
	Hogar	9.80			
	Pensionada	9.4			
	Empleada	9.67			
Máximo nivel educativo alcanzado por el padre	Ninguno	9.8	7.0018	1	0.07184
	Primaria	9.6			
	Secundaria	9.6			
	Educación superior	9.8			
Forma de pago de la matricula	Beca	10.1	1.2894	1	0.7317
	Padres	9.5			
	Recursos propios	9.7			
	Varias formas de pago	9.6			

Con un nivel de significancia del 5%, según el resumen del test no paramétrico de Kruskal Wallis, se encontró que el valor la mediana de los puntajes de Inglés correspondiente a los estudiantes que cuentan en su hogar con lavadora y el valor

de la mediana de los puntajes de este mismo componente correspondiente a los estudiantes que no cuentan con este servicio son diferentes, lo cual que indica que el tener o no tener lavadora en el hogar se asocia con los puntajes de Inglés. Además, también se encontró que los valores de la mediana de los puntajes de Inglés correspondientes a los estudiantes que trabajan y a los que no trabajan son similares, lo explica que el trabajo no está asociado con los puntajes de Inglés.

En la tabla 20, se muestra un resumen del test no paramétrico de Kruskal Wallis respecto al puntaje en Lectura Crítica vs cada variable cualitativa.

Tabla 21. Test no paramétrico de Kruskal Wallis. Puntaje en Lectura Crítica

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
El hogar cuenta con celular	No	9.7	0.0545	1	0.8155
	Si	9.7			
Tiene computador en su hogar	No	9.65	0.2986	1	0.5848
	Si	9.7			
El hogar cuenta con conexión a internet	No	9.70	1.6362	1	0.2009
	Si	9.85			
Situación del hogar actual	Habitual	9.7	0.0003	1	0.9856
	Temporal	9.8			
Es cabeza de familia	Si	9.65	0.404	1	0.525
	No	9.7			
Género	F	9.8	0.9191	1	0.3377
	M	9.7			
Titulo	Académico	9.7	5.0408	1	0.08043
	Técnico	9.8			
	Normalista	8.9			
Tomó algún curso de preparación para el examen	No	9.7	0.0139	1	0.9063
	Si	9.8			

Con un nivel de significancia del 5%, según el resumen del test no paramétrico de Kruskal Wallis, se encontró que el tener o no tener acceso a internet en el hogar no se relaciona con los puntajes de Lectura Crítica. Además se encontró que el tomar o no curso de preparación no está asociado con los puntajes de Lectura Crítica.

Continuación test no paramétrico de Kruskal Wallis

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
Trabaja actualmente	No	9.65	5.3864	1	0.1456
	Si, como ayudante sin remuneración	9.8			
	Si, para adquirir experiencia y/o recursos para sus gastos personales	9.70			
	Si, por ser práctica obligatoria del plan de estudios	8.30			
El hogar cuenta con lavadora	Si	9.8	0.6382	1	0.4244
	No	9.6			
El hogar cuenta con Servicio cerrado de televisión	Si	9.7	0.1452	1	0.7032
	No	9.7			
El hogar cuenta con servicio de teléfono fijo	No	9.70	0.5776	1	0.4472
	Si	9.75			

Con un nivel de significancia del 5%, según resumen del test no paramétrico de Kruskal Wallis, se encontró que el tener servicio cerrado de televisión en el hogar no se asocia con los puntajes de Lectura Crítica.

Continuación test no paramétrico de Kruskal Wallis

VARIABLES	RANGO	MEDIANA	CHI-CUADRADO	GL	P-VALOR
Ocupación de la madre	Trabajadora independiente	9.5	1.4127	1	0.7026
	Hogar	9.70			
	Pensionada	9.30			
	Empleada	9.8			
Ocupación del padre	Trabajador independiente	9.45	3.19	1	0.3632
	Hogar	9.9			
	Pensionado	9.95			
	Empleado	9.80			
Forma de pago de la matrícula	Beca	10.2	2.4526	1	0.4839
	Padres	9.6			
	Recursos propios	9.7			
	Varias formas de pago	9.7			

Según los resultados anteriores, con un nivel de significancia del 5%, se encontró que los valores de la mediana de los puntajes de Lectura Crítica correspondientes a los estudiantes cuyas madres son empleadas, pensionadas, trabajadoras independientes y amas de casa son similares, lo cual significa que la ocupación de la madre no está asociado con el puntaje de Lectura Crítica y este tampoco se asocia con la ocupación del padre ya que el p-valor es mayor que cualquier nivel de significancia.

Como el puntaje en Comunicación Escrita, siguen una distribución normal se realizó el análisis de la varianza (ANOVA) de un factor, con el fin de ver si dos variables están relacionadas o tienen alguna asociación, en base a si las medias de la variable dependiente (puntaje en Comunicación Escrita) son diferentes o iguales en las categorías o grupos de la variable independiente (variable cualitativa del estudiante). Es decir, señala si las medias entre dos o más grupos son similares o diferentes.

En la tabla 21, se muestra un resumen del análisis de la varianza (ANOVA) de un factor respecto al puntaje en Comunicación Escrita vs cada variable cualitativa.

Análisis de la varianza (ANOVA)

Tabla 22. Análisis de la varianza (ANOVA). Puntaje en Comunicación Escrita

VARIABLES	RANGO	MEDIA	VALOR F	P-VALOR
El hogar cuenta con celular	No	10.23	0.003	0.955
	Si	10.24		
Tiene computador en su hogar	No	10.15	0.874	0.351
	Si	10.17		
El hogar cuenta con conexión a internet	No	10.21	0.487	0.486
	Si	10.30		
Situación del hogar actual	Habitual	10.18	3.083	0.0805
	Temporal	10.40		
Es cabeza de familia	Si	10.07	1.254	0.264
	No	10.26		
Género	F	10.27	0.792	0.374
	M	10.17		

De acuerdo con el resumen anterior del análisis de la varianza (ANOVA), se encontró el valor de la media de los puntajes de Comunicación Escrita correspondiente a los estudiantes que cuentan con celular y el valor de la media de los puntajes de este mismo componente correspondiente a los estudiantes que no cuentan con este servicio son similares, lo cual que indica que el tener o no tener celular no se relaciona con los puntajes de Comunicación Escrita. Además también se encontró que los valores de la media de los puntajes de Comunicación Escrita correspondientes a los hombres y mujeres son similares, lo cual explica que el género no se asocia con el puntaje en Comunicación Escrita.

Continuación del análisis de la varianza (ANOVA)

VARIABLES	RANGO	MEDIA	VALOR F	P-VALOR
Titulo	Académico	10.25	0.323	0.723
	Técnico	10.23		
	Normalista	9.86		
Tomó algún curso de preparación para el examen	No	10.20	1.23	0.269
	Si	11.10		
Trabaja actualmente	No	10.19	2.01	0.113
	Si, como ayudante sin remuneración	10.25		
	Si, para adquirir experiencia y/o recursos para sus gastos personales	10.45		
	Si, por ser práctica obligatoria del plan de estudios	8.6		
El hogar cuenta con lavadora	Si	10.24	0.0002	0.969
	No	10.24		
El hogar cuenta con Servicio cerrado de televisión	Si	10.18	0.952	0.33
	No	10.29		
El hogar cuenta con servicio de teléfono fijo	No	10.23	0.094	0.759
	Si	10.28		
Ocupación de la madre	Trabajadora independiente	9.7	0.022	0.995
	Hogar	9.80		
	Pensionada	9.4		
	Empleada	9.67		
Estado civil	Casado	10.20	0.705	0.55
	Soltero	10.22		
	Separado	9.90		
	Unión libre	10.54		

Continuación del análisis de la varianza (ANOVA)

VARIABLES	RANGO	MEDIA	VALOR F	P-VALOR
Máximo nivel educativo alcanzado por el padre	Ninguno	10.21	0.01	0.999
	Primaria	10.24		
	Secundaria	10.25		
	Educación superior	10.23		
Forma de pago de la matrícula	Beca	10.52	0.792	0.429
	Padres	10.12		
	Recursos propios	10.33		
	Varias formas de pago	10.18		

Según los resultados del análisis de la varianza (ANOVA) anteriores, se encontró que valores de la media de los puntajes de Comunicación Escrita de los estudiantes que tomaron algún curso de preparación para el examen y los que no tomaron algún curso de preparación para el examen son similares, lo cual explica que el tomar o no tomar curso de preparación para el examen Saber Pro no se asocia con puntaje en Comunicación Escrita. Se encontró también que los valores de la media de los puntajes de Comunicación Escrita correspondientes a los estudiantes cuyas madres alcanzaron un nivel de educación superior, secundario, primario y que no realizaron ningún nivel educativo son similares, lo cual significa que la educación de la madre no está asociada con el puntaje de Comunicación. Además también se encontró que los valores de las medias de los estudiantes con título del bachillerato académico, técnico, y normalista son similares, lo cual explica que el tipo de bachillerato del que se graduó el evaluado no está relacionado con el puntaje en Comunicación Escrita.

5. FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO

La construcción de cada modelo se llevó a cabo a través de una etapa de especificación del modelo, luego se hizo una selección del modelo (teniendo en cuenta el modelo óptimo), y posteriormente la evaluación, a través de los residuales, para finalmente interpretar los resultados.

5.1 ESPECIFICACIÓN DE LOS MODELOS

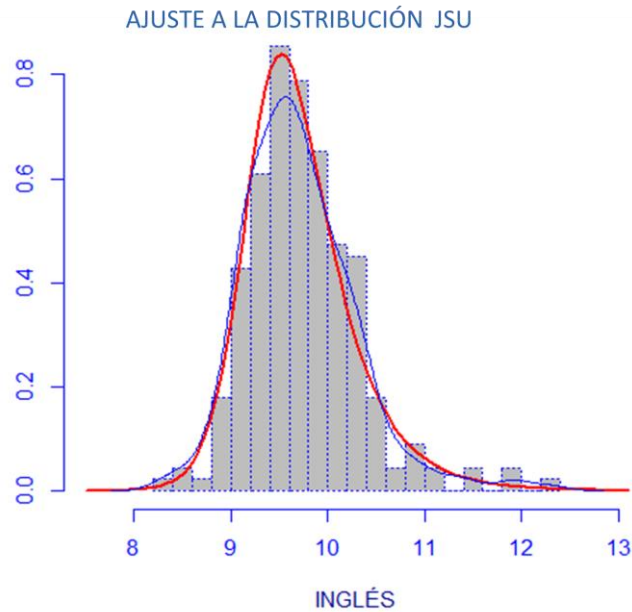
Para el desarrollo de este trabajo se revisó primero el ajuste de los datos de las variables respuesta a una determinada distribución a través del criterio AICG, tal como fue descrito en la sección 2.6.

Después de haber conocido los resultados del criterio AICG de cada distribución y haber seleccionado el valor más pequeño podemos afirmar que para INGLÉS la distribución más adecuada y la que mejor se ajusta a los datos según el criterio AICG es la distribución Johnson Su (JSU).

```
> fitmip$fit
      JSU      ST2      ST1      ST5      SEP1      EGB2      ST3      SEP4
369.4172 369.4206 369.4588 369.5026 369.6632 369.6761 369.8639 370.4218
      SEP2      SN1      RG      TF      SN2      LO      PE      NO
375.6974 376.1939 376.9942 379.2924 379.3166 379.9276 384.7264 397.9140
```

En la gráfica 5-1, se muestra una representación gráfica del ajuste de los datos del puntaje en Inglés a la distribución JSU.

Gráfica 5-1. Ajuste distribución en Inglés

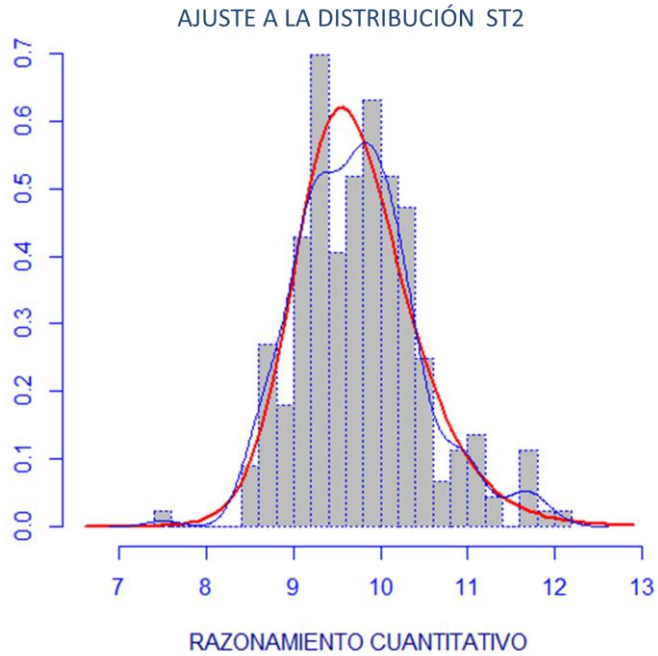


Para RAZONAMIENTO CUANTITATIVO después de haber conocido los resultados del criterio AICG de cada distribución y haber seleccionado el valor más pequeño podemos afirmar que la distribución más adecuada y la que mejor se ajusta a los datos según el criterio AICG es la distribución skew t type 2 (ST2).

```
> fitmrCP$fit
      ST2      JSU      EGB2      ST3      SN1      ST1      ST5      SEP1
475.4979 475.5250 475.5586 475.6768 475.7421 475.7480 475.8145 475.8729
      SEP3      ST4      LO      TF      PE      GT      NO      RG
477.0295 477.0837 478.4461 480.5368 481.5202 482.1937 482.8068 492.7435
```

En la gráfica 5-2 se muestra el ajuste de los datos del puntaje en Razonamiento Cuantitativo a la distribución ST2.

Gráfica 5-2. Ajuste distribución en Razonamiento Cuantitativo

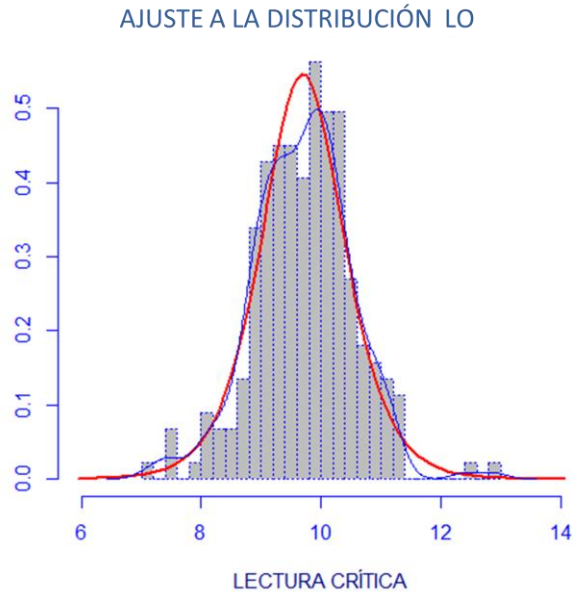


Así mismo para LECTURA CRÍTICA después de haber conocido los resultados del criterio AICG de cada distribución y haber seleccionado el valor más pequeño, podemos afirmar que la distribución más adecuada y la que mejor se ajusta a los datos según el criterio AICG es la distribución logística LO.

```
> fitmlc$fit
      LO      TF      GT      ST3      ST2      ST5      ST1      JSU
543.8337 545.4979 546.0822 547.0729 547.2485 547.3000 547.3494 547.3603
      SEP2  SHASHo  SEP3      NO      SEP4      SN2      SN1      RG
548.8488 549.2695 549.4770 549.8421 549.8712 551.7406 551.7555 599.0619
```

En la gráfica 5-3 se muestra el ajuste de los datos del puntaje en Lectura Crítica a la distribución LO.

Gráfica 5-3. Ajuste distribución en Lectura Crítica

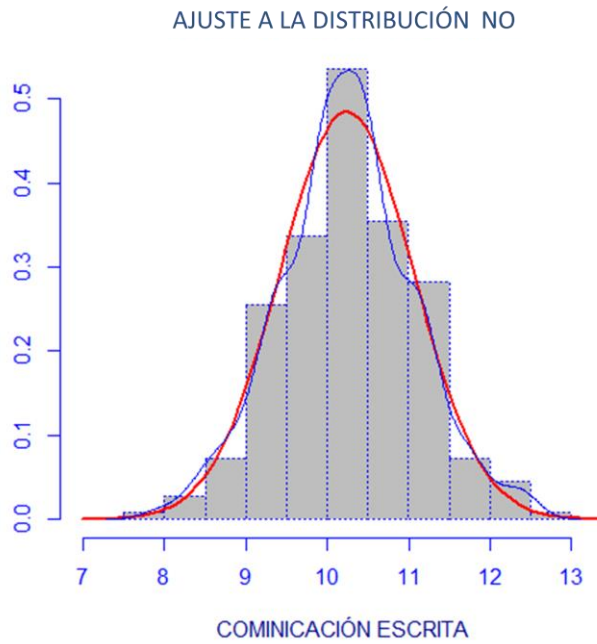


Después de haber conocido los resultados del criterio AICG de cada distribución y haber seleccionado el valor más pequeño podemos afirmar que para COMUNICACIÓN ESCRITA la distribución más adecuada y la que mejor se ajusta a los datos según el criterio AICG es la distribución normal (NO).

```
> fitcmep$fit
      NO      PE      LO      TF      SN1      SN2      SEP4      SEP3
542.6054 543.2447 543.4659 544.3126 544.4649 544.5339 545.1715 545.1858
      EGB2      JSU      ST1      ST4      ST2      ST5      ST3      RG
546.1870 546.2008 546.2277 546.2295 546.2313 546.2327 546.2514 576.1806
```

En la gráfica 5-4 se muestra el ajuste de los datos del puntaje en Comunicación Escrita a la distribución NO.

Gráfica 5-4. Ajuste distribución en Comunicación Escrita



5.2 FACTORES DEMOGRÁFICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

La siguiente tabla muestra las variables explicativas para tener en cuenta en todos los modelos demográficos.

Tabla 23. Componente sistemático. Factor demográfico

NOTACIÓN	NOMBRE
X_1	Edad
X_2	Número de personas que conforman el hogar
X_3	Número de personas de las que se encuentra a cargo el estudiante
X_4	Género
X_5	Es cabeza de familia
X_6	Situación del hogar actual
X_7	Estado civil

El predictor lineal queda determinado por.

$$\eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_r X_4 + \delta_j X_5 + \rho_k X_6 + \varphi_l X_7$$

donde

β_0 . Intercepto

β_1 . Efecto de la edad del estudiante.

β_2 . Efecto del número de personas que conforman el hogar.

β_3 . Efecto del número de personas de las que se encuentra a cargo el estudiante.

γ_r . Efecto del género del estudiante, con $r = 1,2$.

δ_j . Efecto del si es o no cabeza de familia, con $j = 1,2$.

ρ_k . Efecto de la situación del hogar actual del estudiante, con $k = 1,2$.

φ_l . Efecto del estado civil del estudiante, con $l = 1,2,3,4$.

MODELO PARA RAZONAMIENTO CUANTITATIVO

A continuación se presenta el modelo óptimo para RAZONAMIENTO CUANTITATIVO. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-1. Modelo óptimo para Razonamiento Cuantitativo

```
Family: c("ST2", "Skew t (Azzalini type 2)")

Call: gamlss(formula = MOD_RAZONA_CUANTITATIVO_PUNT ~ EDAD + ESTU_SN_CABEZA_FMLIA,
  sigma.formula = ~ESTU_GENERO + EDAD + ESTU_HOGAR_ACTUAL +
    FAMI_NUM_PERS_CARGO + FAMI_NUM_PERS_GRUP_FAM + ESTU_SN_CABEZA_FMLIA,
  family = ST2, data = na.omit(spl))

Fitting method: RS()

-----
Mu link function: identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.20824    0.44371   23.006 <2e-16 ***
EDAD         -0.03540    0.01718   -2.060  0.0411 *
ESTU_SN_CABEZA_FMLIA[T.SI]  0.33735    0.15279    2.208  0.0288 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Continuación modelo anterior

```

-----
Sigma link function: log
Sigma Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.80857    0.86385   2.094 0.03800 *
ESTU_GENERO[T.M] 0.39020    0.12542   3.111 0.00224 **
EDAD          -0.08100    0.03347  -2.420 0.01675 *
ESTU_HOGAR_ACTUAL[T.TEMPORAL] -0.20146    0.13503  -1.492 0.13784
FAMI_NUM_PERS_CARGO -0.18120    0.13031  -1.390 0.16647
FAMI_NUM_PERS_GRUP_FAM -0.06215    0.04071  -1.527 0.12899
ESTU_SN_CABEZA_FMLIA[T.SI] 0.36416    0.22733   1.602 0.11131
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nu link function: identity
Nu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.7187     0.1359   5.289 4.15e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function: log
Tau Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.223e+01  8.032e-07 40129816 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 12
Residual Deg. of Freedom: 143
                        at cycle: 20

Global Deviance: 278.734
AIC: 302.734
SBC: 339.2551

```

Teniendo en cuenta que la variable de Razonamiento Cuantitativo en este caso variable de respuesta se distribuye como Skew t tipo 2 (ST2) y que la función enlace **respecto a la media** es la idéntica se afirma que.

- Con un nivel de significancia inferior al 1%, la variable edad influye en el puntaje en Razonamiento Cuantitativo, es decir, por cada año más que cumpla el estudiante, el puntaje promedio de Razonamiento Cuantitativo disminuye en 0.0354 unidades, dejando las demás variables constantes.
- Con un nivel de significancia inferior al 1%, el puntaje promedio en Razonamiento Cuantitativo aumenta 0.33735 unidades, cuando un

estudiante es cabeza de familia respecto a uno que no lo es, dejando las demás variables constantes.

Respecto a la varianza, teniendo en cuenta que la variable de Razonamiento Cuantitativo en este caso variable de respuesta se distribuye como Skew t tipo 2 (ST2) y que la función enlace respecto a la varianza es el log, se afirma que.

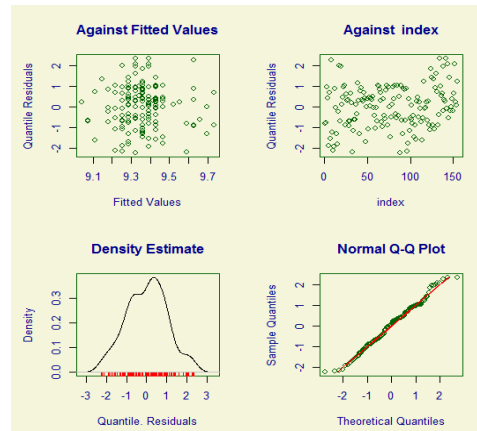
- Con un nivel de significancia inferior al 1%, se puede afirmar que, la edad influye en el puntaje en Razonamiento Cuantitativo, es decir, por cada año más que cumpla el estudiante, el puntaje en Razonamiento Cuantitativo varía en $e^{(-0.08100)} = 0.9221$ unidades, dejando las demás variables constantes.
- Con un nivel de significancia inferior al 0.1%, el puntaje en Razonamiento Cuantitativo varía $e^{(0.39020)} = 1.4772$ unidades, cuando el estudiante es de género masculino respecto a si es de género femenino, dejando las demás variables constantes.

Diagnóstico del modelo óptimo.

Teniendo en cuenta la estimación de los residuos y *QQ – plot*, se puede afirmar que el modelo es adecuado para explicar el puntaje en Razonamiento Cuantitativo.

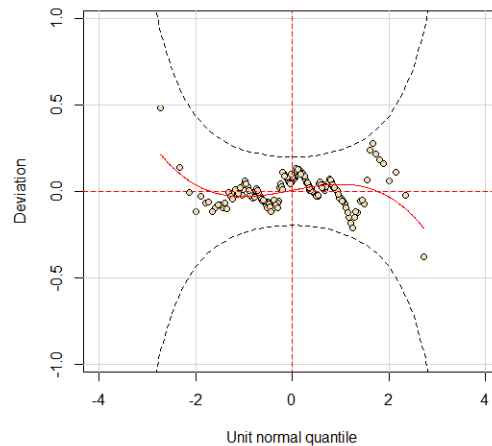
```
> plot(seleccion.rc.d)
*****
          Summary of the Quantile Residuals
          mean      = 0.005360753
          variance  = 1.01275
          coef. of skewness = 0.0003718749
          coef. of kurtosis = 2.647489
          Filliben correlation coefficient = 0.9955619
*****
```

Gráfica 5-5. Diagnóstico modelo para puntaje en Razonamiento Cuantitativo - factor demográfico.



Teniendo en cuenta que en la gráfica 5-6 *worm plot*, , los puntos representan los residuos y la línea discontinua roja (horizontal) son los valores esperados, se puede observar qué tan lejos están unos de otros. Como la mayoría de los residuos aproximadamente el 95% se encuentran entre las dos líneas discontinuas negras y aproximadamente el 5% fuera, se puede afirmar que el modelo es adecuado para explicar la variable de respuesta en este caso el puntaje en Razonamiento Cuantitativo.

Gráfica 5-6. Diagnóstico del modelo para puntaje en Razonamiento Cuantitativo - factor demográfico



MODELO PARA PUNTAJE EN INGLÉS

A continuación se presenta el modelo óptimo para EL PUNTAJE EN INGLÉS. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-2. Modelo óptimo para Inglés

```
*****
Family: c("JSU", "Johnson SU")

Call:  gamlss(formula = MOD_INGLES_PUNT ~ 1, sigma.formula = ~1, family = JSU,
             data = na.omit(sp1))

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.71555    0.03879   250.5  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.54017    0.06122  -8.824 2.27e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function:  identity
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.2656    0.3234   3.913 0.000136 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function:  log
Tau Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.6238    0.1282   4.866 2.79e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 4
      Residual Deg. of Freedom: 151
                          at cycle: 15

Global Deviance:    246.6746
                   AIC:    254.6746
                   SBC:    266.8483
*****
```

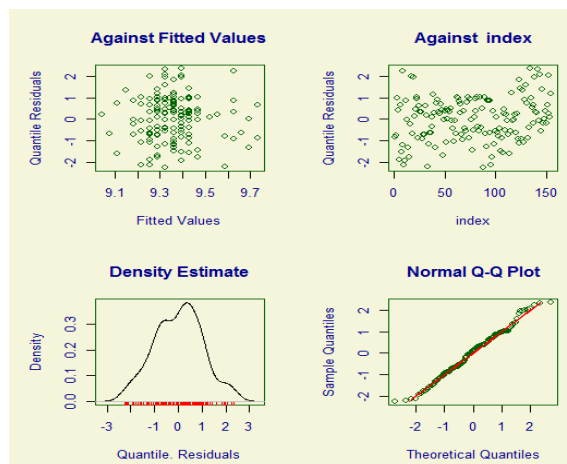
En los resultados estadísticos de los anteriores modelos se muestra que ninguna de las variables demográficas explica la media y la varianza del puntaje en Inglés.

Diagnóstico del modelo óptimo.

Teniendo en cuenta la estimación de densidad de los residuos y *QQ – plot*, se puede afirmar que la distribución es la adecuada para la variable respuesta, ya que hay normalidad en los residuales, la cual se justifica a través de una prueba de hipótesis mediante el AICG (Gráfica 5-8).

```
> plot(seleccion.rc.d)
*****
Summary of the Quantile Residuals
      mean = 0.005360753
  variance = 1.01275
coef. of skewness = 0.0003718749
coef. of kurtosis = 2.647489
Filliben correlation coefficient = 0.9955619
*****
```

Gráfica 5-7. Diagnóstico modelo para puntaje en Inglés - factor demográfico

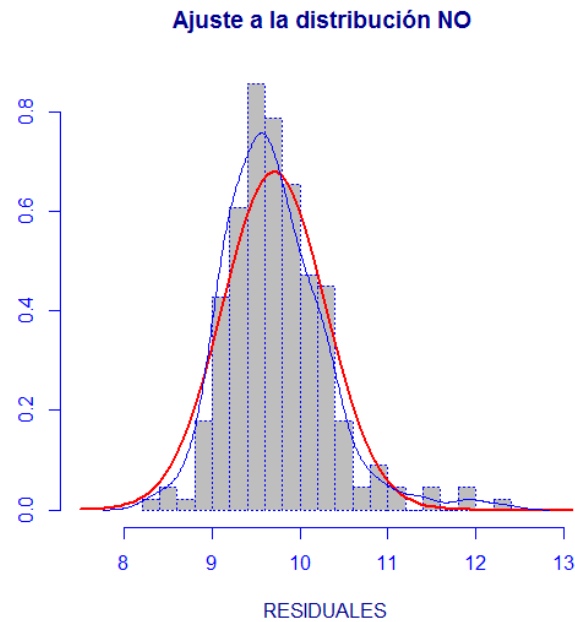


```

> fitmip$fit
      NO      SN2      PE      TF      SN1      LO      SEP1      SEP2
443.7322 445.7154 445.7241 445.7253 445.7322 446.8944 447.6391 447.7050
      SHASHo    ST3      SEP3      ST1      EGB2      ST2      JSU      SEP4
447.7056 447.7081 447.7088 447.7228 447.7229 447.7229 447.7229 447.7233
      ST5      ST4      GU      RG
447.7236 447.7248 468.5084 469.9685

```

Gráfica 5-8. Ajuste de la distribución de los residuales



MODELO PARA LECTURA CRÍTICA

A continuación se presenta el modelo óptimo para LECTURA CRÍTICA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5- 3. Modelo óptimo Lectura Crítica

```

*****
Family: c("LO", "Logistic")

Call:  gamlss(formula = MOD_LECTURA_CRITICA ~ EDAD + FAMI_NUM_PERS_CARGO +
  FAMI_NUM_PERS_GRUP_FAM, sigma.formula = ~ESTU_GENERO + EDAD +
  ESTU_HOGAR_ACTUAL + FAMI_NUM_PERS_CARGO + FAMI_NUM_PERS_GRUP_FAM +
  ESTU_SN_CABEZA_FMLIA, family = LO, data = na.omit(sp1))

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.26154   0.58357  19.298 <2e-16 ***
EDAD           -0.05525   0.02184  -2.529  0.0125 *
FAMI_NUM_PERS_CARGO  0.19786   0.07593   2.606  0.0101 *
FAMI_NUM_PERS_GRUP_FAM -0.06679   0.03501  -1.908  0.0583 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.27311   0.72685  -0.376  0.708
ESTU_GENERO[T.M]  0.06771   0.14353   0.472  0.638
EDAD           -0.03131   0.02792  -1.121  0.264
ESTU_HOGAR_ACTUAL[T.TEMPORAL]  0.01757   0.15540   0.113  0.910
FAMI_NUM_PERS_CARGO -0.01871   0.11682  -0.160  0.873
FAMI_NUM_PERS_GRUP_FAM  0.04751   0.03935   1.208  0.229
ESTU_SN_CABEZA_FMLIA[T.SI] -0.10845   0.25907  -0.419  0.676

-----
No. of observations in the fit:  155
Degrees of Freedom for the fit:  11
Residual Deg. of Freedom:  144
                             at cycle:  3

Global Deviance:  356.2244
AIC:  378.2244
SBC:  411.7021
*****

```

Teniendo en cuenta que la variable de Lectura Crítica en este caso variable de respuesta se distribuye como una logarítmica (LO) y que la función enlace **respecto a la media** es la idéntica, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Con un nivel de significancia inferior al 1% se puede afirmar que la edad influye en el puntaje en Lectura Crítica, es decir, por cada año más que cumpla el estudiante, el puntaje promedio de Lectura Crítica disminuye en 0.05525 unidades.
- Con un nivel de significancia inferior al 1%, el puntaje promedio de Lectura Crítica aumenta 0.19786 unidades, por una persona más que el estudiante se encuentre a cargo.
- Con un nivel de significancia inferior al 5%, se puede afirmar que, el puntaje promedio de Lectura Crítica disminuye 0.06679 unidades, por una persona más que conforme el hogar de los estudiantes.

Respecto a la varianza, teniendo en cuenta que la variable de Lectura Crítica en este caso variable de respuesta se distribuye como una logarítmica (LO) y que la función enlace respecto a la varianza es el log, se afirma que.

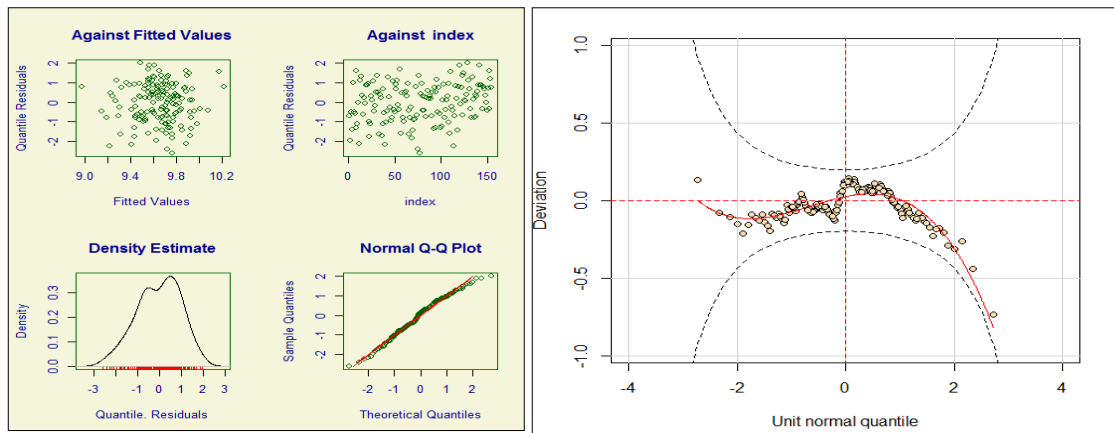
La edad, el género, el número de personas que conforman el hogar del estudiante, el número de personas de las que se encuentra a cargo, explican la variabilidad del puntaje en Lectura Crítica.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar el puntaje de Lectura Crítica.

```
> plot(seleccion.lc.d)
*****
      Summary of the Quantile Residuals
      mean      = -0.0321781
      variance  =  0.9813315
      coef. of skewness = -0.2929289
      coef. of kurtosis  =  2.474367
      Filliben correlation coefficient =  0.9931915
*****
```

Gráfica 5-9. Diagnóstico modelo para el puntaje en Lectura Crítica - factor demográfico (*plot* y el *worm plot*)



MODELO PARA COMUNICACIÓN ESCRITA

A continuación se presenta el modelo óptimo para COMUNICACIÓN ESCRITA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-4. Modelo óptimo para Comunicación Escrita

```
*****
Family: c("NO", "Normal")

Call: gamlss(formula = MOD_COMUNICA_ESCRITA_PUNT ~ EDAD + ESTU_ESTADO_CIVIL,
  sigma.formula = ~ESTU_GENERO + EDAD + ESTU_HOGAR_ACTUAL +
    FAMI_NUM_PERS_CARGO + FAMI_NUM_PERS_GRUP_FAM + ESTU_SN_CABEZA_FMLIA,
  family = NO, data = na.omit(sp1))

Fitting method: RS()

-----
Mu link function: identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.91724    0.74866   15.918  <2e-16 ***
EDAD           -0.05203    0.02510   -2.073  0.0399 *
ESTU_ESTADO_CIVIL[T.SLT] -0.47496    0.40013   -1.187  0.2371
ESTU_ESTADO_CIVIL[T.SP] -0.50826    0.77153   -0.659  0.5110
ESTU_ESTADO_CIVIL[T.UL]  0.17806    0.46300    0.385  0.7011
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Continuacion modelo óptimo Comunicación Escrita

```

-----
Sigma link function:  log
Sigma Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.061054  0.614602  -1.726  0.0864 .
ESTU_GENERO[T.M]  0.091826  0.121360   0.757  0.4505
EDAD           0.026463  0.023608   1.121  0.2641
ESTU_HOGAR_ACTUAL[T.TEMPORAL] 0.160835  0.131400   1.224  0.2229
FAMI_NUM_PERS_CARGO -0.003889  0.098777  -0.039  0.9686
FAMI_NUM_PERS_GRUP_FAM  0.019877  0.033271   0.597  0.5511
ESTU_SN_CABEZA_FMLIA[T.SI] -0.223326  0.219059  -1.019  0.3096
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 12
  Residual Deg. of Freedom: 143
                        at cycle: 3

Global Deviance:    359.0618
                   AIC:    383.0618
                   SBC:    419.5829
*****

```

Teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace **respecto a la media** es la idéntica, se afirma que.

Con un nivel de significancia inferior al 5% se puede afirmar que la edad influye en el puntaje medio de Comunicación Escrita, es decir, por cada año más que cumpla el estudiante, el puntaje medio de Razonamiento Cuantitativo disminuye en 0.05203 unidades, dejando las demás variables constantes.

Respecto a la varianza, según los resultados estadísticos anteriores, teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace respecto a la varianza es el log, entonces se afirma que.

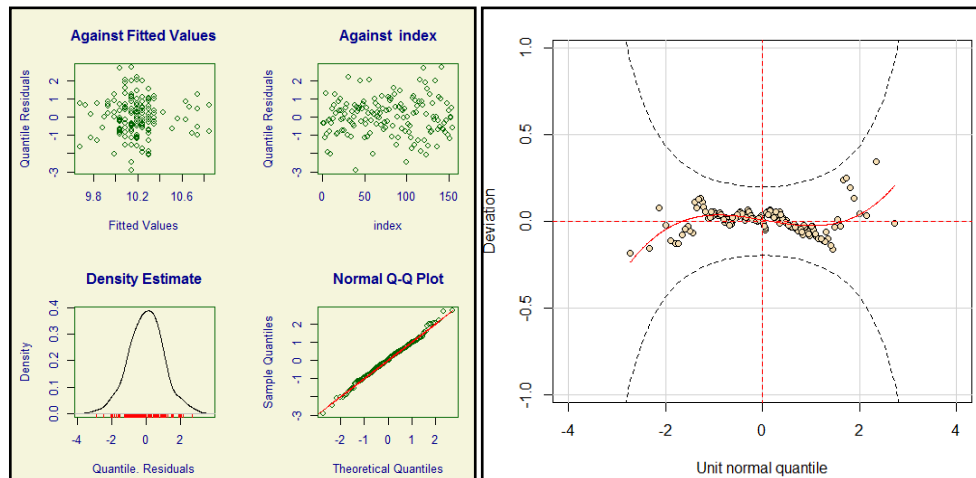
El puntaje de Comunicación Escrita varía $e^{(-1.0610)} = 0.364$ unidades, cuando no depende de ninguna variable.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar El puntaje de Comunicación Escrita.

```
> plot(seleccion.ce.d)
*****
                Summary of the Quantile Residuals
                mean      = 0.005164114
                variance  = 1.006466
                coef. of skewness = -0.009443475
                coef. of kurtosis  = 3.197293
                Filliben correlation coefficient = 0.9975257
                *****
```

Gráfica 5-10. Diagnóstico modelo para el puntaje en Comunicación Escrita - factor demográfico (*plot* y el *worm plot*)



En cuanto al **DESEMPEÑO EN INGLÉS** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo y su interpretación.

Tabla 24. Modelo óptimo para el desempeño en Inglés - factor demográfico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE	INTERVALO DE CONF 95.00%	
ESTU_HOGAR_ACTUAL [T. TEMPORAL]	-0,2998	0,199	-1,50653	0,7410	0,93404	0,501	1,093
A- A1	-0,5267	0,1185	-4,44473	0,5906			
A1 A2	0,5917	0,1209	4,89413	1,8071			
A2 B+	1,5048	0,1689	8,90941	4,5033			
B+ B1	1,5647	0,1749	8,94626	4,7812			

Según la razón de probabilidad (OR) y el intervalo de confianza al 95%, es más probable que un estudiante con una situación del hogar temporal tenga valores más altos en el desempeño de Inglés que un estudiante con una situación del hogar habitual o permanente.

En cuanto al **DESEMPEÑO EN COMUNICACIÓN ESCRITA**, se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo y su interpretación.

Tabla 25. Modelo óptimo para el Desempeño en Comunicación Escrita - f actor demográfico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE	INTERVALO DE CONF 95.00%	
ESTU_HOGAR_ACTUAL [T. TEMPORAL]	0,3574	0,926	0,386	1,4296	0,3499	0,9807	2,0863
2 3	-2,1505	0,2782	-7,73	0,1164			
3 4	-1,2151	0,1459	-8,328	0,2967			
4 5	0,0686	0,1128	0,608	1,0710			
5 6	1,244	0,1431	8,693	3,4695			
6 7	2,3639	0,286	8,265	10,6323			

Según la razón de probabilidad (OR) y el intervalo de confianza al 95%, es más probable que un estudiante con una situación del hogar temporal tenga valores

más altos en el desempeño de Comunicación Escrita que un estudiante con una situación del hogar habitual o permanente.

5.3 FACTORES SOCIOECONÓMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

La siguiente tabla muestra las variables explicativas para tener en cuenta en todos los modelos socioeconómicos.

Tabla 26. Componente sistemático. Factor socioeconómico

Notación	Nombre
X_1	El hogar cuenta con celular
X_2	El hogar cuenta con conexión a internet
X_3	El hogar cuenta con Servicio cerrado de televisión
X_4	El hogar cuenta con servicio de teléfono fijo
X_5	El hogar cuenta con lavadora
X_6	El hogar cuenta con computador
X_7	Valor anual de la matrícula del año anterior
X_8	Ocupación del padre
X_9	Ocupación de la madre
X_{10}	Estrato
X_{11}	Ingresos familiares mensuales representado en salarios mínimos mensuales
X_{12}	Trabaja actualmente
X_{13}	Forma de pago de la matrícula

El modelo lineal queda determinado por el siguiente predictor lineal

$$\eta_i = \beta_0 + \rho_r X_1 + \lambda_j X_2 + v_k X_3 + \varrho_l X_4 + \partial_m X_5 + \theta_n X_6 + \gamma_f X_7 + \delta_t X_8 + \vartheta_g X_9 + \varphi_z X_{10} + \tau_d X_{11} + \pi_h X_{12} + \omega_p X_{13}$$

donde

β_0 . Intercepto

ρ_r . Efecto del sí o no cuenta con celular, con $r = 1, 2$.

λ_j . Efecto del sí o no cuenta con conexión a internet en el hogar, con $j = 1, 2$.

v_k . Efecto del sí o no cuenta con servicio cerrado de televisión en el hogar, con $k = 1, 2$.

- ϱ_l . Efecto del sí o no cuenta con servicio de teléfono fijo en el hogar, con $l = 1, 2$.
- ∂_m . Efecto del sí o no cuenta con lavadora en el hogar, con $m = 1, 2$.
- θ_n . El hogar cuenta con computador, con $n = 1, 2$.
- γ_f . Efecto del valor anual de la matrícula del año anterior del estudiante, con $f = 1, 2, 3, 4$.
- δ_t . Efecto de la ocupación del padre del estudiante, con $t = 1, 2, 3, 4$.
- ϑ_g . Efecto de la ocupación de la madre del estudiante, con $g = 1, 2, 3, 4$.
- φ_z . Efecto del estrato, con $z = 1, 2, 3, 4$.
- τ_d . Efecto de los Ingresos familiares mensuales familiares representado en salarios mínimos mensuales del estudiante, con $d = 1, 2, 3, 4$.
- π_h . Efecto del sí o no trabaja actualmente, con $h = 1, 2, 3, 4$.
- ω_p . Forma de pago de la matrícula, con $p = 1, 2, 3, 4$.

MODELO RAZONAMIENTO CUANTITATIVO

A continuación se presenta el modelo óptimo para RAZONAMIENTO CUANTITATIVO. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-5. Modelo óptimo para Razonamiento Cuantitativo

```

*****
Family:  c("ST2", "Skew t (Azzalini type 2)")

Call:  gamlss(formula = MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_CELULAR +
  FAMI_COD_OCUP_MADRE, sigma.formula = ~ECON_SN_TELEFONIA +
  ECON_SN_LAVADORA, family = ST2, data = na.omit(sp1))

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.0327    0.1631  55.390 < 2e-16 ***
ECON_SN_CELULAR[T.SI]  0.4068    0.1341   3.034  0.00284 **
FAMI_COD_OCUP_MADRE[T.HOGAR] -0.2700    0.1276  -2.116  0.03602 *
FAMI_COD_OCUP_MADRE[T.PENSIONADO.Y.A] -0.3567    0.2646  -1.348  0.17969
FAMI_COD_OCUP_MADRE[T.TRABAJADOR.I.(PI.E)] -0.1122    0.1893  -0.593  0.55420
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Continuación Modelo óptimo para Razonamiento Cuantitativo

```
Sigma link function: log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.33141    0.09356  -3.542 0.000527 ***
ECON_SN_TELEFONIA[T.SI]  0.27667    0.23528   1.176 0.241467
ECON_SN_LAVADORA[T.SI] -0.15737    0.13934  -1.129 0.260511
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Nu link function: identity
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4285     0.2701   5.288 4.17e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function: log
Tau Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.7673     0.3862   4.576 9.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 10
Residual Deg. of Freedom: 145
                        at cycle: 20

Global Deviance:    296.9021
AIC:                316.9021
SBC:                347.3364
*****
```

Teniendo en cuenta que la variable de Razonamiento Cuantitativo en este caso variable de respuesta se distribuye como Skew t tipo 2 (ST2) y que la función enlace **respecto a la media** es la idéntica se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Con un nivel de significancia inferior al 0.1%, se puede afirmar que si un estudiante tiene celular influye positivamente en la media del puntaje en Razonamiento Cuantitativo respecto a uno que no lo tiene.
- Con un nivel de significancia inferior al 1%, para aquel estudiante cuya madre es ama de casa influye negativamente en la media del puntaje en Razonamiento Cuantitativo respecto a otro estudiante cuya madre de familia es empleada de cualquier cargo.

Respecto a la varianza, según los resultados del modelo anterior, se puede afirmar que, el tener lavadora en el hogar y teléfono fijo respecto a si no se tienen estos servicios, explica la variabilidad del puntaje en Razonamiento Cuantitativo, siempre y cuando las demás variables sean constantes.

Diagnóstico del modelo óptimo.

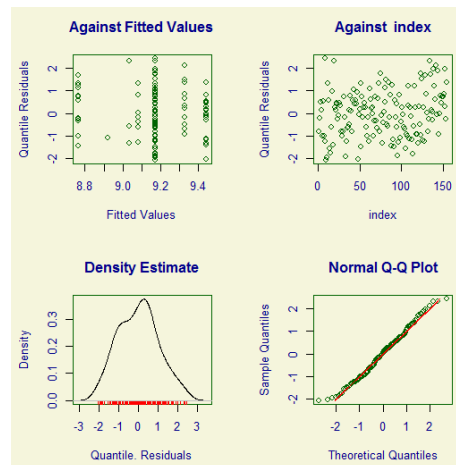
Teniendo en cuenta la estimación de densidad de los residuos y *QQ – plot*(grafica 5-, se puede afirmar que el modelo es adecuado para explicar el puntaje en Razonamiento Cuantitativo.

```

*****
> plot(seleccion.rc.s)
*****
Summary of the Quantile Residuals
      mean      = 0.003373019
      variance   = 1.027922
      coef. of skewness = 0.1485714
      coef. of kurtosis  = 2.455688
Filliben correlation coefficient = 0.9951485
*****

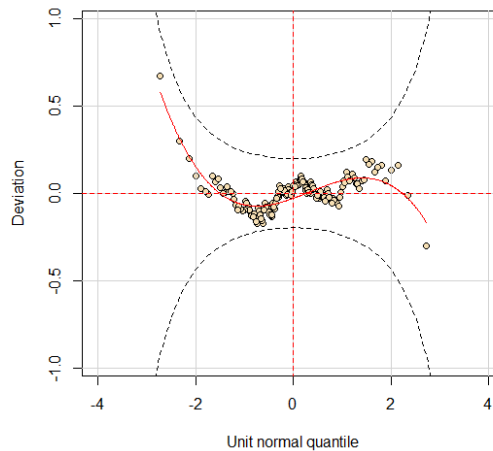
```

Gráfica 5-11. Diagnóstico modelo para el puntaje en Razonamiento Cuantitativo - factor socioeconómico



Teniendo en cuenta que en la gráfica 5-12 *worm plot*, los puntos representan los residuos y la línea discontinua roja (horizontal) son los valores esperados, se puede observar que tan lejos están unos de otros. Como la mayoría de los puntos (residuos) aproximadamente el 95% se encuentran entre las dos líneas discontinuas negras y aproximadamente el 5% fuera, se puede afirmar que el modelo es adecuado para explicar la variable de respuesta.

Gráfica 5-12. Diagnóstico modelo para el puntaje en Razonamiento Cuantitativo - factor socioeconómico



MODELO PARA PUNTAJE EN INGLÉS

A continuación se presenta el modelo óptimo para EL PUNTAJE EN INGLÉS. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-6. Modelo óptimo para el puntaje en Inglés

```

-----
Mu link function: identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.79994	0.16283	60.184	< 2e-16 ***
factor(INST_VLR_MATRICULA_ANT) [T.2]	0.35531	0.07763	4.577	1.04e-05 ***
factor(INST_VLR_MATRICULA_ANT) [T.3]	1.13516	0.14681	7.732	1.98e-12 ***
ECON_SN_CELULAR[T.SI]	-0.49235	0.12775	-3.854	0.000177 ***
ECON_SN_SERVICIO_TV[T.SI]	-0.07324	0.04791	-1.529	0.128649
ECON_SN_TELEFONIA[T.SI]	0.05039	0.06969	0.723	0.470869
ECON_SN_LAVADORA[T.SI]	0.06347	0.04787	1.326	0.187082
ESTU_DORMITORIOS	0.03979	0.02509	1.586	0.115008
FAMI_COD_OCUP_PADRE [T.HOGAR]	0.09462	0.10194	0.928	0.354903
FAMI_COD_OCUP_PADRE [T.PENSIONADO.Y.A]	0.39202	0.09176	4.272	3.58e-05 ***
FAMI_COD_OCUP_PADRE [T.TRABAJADOR.I. (PI.E)]	0.33891	0.04597	7.373	1.40e-11 ***
FAMI_COD_OCUP_MADRE [T.HOGAR]	-0.01132	0.06330	-0.179	0.858346
FAMI_COD_OCUP_MADRE [T.PENSIONADO.Y.A]	0.34016	0.11582	2.937	0.003884 **
FAMI_COD_OCUP_MADRE [T.TRABAJADOR.I. (PI.E)]	-0.17746	0.08400	-2.112	0.036445 *
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2]	0.10919	0.05191	2.104	0.037231 *
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3]	-0.17309	0.07608	-2.275	0.024428 *
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4]	-0.07368	0.12045	-0.612	0.541772

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.39494	0.31083	-1.271	0.205998
factor(INST_VLR_MATRICULA_ANT) [T.2]	0.52700	0.12242	4.305	3.14e-05 ***
factor(INST_VLR_MATRICULA_ANT) [T.3]	1.11732	0.39017	2.864	0.004837 **
ECON_SN_CELULAR[T.SI]	-0.52313	0.14407	-3.631	0.000396 ***
ECON_SN_INTERNET[T.SI]	-0.27947	0.12000	-2.329	0.021308 *
ECON_SN_SERVICIO_TV[T.SI]	-0.34285	0.08632	-3.972	0.000114 ***
ESTU_DORMITORIOS	0.11817	0.06654	1.776	0.077927 .
FAMI_COD_OCUP_PADRE [T.HOGAR]	-0.05929	0.34132	-0.174	0.862352
FAMI_COD_OCUP_PADRE [T.PENSIONADO.Y.A]	0.43158	0.16510	2.614	0.009933 **
FAMI_COD_OCUP_PADRE [T.TRABAJADOR.I. (PI.E)]	0.30962	0.10072	3.074	0.002543 **
FAMI_COD_OCUP_MADRE [T.HOGAR]	0.28410	0.14736	1.928	0.055905 .
FAMI_COD_OCUP_MADRE [T.PENSIONADO.Y.A]	0.39193	0.28434	1.378	0.170291
FAMI_COD_OCUP_MADRE [T.TRABAJADOR.I. (PI.E)]	-0.60392	0.20805	-2.903	0.004304 **
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2]	0.35290	0.10926	3.230	0.001546 **
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3]	-0.82700	0.24125	-3.428	0.000801 ***
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4]	-1.63482	0.31173	-5.244	5.71e-07 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: identity
Nu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.8880	0.1431	13.19	<2e-16 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function: log
Tau Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.00273	0.02363	-0.116	0.908

```

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 35
Residual Deg. of Freedom: 120
at cycle: 20

Global Deviance: 172.4313
AIC: 242.4313
SBC: 348.9511
*****

```

Según los resultados del modelo estadístico anterior **respecto a la media**, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- El valor de la matrícula, el tener celular, la ocupación del padre, la ocupación de la madre y los ingresos familiares mensuales familiares influyen en el puntaje promedio de Inglés.
- Cuando un estudiante cuenta en su hogar con celular, el puntaje promedio de Inglés disminuye 0.4923 unidades respecto a uno que no cuenta con este dispositivo electrónico.
- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 500 mil y menos de un millón el puntaje promedio de Inglés aumenta en 0.35531 unidades respecto a otro estudiante que pago menos de 500 mil.
- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 1 millón y tres millones el puntaje promedio de Inglés aumenta en 1.13516 unidades respecto a un estudiante que pago menos de 500 mil.
- Al comparar dos estudiantes que tan sólo difieren en que uno tiene padre de familia pensionado y otro tiene padre familia empleado de diferentes cargos, el puntaje promedio de Inglés aumenta en 0.39202 unidades para aquel estudiante con padre pensionado.
- El puntaje promedio de Inglés aumenta en 0.33891 unidades para aquel estudiante cuyo padre es trabajador independiente o un profesional independiente respecto a otro estudiante cuyo padre es empleado de diferentes cargos.
- Con un nivel de significancia inferior al 0.1%, el puntaje promedio de Inglés aumenta en 0.34016 unidades para aquel estudiante cuya madre de familia es pensionada respecto a otro estudiante cuya madre de familia es empleada de diferentes cargos.

- Con un nivel de significancia inferior al 1%, el puntaje promedio de Inglés disminuye en 0.33891 unidades para aquel estudiante cuya madre de familia es trabajadora independiente o una profesional independiente respecto a otro estudiante cuya madre es empleada de diferentes cargos.
- Con un nivel de significancia inferior al 1%, el puntaje medio en Inglés aumenta en 0.10919 unidades para aquel estudiante cuyos los ingresos familiares mensuales familiares de un estudiante están entre uno y menos de dos salarios mínimos el puntaje medio en Inglés aumenta en 0.10919 unidades respecto a otro estudiante cuyos ingresos son menos de un salario mínimo.
- Con un nivel de significancia inferior al 1%, se puede afirmar que, si los ingresos familiares mensuales familiares de un estudiante están entre dos y menos de tres salarios mínimos su puntaje medio en Inglés disminuye en 0.17309 unidades respecto a otro estudiante con ingresos familiares mensuales de menos de un salario mínimo.

Para el modelo ajustado **respecto a la varianza**, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- El valor de la matrícula, el tener celular, tener internet, tener TV, la ocupación del padre, la ocupación de la madre y los ingresos familiares mensuales familiares influyen en la variabilidad del puntaje en Inglés.
- Cuando un estudiante cuenta en su hogar con celular el puntaje en Inglés varía en $e^{(-0.52313)} = 0.5926$ unidades, respecto a uno que no cuenta con este móvil.
- Cuando un estudiante cuenta en su hogar con servicio cerrado de televisión el puntaje en Inglés varía en $e^{(-0.34285)} = 0.7097$ unidades, respecto a otro que no cuenta en su hogar con este electrodoméstico.

- El puntaje en Inglés en un estudiante que cuenta en su hogar con conexión de internet varía en $e^{(-0.2794)} = 0.756184$ unidades respecto a otro que no cuenta con este servicio.
- Si el valor anual de la matrícula del año anterior de un estudiante fue entre 500 mil y menos de un millón, el puntaje en Inglés varía en $e^{(0.527)} = 1.6938$ unidades, respecto a otro estudiante que pago menos de 500 mil.
- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 1 millón y tres millones, el puntaje en Inglés varía en $e^{(1.1173)} = 3.0566$ unidades, respecto a otro estudiante que pago menos de 500 mil.
- Para un estudiante cuyo padre es pensionado el puntaje en Inglés varía en $e^{(0.43158)} = 1.539688$ unidades, respecto a otro estudiante cuyo padre es empleado de diferentes cargos.
- Para un estudiante cuyo padre es trabajador independiente o un profesional independiente, el puntaje en Inglés varía en $e^{(0.3096)} = 1.3629$ unidades, respecto a otro estudiante cuyo padre es empleado de diferentes cargos.
- Para un estudiante cuya madre de familia es trabajadora independiente o una profesional independiente, el puntaje en Inglés varía en $e^{(-0.6039)} = 0.54667$ unidades, respecto a otro estudiante cuya madre de familia es empleada de diferentes cargos.
- Para un estudiante cuyos ingresos familiares mensuales están entre uno y menos de dos salarios mínimos, el puntaje en Inglés varía en $e^{(0.3529)} = 1.4223188$ unidades, respecto a otro estudiante cuyos ingresos familiares mensuales son menos de un salario mínimo.
- Para un estudiante cuyos ingresos familiares mensuales están entre dos y menos de tres salarios mínimos, el puntaje en Inglés varía en $e^{(-0.82700)} = 0.437359$ unidades, respecto a otro estudiante cuyos ingresos son de menos un salario mínimo.

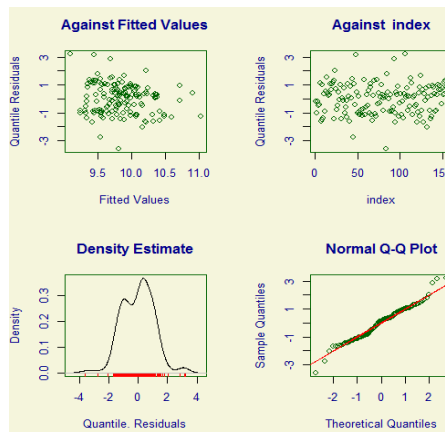
- Para un estudiante cuyos ingresos familiares mensuales están entre tres y menos de siete salarios mínimos, el puntaje en Inglés varía en $e^{(-1.6348)} = 0.19499$ unidades, respecto a otro estudiante cuyos ingresos familiares mensuales son de menos de un salario mínimo.

Diagnóstico del modelo óptimo.

Teniendo en cuenta la estimación de densidad de los residuos y *QQ – plot*, se puede afirmar que la distribución Johnson Su (JSU) es la adecuada para el puntaje en Inglés.

```
> plot(seleccion.pi.s)
*****
Summary of the Quantile Residuals
      mean      = -0.01927569
  variance      =  1.136685
coef. of skewness =  0.03750138
coef. of kurtosis =  3.521936
Filliben correlation coefficient =  0.9893339
*****
```

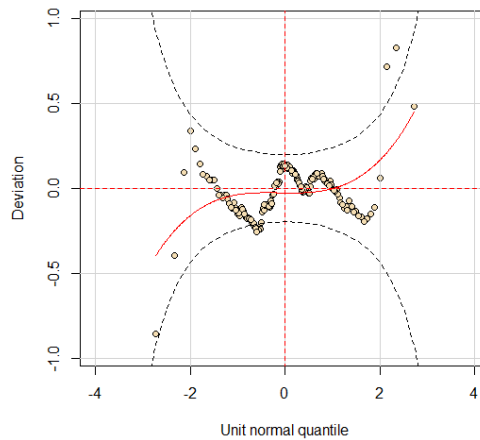
Gráfica 5-13. Diagnóstico del modelo para el puntaje en Inglés - factor socioeconómico



Teniendo en cuenta que en la gráfica 5-14 *worm plot*, los puntos representan los residuos y la línea discontinua roja (horizontal) son los valores esperados, se puede observar que tan lejos están unos de otros. Como la mayoría de los puntos

(residuos) aproximadamente el 95% se encuentran entre las dos líneas discontinuas negras el 5% fuera, se puede afirmar que el modelo es adecuado para explicar la variable de respuesta.

Gráfica 5-14. Diagnóstico del modelo para el puntaje en Inglés - factor socioeconómico



MODELO PARA LECTURA CRÍTICA

A continuación se presenta el modelo óptimo para LECTURA CRÍTICA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-7. Modelo óptimo para Lectura Crítica

```

-----
Mu link function:  identity
Mu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.84859   0.11494  85.686 < 2e-16 ***
ESTU_TRABAJA[T.EXPERIENCIA.GASTOS.P] -0.36984   0.18257  -2.026  0.04459 *
ESTU_TRABAJA[T.NO] -0.44493   0.10780  -4.127  6.11e-05 ***
ESTU_TRABAJA[T.PRACTICA.O] -1.64023   0.93811  -1.748  0.08246 .
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2]  0.09138   0.13319   0.686  0.49373
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3] -0.21318   0.17026  -1.252  0.21251
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4]  0.46728   0.16166   2.891  0.00443 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

Sigma link function:  log
Sigma Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.536657  0.610851  -0.879  0.3812
factor(INST_VLR_MATRICULA_ANT) [T.2]  0.346133  0.181106  1.911  0.0581 .
factor(INST_VLR_MATRICULA_ANT) [T.3] -0.005182  0.442229  -0.012  0.9907
FORMA_PAGO_MATRICULA[T.PP] -0.273971  0.483961  -0.566  0.5723
FORMA_PAGO_MATRICULA[T.VFP] -0.246066  0.471908  -0.521  0.6029
ECON_SN_CELULAR[T.SI] -0.159064  0.223516  -0.712  0.4779
ECON_SN_INTERNET[T.SI] -0.037163  0.198950  -0.187  0.8521
ECON_SN_SERVICIO_TV[T.SI] -0.177714  0.159376  -1.115  0.2668
ECON_SN_TELEFONIA[T.SI]  0.290489  0.232100  1.252  0.2129
ECON_SN_LAVADORA[T.SI] -0.146578  0.162177  -0.904  0.3677
ECON_SN_COMPUTADOR[T.SI]  0.138220  0.179457  0.770  0.4425
ESTU_DORMITORIOS  0.071974  0.076167  0.945  0.3464
FAMI_COD_OCUP_PADRE [T.HOGAR] -0.095308  0.487161  -0.196  0.8452
FAMI_COD_OCUP_PADRE [T.PENSIONADO.Y.A] -0.026530  0.219769  -0.121  0.9041
FAMI_COD_OCUP_PADRE [T.TRABAJADOR.I. (PI.E)] -0.043588  0.164856  -0.264  0.7919
FAMI_COD_OCUP_MADRE [T.HOGAR] -0.115194  0.196793  -0.585  0.5593
FAMI_COD_OCUP_MADRE [T.PENSIONADO.Y.A]  0.406352  0.409126  0.993  0.3224
FAMI_COD_OCUP_MADRE [T.TRABAJADOR.I. (PI.E)] -0.422691  0.270951  -1.560  0.1211
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2] -0.030112  0.182029  -0.165  0.8689
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3] -0.392362  0.281401  -1.394  0.1655
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4] -0.829363  0.377879  -2.195  0.0299 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

No. of observations in the fit:  155
Degrees of Freedom for the fit:  28
Residual Deg. of Freedom:  127
                             at cycle:  4

Global Deviance:  335.3663
AIC:  391.3663
SBC:  476.5822
*****

```

Según los resultados estadísticos anteriores para el modelo ajustado del puntaje en Lectura Crítica **respecto a la media**, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Cuando un estudiante trabaja para adquirir experiencia y sus recursos para gastos personales el puntaje promedio de Lectura Crítica disminuye 0.36984 unidades respecto a otro estudiante que trabaja como ayudante sin remuneración.
- Cuando un estudiante no trabaja el puntaje promedio de Lectura Crítica disminuye 0.44493 unidades respecto a otro estudiante que trabaja como ayudante sin remuneración.
- Cuando un estudiante trabaja por ser práctica obligatoria del plan de estudios el puntaje promedio de Lectura Crítica disminuye 1.64023 unidades respecto a otro estudiante que trabaja como ayudante sin remuneración.
- Para un estudiante cuyos ingresos familiares mensuales están entre tres y menos de siete salarios mínimos su puntaje promedio en Lectura Crítica aumenta en 0.4672 unidades respecto a otro estudiante cuyos ingresos familiares mensuales son menos de un salario mínimo.

Teniendo en cuenta que la variable de Lectura Crítica en este caso variable de respuesta se distribuye como una logarítmica (LO) y que la función enlace **respecto a la varianza** es el log, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 500 mil y menos de un millón, el puntaje en Lectura Crítica varía en

$e^{(0.346133)} = 1.41359$ unidades, respecto a otro estudiante que pago menos de 500 mil.

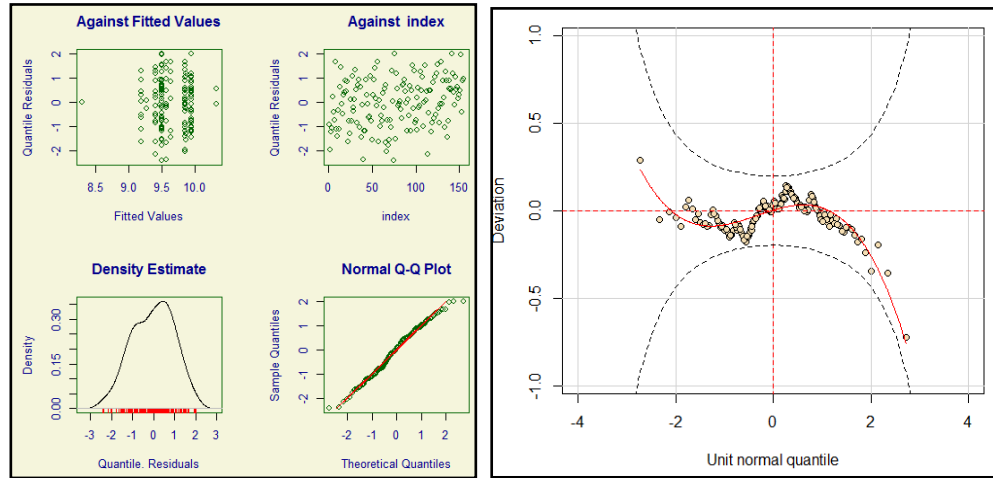
- Para un estudiante cuyos ingresos familiares mensuales están entre tres y menos de siete salarios mínimos, el puntaje en Lectura Crítica varía en $e^{(-0.8293)} = 0.4363$ unidades, respecto a otro estudiante cuyos ingresos familiares mensuales son de menos de un salario mínimo.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar El puntaje de Lectura Crítica.

```
> plot(seleccion.lcs)
*****
          Summary of the Quantile Residuals
              mean      = -0.03196543
              variance   =  0.9728019
              coef. of skewness = -0.1754809
              coef. of kurtosis =  2.340982
Filliben correlation coefficient =  0.9943812
*****
```

Gráfica 5-15. Diagnóstico modelo para el puntaje en Lectura Crítica - factor demográfico (*plot* y el *worm plot*)



MODELO PARA COMUNICACIÓN ESCRITA

A continuación se presenta el modelo óptimo para COMUNICACIÓN ESCRITA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-8. Modelo óptimo Comunicación escrita

```
-----
Mu link function: identity
Mu Coefficients:

                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    10.21594    0.20129   50.752 < 2e-16 ***
ESTU_TRABAJA[T.EXPERIENCIA.GASTOS.P] -0.27551    0.09185   -2.999 0.003226 **
ESTU_TRABAJA[T.NO]              -0.10742    0.06463   -1.662 0.098842 .
ESTU_TRABAJA[T.PRACTICA.O]      -2.05376    0.80874   -2.539 0.012245 *
factor(ESTU ESTRATO) [T.2]       0.01252    0.14987    0.084 0.933547
factor(ESTU ESTRATO) [T.3]       0.05836    0.15838    0.368 0.713116
factor(ESTU ESTRATO) [T.4]      -1.03527    0.18660   -5.548 1.49e-07 ***
ECON_SN_CELULAR[T.SI]           0.53181    0.11010    4.830 3.67e-06 ***
ECON_SN_INTERNET[T.SI]          0.13486    0.08011    1.683 0.094634 .
ECON_SN_SERVICIO_TV[T.SI]       -0.21256    0.07053   -3.014 0.003088 **
ECON_SN_TELEFONIA[T.SI]         0.39010    0.13624    2.863 0.004868 **
ECON_SN_LAVADORA[T.SI]          -0.16188    0.06795   -2.382 0.018614 *
FAMI_COD_OCUP_PADRE[T.HOGAR]    -1.88134    0.21025   -8.948 2.60e-15 ***
FAMI_COD_OCUP_PADRE[T.PENSIONADO.Y.A] -0.37269    0.07137   -5.222 6.60e-07 ***
FAMI_COD_OCUP_PADRE[T.TRABAJADOR.I.(PI.E)] 0.06848    0.07313    0.936 0.350781
FAMI_COD_OCUP_MADRE[T.HOGAR]    -0.16715    0.10141   -1.648 0.101618
FAMI_COD_OCUP_MADRE[T.PENSIONADO.Y.A] 0.54134    0.15403    3.515 0.000601 ***
FAMI_COD_OCUP_MADRE[T.TRABAJADOR.I.(PI.E)] -0.13836    0.10860   -1.274 0.204833
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2] -0.10651    0.07420   -1.436 0.153469
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3] -0.11860    0.07914   -1.499 0.136344
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4] -0.94054    0.15146   -6.210 6.21e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Continuación modelo óptimo Comunicación escrita

```

-----
Sigma link function: log
Sigma Coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -3.669e-02  5.232e-01  -0.070  0.944198
factor(ESTU_ESTRATO) [T.2]  -9.005e-02  1.846e-01  -0.488  0.626513
factor(ESTU_ESTRATO) [T.3]  -3.440e-01  2.461e-01  -1.398  0.164528
factor(ESTU_ESTRATO) [T.4]  -3.673e+01  7.664e-01  -47.926 < 2e-16 ***
factor(INST_VLR_MATRICULA_ANT) [T.2]  -3.754e-01  1.539e-01  -2.439  0.016048 *
factor(INST_VLR_MATRICULA_ANT) [T.3]   8.223e-01  3.783e-01   2.174  0.031520 *
FORMA_PAGO_MATRICULA[T.PP]    4.689e-01  4.098e-01   1.144  0.254575
FORMA_PAGO_MATRICULA[T.VFP]   1.281e-01  4.003e-01   0.320  0.749376
ECON_SN_CELULAR[T.SI]        -8.026e-04  1.894e-01  -0.004  0.996625
ECON_SN_INTERNET[T.SI]       -1.726e+00  1.689e-01  -10.223 < 2e-16 ***
ECON_SN_SERVICIO_TV[T.SI]     7.529e-01  1.352e-01   5.567  1.40e-07 ***
ECON_SN_TELEFONIA[T.SI]      1.851e+00  1.964e-01   9.426 < 2e-16 ***
ECON_SN_LAVADORA[T.SI]       -1.533e-01  1.375e-01  -1.115  0.267016
ECON_SN_COMPUTADOR[T.SI]      5.189e-01  1.572e-01   3.302  0.001239 **
ESTU_DORMITORIOS            -3.001e-01  6.524e-02  -4.600  9.84e-06 ***
FAMI_COD_OCUP_PADRE [T.HOGAR]  -2.717e-01  4.125e-01  -0.659  0.511190
FAMI_COD_OCUP_PADRE [T.PENSIONADO.Y.A]  4.997e-03  1.895e-01   0.026  0.979002
FAMI_COD_OCUP_PADRE [T.TRABAJADOR.I.(PI.E)]  5.447e-01  1.408e-01   3.869  0.000171 ***
FAMI_COD_OCUP_MADRE [T.HOGAR]  -3.495e-01  1.696e-01  -2.061  0.041249 *
FAMI_COD_OCUP_MADRE [T.PENSIONADO.Y.A]   3.094e-01  3.479e-01   0.889  0.375436
FAMI_COD_OCUP_MADRE [T.TRABAJADOR.I.(PI.E)] -7.730e-01  2.322e-01  -3.328  0.001134 **
factor(FAMI_ING_FMLIAR_MENSUAL) [T.2]    1.524e-01  1.546e-01   0.986  0.325871
factor(FAMI_ING_FMLIAR_MENSUAL) [T.3]    3.381e-03  2.398e-01   0.014  0.988771
factor(FAMI_ING_FMLIAR_MENSUAL) [T.4]   -1.071e+00  3.197e-01  -3.349  0.001058 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 45
  Residual Deg. of Freedom: 110
                        at cycle: 15

Global Deviance:    220.6976
                   AIC:    310.6976
                   SBC:    447.6517
*****

```

Teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace **respecto a la media** es la idéntica, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Cuando un estudiante trabaja por adquirir experiencia y/o recursos para sus gastos personales, el puntaje medio en Comunicación Escrita disminuye en 0.27551 unidades respecto a otro que trabaja como ayudante sin remuneración.

- Cuando un estudiante no trabaja, el puntaje medio en Comunicación Escrita disminuye en 0.10742 unidades respecto a otro que trabaja como ayudante sin remuneración.
- Cuando un estudiante está realizando la práctica del plan de estudios, el puntaje medio en Comunicación Escrita disminuye en 2.05376 unidades respecto a otro que trabaja como ayudante sin remuneración.
- Cuando un estudiante se encuentra en estrato 4, el puntaje medio en Comunicación Escrita disminuye 1.03527 respecto a otro estudiante de estrato uno.
- Cuando un estudiante posee celular, el puntaje medio en Comunicación Escrita aumenta 0.53181 unidades respecto a otro que no posee celular.
- Cuando un estudiante cuenta en su hogar con TV el puntaje medio en Comunicación Escrita disminuye en 0.21256 unidades respecto a otro no tiene este electrodoméstico.
- Cuando un estudiante cuenta en su hogar con teléfono, el puntaje medio en Comunicación Escrita aumenta en 0.3901 unidades respecto a otro que no cuenta en su hogar con este electrodoméstico.
- Cuando un estudiante cuenta en su hogar con lavadora el puntaje medio en Comunicación Escrita disminuye 0.16188 unidades respecto a otro que no cuenta con este electrodoméstico en su hogar.
- El puntaje medio de Comunicación Escrita disminuye en 1.88134 unidades para aquel estudiante cuyo padre de familia su ocupación es el hogar respecto a otro estudiante cuyo padre es empleado de diferentes cargos.
- El puntaje medio de Comunicación Escrita disminuye en 0.37269 unidades para aquel estudiante cuyo padre es pensionado respecto a otro estudiante cuyo padre de familia es empleado de diferentes cargos.
- El puntaje medio de Comunicación Escrita aumenta en 0.06848 unidades para aquel estudiante cuyo padre es trabajador independiente o un

profesional independiente respecto a otro estudiante cuyo padre es empleado de diferentes cargos.

- El puntaje medio de Comunicación Escrita disminuye en 0.13836 unidades para aquel estudiante cuya la madre de familia es pensionada respecto a otro estudiante cuya madre es empleada de diferentes cargos.
- El puntaje medio en Comunicación Escrita disminuye en 0.94054 unidades para un estudiante cuyos los ingresos familiares mensuales familiares están entre tres y menos de siete salarios mínimos respecto a otro estudiante cuyos ingresos familiares mensuales son menos de un salario mínimo.

Respecto a la varianza, teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace respecto a la varianza es el log, se afirma que al comparar dos estudiantes manteniendo las demás variables constantes se obtiene que.

- Para un estudiante cuyo estrato socioeconómico de la residencia según la factura de energía es el nivel 4, el puntaje en Comunicación Escrita varía en $e^{(-36.73)} = 1.118 \times 10^{-16}$ unidades, respecto a otro estudiante que pertenece a un estrato socioeconómico de nivel uno.
- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 500 mil y menos de un millón, el puntaje en Comunicación Escrita varía en $e^{(-0.3754)} = 0.6870$ unidades, respecto a otro estudiante que pago menos de 500 mil.
- Para un estudiante cuyo valor anual de la matrícula del año anterior fue entre 1 millón y 3 millones, el puntaje en Comunicación Escrita varía en $e^{(0.8223)} = 2.2757$ unidades, respecto a otro estudiante que pago menos de 500 mil.

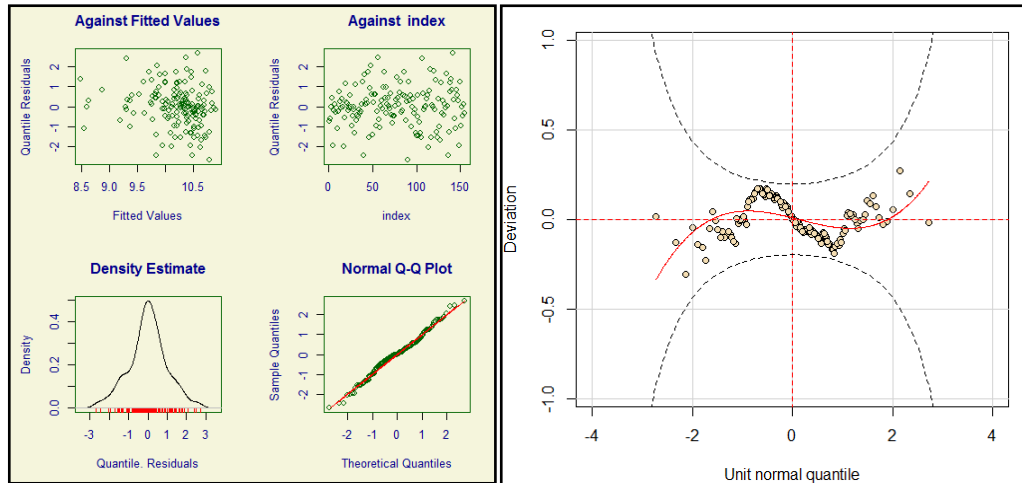
- Cuando un estudiante cuenta en el hogar con servicio cerrado de televisión su puntaje en Comunicación Escrita varía en $e^{(0.7529)} = 2.12314$ unidades, respecto a otro estudiante que no cuenta con este servicio.
- Cuando un estudiante cuenta en el hogar con conexión a internet su puntaje en Comunicación Escrita varía en $e^{(-0.726)} = 0.178$ $e^{(-1.726)} = 0.178$ unidades respecto a otro que no cuenta con este servicio.
- Cuando un estudiante cuenta en el hogar con servicio de teléfono fijo, el puntaje en Comunicación Escrita varía en $e^{(1.857)} = 6.4045$ unidades respecto a otro estudiante que no cuenta con este servicio.
- Cuando un estudiante cuenta en el hogar con computador el puntaje en Comunicación Escrita varía en $e^{(0.5189)} = 1.68$ unidades respecto a otro estudiante que no cuenta con este artefacto tecnológico.
- Para un estudiante cuyo padre de familia es un trabajador independiente o un profesional independiente, el puntaje en Comunicación Escrita varía en $e^{(0.5447)} = 1.724091$ unidades respecto a otro estudiante cuyo padre es empleado de diferentes cargos.
- Para un estudiante cuya madre de familia es madre de hogar el puntaje en Comunicación Escrita varía en $e^{(-0.3495)} = 0.7050$ unidades, respecto a otro estudiante cuya madre es empleada de diferentes cargos.
- Para un estudiante cuya madre de familia es trabajadora independiente o una profesional independiente, el puntaje en Comunicación Escrita varía en $e^{(-0.73)} = 0.4019$ unidades, respecto a otro estudiante cuya madre es empleada de diferentes cargos.
- Para un estudiante cuyos ingresos familiares mensuales están entre tres y menos de siete salarios mínimos el puntaje en Comunicación Escrita varía en $e^{(-1.071)} = 0.34266$ unidades, respecto a otro estudiante cuyos ingresos familiares mensuales son menos de un salario mínimo.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar el puntaje de Comunicación Escrita.

```
> plot(seleccion.ce.s)
*****
                Summary of the Quantile Residuals
                mean      = -0.0006066627
                variance  =  0.9999971
                coef. of skewness = -0.05726902
                coef. of kurtosis =  3.257703
                Filliben correlation coefficient = 0.9946491
                *****
```

Gráfica 5-16. Diagnóstico del modelo para el puntaje en Comunicación Escrita - factor socioeconómico



En cuanto al DESEMPEÑO EN INGLÉS se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo y su interpretación.

Tabla 27. Modelo óptimo para el desempeño en Inglés - factor socioeconómico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE	INTERVALO DE CONF 95.00%	
INST_VLR_MATRICULA_ANT [T.2]	0,1282	0,2098	0,6111	1,1368	0,27056	0,7532	1,7142
INST_VLR_MATRICULA_ANT [T.3]	1,7696	0,5462	3,2398	5,8685	0,000598	2,0613	17,727
ECON_SN_LAVADORA [T.SI]	0,4048	0,1835	2,2060	1,4990	0,013692	1,0471	2,1497
A- A1	-0,1718	0,1573	-1,0922	0,8421			
A1 A2	1,0014	0,1709	5,8596	2,7221			
A2 B+	1,9999	0,2205	9,0698	7,3883			
B+ B1	2,0704	0,2274	9,1047	7,9280			

Según la razón de probabilidad (OR) y el intervalo de confianza al 95%, es más probable que un estudiante con valor anual de la matrícula del año anterior entre 500 mil y menos de un millón, tenga valores más altos en el desempeño de Inglés que un estudiante que pago menos de 500 mil en la matricula del año anterior. También se puede afirmar que si el valor anual de la matrícula del año anterior de un estudiante fue entre un (1) millón y tres (3) millones, es más probable que tenga valores más altos en el desempeño de Inglés respecto a un estudiante que pago menos de 500 mil en la matricula del año anterior.

En cuanto al DESEMPEÑO EN COMUNICACIÓN ESCRITA, se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo y su interpretación.

Tabla 28. Modelo óptimo para el desempeño en Comunicación Escrita - factor socioeconómico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE	INTERVALO DE CONF 95.00%	
FAMI_COD_OCUP_PADRE [T. HOGAR]	-1,0506	0,5611	-1,872	0,3497	0,9694	0,1156	1,0444
FAMI_COD_OCUP_PADRE [T. PENSIONADO]	0,0963	0,2446	0,394	1,1011	0,34679	0,6818	1,7789
FAMI_COD_OCUP_PADRE [T. TRABAJADOR.]	0,5898	0,196	3,009	1,8036	0,00131	1,2297	2,6512
2 3	-2,1719	0,3231	-6,722	0,114			
3 4	-1,1316	0,1735	-6,522	0,3225			
4 5	0,2105	0,1493	1,41	1,2343			
5 6	1,4234	0,1796	7,925	4,1512			
6 7	2,5933	0,3183	8,147	13,374			

Según la razón de probabilidad (OR) y el intervalo de confianza al 95%, es más probable que un estudiante con un padre pensionado tenga valores más altos en el desempeño de Comunicación Escrita que un estudiante con un padre que es empleado de diferentes cargos. También se puede afirmar que, si un estudiante tiene un padre con ocupación de trabajador independiente, es probable que tenga valores más altos en el desempeño de Comunicación Escrita que un estudiante con un padre que es empleado de diferentes cargos.

5.4 FACTORES ACADÉMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

La siguiente tabla muestra las variables explicativas para tener en cuenta en todos los modelos académicos.

Tabla 29. Componente sistemático factor académico

Notación	Nombre
X_1	Tomó algún curso de preparación para el examen
X_2	Tipo de bachillerato del que se graduó el evaluado
X_3	Máximo nivel educativo alcanzado por el padre
X_4	Máximo nivel educativo alcanzado por la madre

El modelo lineal queda determinado por el siguiente predictor lineal

$$\eta_i = \beta_0 + \rho_j X_1 + \pi_r X_2 + \gamma_k X_3 + \delta_m X_4$$

donde

β_0 . Intercepto

ρ_j . Efecto del sí o no tomó algún curso de preparación para el examen, con

$j = 1, 2$.

π_r . Efecto del tipo de bachillerato del que se graduó el evaluado, con $r = 1, 2, 3$.

γ_k . Efecto del máximo nivel educativo alcanzado por el padre del estudiante, con $k = 1, 2, 3, 4$.

δ_m . Efecto del máximo nivel educativo alcanzado por la madre del estudiante, con $m = 1, 2, 3, 4$.

MODELO PARA RAZONAMIENTO CUANTITATIVO

A continuación se presenta el modelo óptimo para RAZONAMIENTO CUANTITATIVO. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-9. Modelo óptimo para Razonamiento Cuantitativo

```

Mu link function: identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.3941	0.1571	59.786	< 2e-16	***
FAMI_COD_EDUCA_PADRE[T.NINGUNO]	1.5362	0.2650	5.797	3.93e-08	***
FAMI_COD_EDUCA_PADRE[T.PRIMARIA]	0.1919	0.1717	1.118	0.26540	
FAMI_COD_EDUCA_PADRE[T.SECUNDARIA]	0.3677	0.1681	2.188	0.03027	*
FAMI_COD_EDUCA_MADRE[T.NINGUNO]	-0.8035	0.2570	-3.126	0.00213	**
FAMI_COD_EDUCA_MADRE[T.PRIMARIA]	-0.2439	0.1825	-1.336	0.18357	
FAMI_COD_EDUCA_MADRE[T.SECUNDARIA]	-0.2333	0.1704	-1.369	0.17293	

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.53805	0.18708	-2.876	0.00461	**
FAMI_COD_EDUCA_MADRE[T.NINGUNO]	-2.38714	1.54381	-1.546	0.12413	
FAMI_COD_EDUCA_MADRE[T.PRIMARIA]	-0.03214	0.21081	-0.152	0.87901	
FAMI_COD_EDUCA_MADRE[T.SECUNDARIA]	-0.02549	0.22048	-0.116	0.90810	

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Continuación modelo óptimo para Razonamiento Cuantitativo

```
-----  
Nu link function: identity  
Nu Coefficients:  
      Estimate Std. Error t value Pr(>|t|)  
(Intercept)  0.5496      0.1298   4.233 3.95e-05 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
-----  
Tau link function: log  
Tau Coefficients:  
      Estimate Std. Error t value Pr(>|t|)  
(Intercept)  1.7533      0.3688   4.754 4.54e-06 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
-----  
No. of observations in the fit: 155  
Degrees of Freedom for the fit: 13  
      Residual Deg. of Freedom: 142  
              at cycle: 20  
  
Global Deviance:    288.1991  
      AIC:           314.1991  
      SBC:           353.7636  
*****
```

Teniendo en cuenta que la variable de Razonamiento Cuantitativo en este caso variable de respuesta se distribuye como Skew t tipo 2 (ST2) y que la función enlace **respecto a la media** es la idéntica se afirma que.

- Al comparar dos estudiantes que tan sólo difieren en que uno cuyo padre no realizo ningún nivel educativo y otro cuyo padre alcanzó un nivel de educación superior, el puntaje promedio de Razonamiento Cuantitativo aumenta en 1.5362 unidades para aquel estudiante donde el padre no realizo ningún nivel educativo.
- El puntaje promedio en Razonamiento Cuantitativo aumenta 0.3677 unidades para aquel estudiante cuyo padre de familia alcanzó un nivel educativo de básica secundaria respecto a otro estudiante cuyo padre de familia alcanzó un nivel de educación superior, siempre y cuando las demás variables sean constantes.
- El puntaje promedio de Razonamiento Cuantitativo disminuye en 0.8035 unidades para aquel estudiante cuya madre no realizo ningún nivel

educativo respecto otro estudiante cuya madre si alcanzó un nivel de educación superior, siempre y cuando las demás variables sean constantes.

Teniendo en cuenta que la variable de Razonamiento Cuantitativo en este caso variable de respuesta se distribuye como Skew t tipo 2 (ST2) y que la función enlace respecto a la varianza es el log, se afirma que.

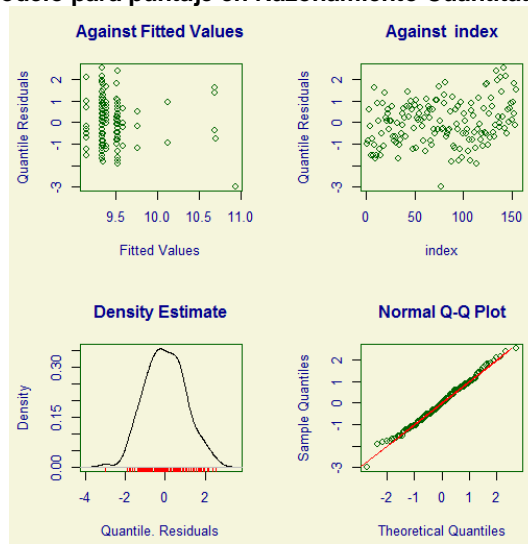
- El máximo nivel educativo alcanzado por la madre y el padre del estudiante es el único que explica la variabilidad del puntaje en Razonamiento Cuantitativo.

Diagnóstico del modelo óptimo.

Teniendo en cuenta la estimación de densidad de los residuos y *QQ – plot*, se puede afirmar que el modelo es adecuado para explicar el puntaje en Razonamiento Cuantitativo.

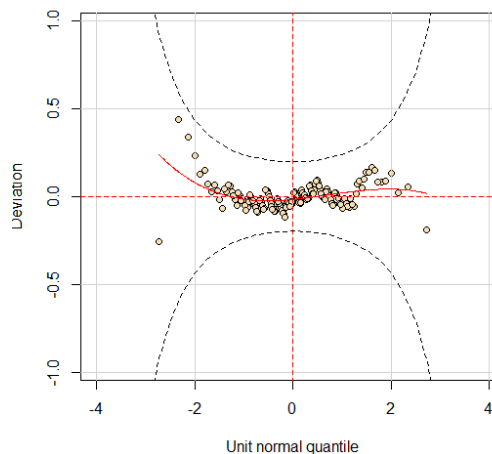
```
> plot(seleccion.rc.a)
*****
          Summary of the Quantile Residuals
              mean      = 0.002589606
              variance   = 1.006206
              coef. of skewness = 0.09307186
              coef. of kurtosis  = 2.773277
Filliben correlation coefficient = 0.9969588
```

Gráfica 5-17. Diagnóstico modelo para puntaje en Razonamiento Cuantitativo - factor académico



Teniendo en cuenta que en la gráfica 5-18 *worm plot*, los puntos representan los residuos y la línea discontinua roja (horizontal) son los valores esperados, se puede observar qué tan lejos están unos de otros. Como la mayoría de los residuos aproximadamente el 95% se encuentran entre las dos líneas discontinuas negras y aproximadamente el 5% fuera, se puede afirmar que el modelo es adecuado para explicar la variable de respuesta en este caso el puntaje en Razonamiento Cuantitativo.

Gráfica 5-18. Diagnóstico del modelo para puntaje en Razonamiento Cuantitativo – factor académico



MODELO PARA PUNTAJE EN INGLÉS

A continuación se presenta el modelo óptimo para EL PUNTAJE EN INGLÉS. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-10. Modelo óptimo para Inglés

```
-----
Mu link function: identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.86547    0.09456 104.333  <2e-16 ***
FAMI_COD_EDUCA_PADRE[T.NINGUNO]  0.10678    0.18143   0.589  0.5570
FAMI_COD_EDUCA_PADRE[T.PRIMARIA] -0.26627    0.10680  -2.493  0.0137 *
FAMI_COD_EDUCA_PADRE[T.SECUNDARIA] -0.06297    0.12179  -0.517  0.6059
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.56107    0.05841  -9.606  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: identity
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.5331    0.3551   4.317 2.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function: log
Tau Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6461    0.1111   5.816 3.38e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 155
Degrees of Freedom for the fit: 7
Residual Deg. of Freedom: 148
                        at cycle: 20

Global Deviance: 236.1872
AIC: 250.1872
SBC: 271.4912
*****
```


Teniendo en cuenta que la variable de Inglés en este caso variable de respuesta se distribuye como (JSU) y que la función enlace **respecto a la media** es la idéntica se afirma que.

- El puntaje medio en Inglés disminuye 0.2662 unidades para aquel estudiante cuyo padre alcanzó un nivel educativo de básica primaria respecto a otro estudiante cuyo padre de familia alcanzó un nivel de educación superior alcanzado, siempre y cuando las demás variables sean constantes.

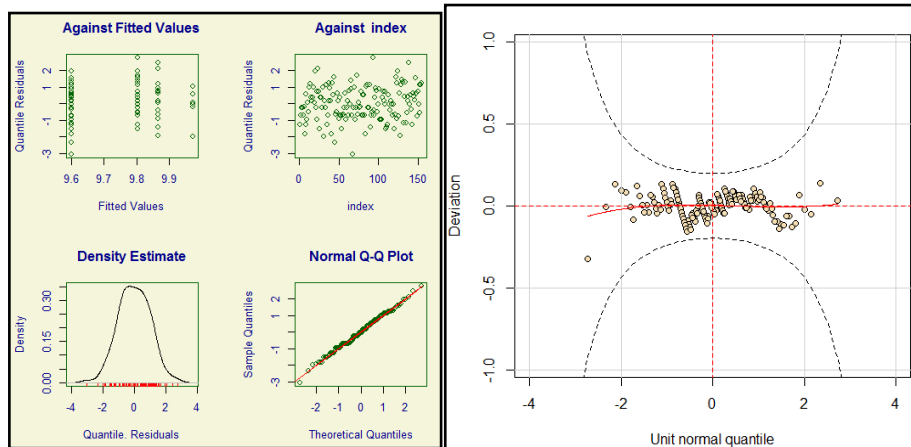
Según el modelo óptimo anterior, se afirma que ninguna variable académica explica la variabilidad del puntaje en Inglés.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar el puntaje de Inglés.

```
> plot(seleccion.pi.a)
*****
          Summary of the Quantile Residuals
              mean = -4.589725e-05
              variance = 1.005067
              coef. of skewness = -0.02784244
              coef. of kurtosis = 2.923871
Filliben correlation coefficient = 0.9975158
*****
```

Gráfica 5-19. Diagnóstico del modelo para el puntaje en Inglés - factor académico



MODELO PARA LECTURA CRÍTICA

A continuación se presenta el modelo óptimo para LECTURA CRÍTICA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-11. Modelo óptimo Lectura Crítica

```
Family: c("LO", "Logistic")

Call: gamlss(formula = MOD_LECTURA_CRITICA ~ FAMI_COD_EDUCA_MADRE,
  sigma.formula = ~FAMI_COD_EDUCA_PADRE + FAMI_COD_EDUCA_MADRE,
  family = LO, data = na.omit(sp1))

Fitting method: RS()

-----
Mu link function: identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.54059    0.18592  51.314 < 2e-16 ***
FAMI_COD_EDUCA_MADRE[T.NINGUNO]  0.95943    0.24446   3.925 0.000132 ***
FAMI_COD_EDUCA_MADRE[T.PRIMARIA]  0.05629    0.20593   0.273 0.784972
FAMI_COD_EDUCA_MADRE[T.SECUNDARIA] 0.16948    0.21003   0.807 0.420975
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

-----
Sigma link function:  log
Sigma Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.74335   0.21004  -3.539 0.000537 ***
FAMI_COD_EDUCA_PADRE[T.NINGUNO]  0.46574   0.42098   1.106 0.270386
FAMI_COD_EDUCA_PADRE[T.PRIMARIA] -0.03884   0.22947  -0.169 0.865809
FAMI_COD_EDUCA_PADRE[T.SECUNDARIA] 0.02774   0.22853   0.121 0.903539
FAMI_COD_EDUCA_MADRE[T.NINGUNO]  -1.76583   0.72326  -2.441 0.015807 *
FAMI_COD_EDUCA_MADRE[T.PRIMARIA]  -0.06924   0.24349  -0.284 0.776523
FAMI_COD_EDUCA_MADRE[T.SECUNDARIA] -0.09347   0.23252  -0.402 0.688279
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----

No. of observations in the fit: 155
Degrees of Freedom for the fit: 11
  Residual Deg. of Freedom: 144
                        at cycle: 3

Global Deviance:    359.9042
      AIC:          381.9042
      SBC:          415.3818
*****

```

Teniendo en cuenta que la variable de Lectura Crítica en este caso variable de respuesta se distribuye como una logarítmica (LO) y que la función enlace **respecto a la media** es la idéntica, se afirma que.

- El puntaje medio de Lectura Crítica aumenta en 0.9594 unidades para aquel estudiante cuya madre no realizó ningún nivel educativo respecto a otro estudiante cuya madre de familia alcanzó un nivel de educación superior, siempre y cuando las demás variables sean constantes.

Teniendo en cuenta que la variable de Lectura Crítica en este caso variable de respuesta se distribuye como una logarítmica (LO) y que la función enlace **respecto a la varianza** es el log, se afirma que.

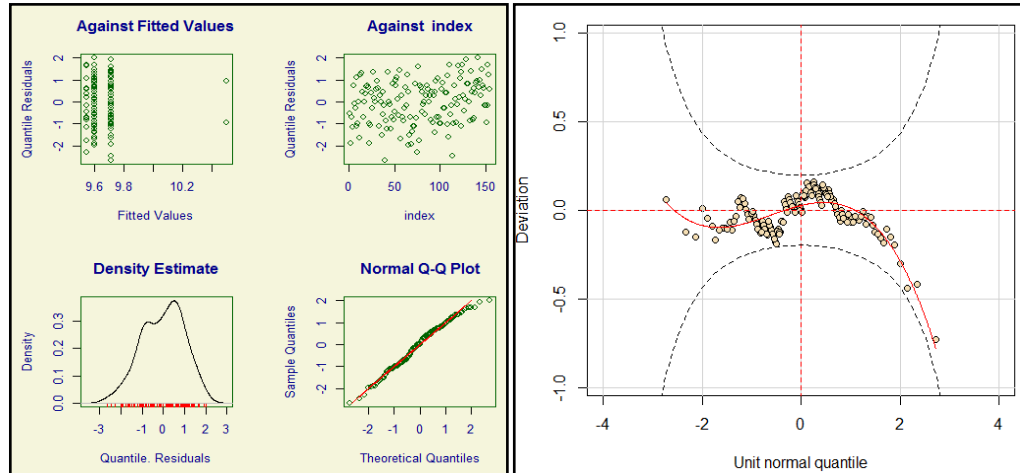
- Para un estudiante cuya madre de familia no realizó ningún nivel educativo el puntaje en Lectura Crítica varía en $e^{(-1.7658)} = 0.1710$ unidades respecto a otro estudiante que alcanzó un nivel de educación superior, siempre y cuando las demás variables sean constantes.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar el puntaje de Lectura Crítica.

```
> plot(seleccion.lc.a)
*****
                Summary of the Quantile Residuals
                mean      = -0.02517027
                variance  =  0.9788812
                coef. of skewness = -0.2628203
                coef. of kurtosis =  2.46649
                Filliben correlation coefficient =  0.9931615
                *****
```

Gráfica 5-20. Diagnóstico modelo para el puntaje en Lectura Crítica - factor académico (*plot* y el *worm plot*)



MODELO PARA COMUNICACIÓN ESCRITA

A continuación se presenta el modelo óptimo para COMUNICACIÓN ESCRITA. Para la selección del modelo óptimo se utilizó el método hacia adelante con criterio de selección AICG.

Imagen 5-12. Modelo óptimo para Comunicación Escrita

```
-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   10.18567   0.06179   164.8  <2e-16 ***
ESTU_TOMOCURSO[T.SI]  0.91433   0.06179   14.8  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.33531   0.18171  -1.845  0.0670 .
ESTU_TOMOCURSO[T.SI] -36.11588   0.73034 -49.451  <2e-16 ***
ESTU_TITULO_BTO[T.N]  -0.39059   0.51688  -0.756  0.4511
ESTU_TITULO_BTO[T.T]   0.22698   0.11741   1.933  0.0552 .
FAMI_COD_EDUCA_PADRE[T.NINGUNO]  0.16720   0.35767   0.467  0.6409
FAMI_COD_EDUCA_PADRE[T.PRIMARIA]  0.15517   0.20002   0.776  0.4391
FAMI_COD_EDUCA_PADRE[T.SECUNDARIA]  0.18904   0.19846   0.953  0.3424
FAMI_COD_EDUCA_MADRE[T.NINGUNO]  0.12873   0.61775   0.208  0.8352
FAMI_COD_EDUCA_MADRE[T.PRIMARIA] -0.24719   0.20996  -1.177  0.2410
FAMI_COD_EDUCA_MADRE[T.SECUNDARIA] -0.09246   0.20104  -0.460  0.6463
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  155
Degrees of Freedom for the fit:  12
      Residual Deg. of Freedom:  143
                        at cycle:  3

Global Deviance:  292.9248
      AIC:  316.9248
      SBC:  353.4459
*****
```

Teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace **respecto a la media** es la idéntica, se afirma que.

- Al comparar dos estudiantes que tan sólo difieren en que uno que tomo curso de preparación para el examen respecto a otro estudiante que no

tomo este curso, el puntaje medio en Comunicación Escrita aumenta 0.914 unidades para aquel estudiante que tomo curso.

Teniendo en cuenta que la variable Comunicación Escrita en este caso variable de respuesta se distribuye como una normal (NO) y que la función enlace **respecto a la varianza** es el log, se afirma que.

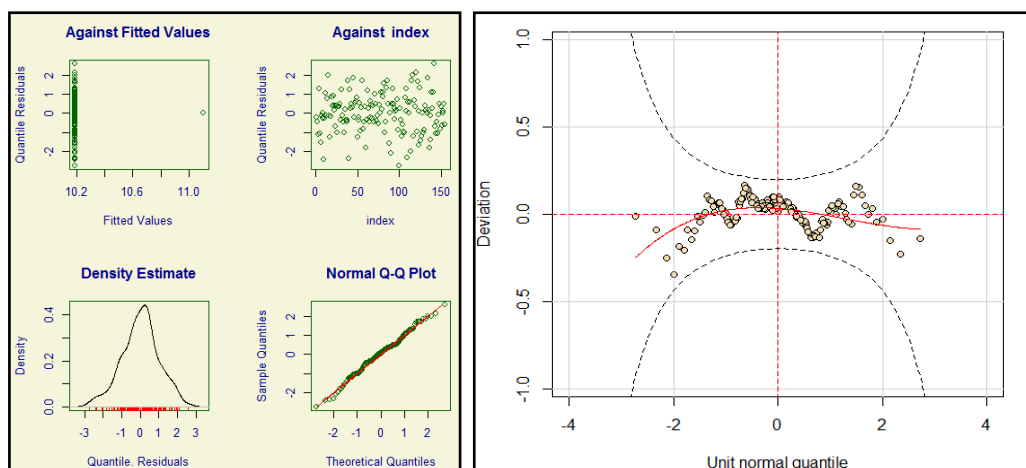
- Al comparar dos estudiantes que tan sólo difieren en que uno con título de bachillerato técnico el puntaje en Lectura Crítica varía en $e^{(0.2269)} = 1.2547$ unidades respecto a otro con título de bachillerato académico.
- Al comparar dos estudiantes que tan sólo difieren en que uno que tomo curso de preparación para el examen el puntaje en Comunicación Escrita varía en $e^{(-36.1158)} = 2.066 \times 10^{-16}$ unidades respecto a otro estudiante que no tomo este curso.

Diagnóstico del modelo óptimo.

Mediante el *plot* y el *worm plot* del modelo, se puede afirmar que la distribución es la adecuada para la variable respuesta y el modelo se ajusta adecuadamente a los datos, lo cual quiere decir que el modelo es adecuado para explicar El puntaje de Comunicación Escrita.

```
> plot(seleccion.ce.a)
*****
                Summary of the Quantile Residuals
                mean      = 0.006405152
                variance  = 0.9999587
                coef. of skewness = -0.1545799
                coef. of kurtosis  = 3.015871
                Filliben correlation coefficient = 0.9964313
                *****
```

Gráfica 5-21. Diagnóstico modelo para el puntaje en Comunicación Escrita - factor académico (plot y el worm plot)



En cuanto al DESEMPEÑO EN INGLÉS se aplicó regresión logística ordinal. a continuación, se presenta el modelo optimo y su interpretación.

Tabla 30. Modelo óptimo para el desempeño en Inglés - factor académico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE	INTERVALO DE CONF 95.00%	
FAMI_COD_EDUCA_PADRE [T. NINGUNO]	-0,2255	0,4594	-0,4909	0,7981	0,6882	0,3237	1,9601
FAMI_COD_EDUCA_PADRE [T. PRIMARIA]	-0,7185	0,2574	-2,7914	0,4875	0,9973	0,2942	0,8070
FAMI_COD_EDUCA_PADRE [T. SECUNDARIA]	-0,2624	0,2745	-0,9559	0,7692	0,8304	0,4490	1,3173
ESTU_TOMOCURSO[T.SI]	-4,8565	37,9338	-0,1280	0,0078	0,5509	NA	1,0480
A- A1	-0,9268	0,2374	-3,9040	0,3958			
A1 A2	0,2328	0,2292	1,0157	1,2621			
A2 B+	1,1732	0,2541	4,6171	3,2323			
B+ B1	1,2350	0,2585	4,7776	3,4384			

Según la razón de probabilidad (OR) y el intervalo de confianza al 95%, es más probable que un estudiante con padre de familia que no realizo ningún nivel educativo tenga valores más bajos en el desempeño de Inglés que un estudiante con padre familia que alcanzó un nivel de educación superior. También se puede

afirmar que, es más probable que aquellos estudiantes con padre familia que alcanzó un nivel educativo de básica primaria o básica secundaria, tengan valores más bajos en el desempeño de Inglés que un estudiante con un padre que alcanzó un nivel de educación superior.

En cuanto al DESEMPEÑO EN COMUNICACIÓN ESCRITA, se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo y su interpretación

Tabla 31. Modelo óptimo para el desempeño en Comunicación Escrita -factor académico

VARIABLE	COEF. ESTIMADO	ERROR	WALD	OR	P-VALUE
2 3	-2,2290	0,2725	-8,1798	0,1076	
3 4	-1,3001	0,1386	-9,3802	0,2725	
4 5	-0,0242	0,1007	-0,2403	0,9761	
5 6	1,1310	0,1279	8,8428	3,0988	
6 7	2,2295	0,2726	8,1787	9,2952	

Ninguna variable académica de las incluidas en este estudio explica el desempeño en Comunicación Escrita.

6. DISCUSIÓN

Según la opinión de los estudiantes universitarios colombianos los factores más influyentes y las razones por las cuales no les va bien en la prueba Saber Pro de acuerdo a las investigaciones de Abuchar y Simanca⁵⁶, son los siguientes.

Las preguntas no se comprenden, salieron temas no vistos, los ejercicios de matemáticas requerían mucha abstracción, la instrucción dada antes (simulacro) no era acorde con las pruebas a las cuales teníamos que enfrentar, en general era sencillo pero no se tenía habilidad de pensamiento, falta mayor enfoque por parte de los docentes, en las universidades no hay tiempo, ni recursos disponibles, para hacer un mayor número de simulacros; en algunas instituciones las preguntas que se realizan se deben a la subjetividad del docente que las prepara sin que necesariamente el conozca cómo se hacen las pruebas, lo cual es importante ya que estas tienen competencias a evaluar como son; interpretativa, argumentativa y propositiva⁵⁷.

Según estos autores es necesario llevar este proceso de preparación de las pruebas Saber Pro, hacia una plataforma informática que brinde el apoyo necesario para optimizar los resultados obtenidos por los estudiantes, brindándoles la posibilidad de una capacitación individual, sin límites geográficos, ni de tiempo y con recursos de fácil manejo y obtención.

⁵⁶ ABUCHAR, A. & SIMANCA, F. (2013). E-Learning en procesos de evaluación académica; pruebas Saber Pro. Universidad Distrital Francisco José de Caldas. Bogotá, Colombia: Revista vínculos.

⁵⁷ *Ibíd.*, p. 362

En esta investigación el objetivo principal fue determinar los factores demográficos, socioeconómicos y académicos que influyen en los resultados (desempeño y puntaje) del componente genérico de la prueba Saber Pro en el programa de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama. A continuación se muestran los factores de **aspecto demográfico** que influyen en los puntajes del componente genérico.

Tabla 32. Factores demográficos que influyen en los puntajes del componente genérico.

VARIABLES DEMOGRÁFICAS	COMPONENTE GENÉRICO			
	P. CE	P. I	P.LC	P.RC
Género(Masculino)	*	*	+	+
Edad	-	*	-	-
Número de personas que conforman el hogar	*	*	-	*
Número de personas de las que se encuentra a cargo	*	*	+	*
Es cabeza de familia (Si)	*	*	*	+
Estado civil (Soltero)	-	*	*	*
Estado civil (Separado)	-	*	*	*
Estado civil (Unión libre)	+	*	*	*

“+” Indica que la variable influye positivamente en dicho puntaje del componente genérico

“-” Indica que la variable influye negativamente en dicho puntaje.

“ * “ Indica que la variable no influye en dicho puntaje.

Según los resultados de los modelos óptimos, ningún factor demográfico influye en los puntajes de Inglés, esto también se puede ver en los resultados de la descripción bivariada sección 4.4, donde se encontró según la prueba de Kruskal Wallis que ninguna variable demográfica está asociada con este puntaje. Es decir, las características demográficas abordadas en este estudio no afectan el puntaje en Inglés en las prueba Saber Pro.

Por otra parte, se puede afirmar que si el estudiante es hombre tanto el puntaje en Razonamiento Cuantitativo como el puntaje en Lectura Crítica son más altos en comparación con los puntajes de las mujeres, siempre y cuando las demás variables consideradas en este estudio son constantes. De igual manera en varias investigaciones se encontró el mismo caso, donde afirman que si el estudiante es hombre su rendimiento en las pruebas Saber Pro es más alto que en las mujeres.

Bahamón y Reyes⁵⁸, realizaron un estudio sobre la caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012, concluyen que tienen mejores resultados en las pruebas Saber Pro los estudiantes solteros ya que dedican más tiempo al estudio en comparación con los estudiantes casados. Para esta investigación según la tabla anterior se encontró que tienen mejores resultados en las pruebas Saber Pro los estudiantes casados en comparación con los estudiantes solteros. En este caso se podría decir que el hecho de ser casado la responsabilidad frente a su formación profesional es mayor respecto a un estudiante soltero ya que existen probablemente variedad de obligaciones a cargo por parte del estudiante casado que lo hacen más dedicado en el estudio.

La edad influye negativamente en los puntajes de Comunicación Escrita, Lectura Crítica y Razonamiento Cuantitativo, quiere decir que por cada año más que cumpla el estudiante, los puntajes de Razonamiento Cuantitativo, Lectura Crítica y Razonamiento Cuantitativo disminuyen, siempre y cuando las demás variables consideradas en este estudio son constantes. Además Bahamón y Reyes⁵⁹ encontraron en su trabajo de investigación que la edad es uno de los factores que

⁵⁸ BAHAMÓN, M., & REYES, R. L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012. *Avances en Psicología Latinoamericana*, 32(3), 459-476.

⁵⁹ *Ibíd.*, p. 37

inciden en el rendimiento de las pruebas Saber Pro, afirman que los puntajes altos se ubican en los estudiantes de 24 años.

Además se observa que por una persona más que el estudiante se encuentre a cargo influye positivamente en el puntaje en Lectura Crítica y además cuando el estudiante es cabeza de familia respecto a uno que no lo es, influye positivamente en el puntaje en Razonamiento Cuantitativo, siempre y cuando las demás variables consideradas en este estudio son constantes. De acuerdo con esto, se puede pensar que a mayor número de necesidades que se encuentre a cargo el estudiante mayor es la responsabilidad en cuando a su rendimiento académico, si el estudiante que presentó la prueba es cabeza de hogar se le atribuye una mayor responsabilidad en su formación profesional. Por otro lado, para Rodríguez⁶⁰ el ser cabeza de familia influye negativamente en el rendimiento académico ya que si el estudiante es cabeza de familia su tiempo para el estudio disminuye y esto a su vez disminuye su rendimiento.

En la tabla 33, se muestran los factores de **aspecto socioeconómico** que influyen en los puntajes del componente genérico.

Tabla 33. Factores socioeconómicos que influyen en los puntajes del componente genérico.

VARIABLES SOCIOECONÓMICAS	COMPONENTE GENÉRICO			
	P. CE	P. I	P.LC	P.RC
Valor anual de la matrícula del año anterior (Entre 500 mil y menos de 1 millón)	*	+	+	*
Valor anual de la matrícula del año anterior (Entre 1 millón y 3 millones)	*	+	*	*
Ocupación del padre (Trabajador independiente)	*	+	*	*
Ocupación del padre (Hogar)	-	*	*	*
Ocupación del padre (Pensionado)	-	+	*	*

⁶⁰ RODRÍGUEZ, Gustavo. Determinantes del desempeño académico universitario. el caso de la Región Caribe Colombiana. ICFES: Saber Pro. Bogotá, D.C., noviembre de 2014.

Continuación tabla anterior

VARIABLES SOCIOECONÓMICAS	P. CE	P. I	P.LC	P.RC
Ocupación de la madre (Trabajadora independiente)	*	-	*	*
Ocupación de la madre (Hogar)	*	*	*	-
Ocupación de la madre (Pensionada)	+	+	*	*
Estrato 2	*	*	*	*
Estrato 3	*	*	*	*
Estrato 4	-	*	*	*
El hogar cuenta con celular (Si)	+	-	*	+
El hogar cuenta con conexión a internet (Si)	+	-	*	*
El hogar cuenta con Servicio cerrado de televisión (Si)	-	-	*	*
El hogar cuenta con servicio de teléfono fijo (Si)	+	*	*	*
El hogar cuenta con lavadora (Si)	-	*	*	*
Ingresos familiares mensuales representado en salarios mínimos mensuales (Entre 1 y Menos de 2 SM)	*	+	*	*
Ingresos familiares mensuales representado en salarios mínimos mensuales (Entre 2 y Menos de 3 SM)	*	-	*	*
Ingresos familiares mensuales representado en salarios mínimos mensuales (Entre 3 y Menos de 7 SM)	-	*	+	*
Trabaja actualmente (No)	-	*	-	*
Trabaja actualmente (Si, por ser práctica obligatoria del plan de estudios)	-	*	-	*
Trabaja actualmente (Si, para adquirir experiencia y/o recursos para sus gastos personales)	-	*	-	*

Si el valor anual de la matrícula del año anterior de un estudiante fue entre más de 500 mil y menos de 3 millones influye positivamente en el puntaje en Inglés respecto a un estudiante que pago menos de 500 mil, es decir que un estudiante que pago una matrícula de más de 500 mil tiene un puntaje más alto respecto a uno que pago menos de un salario mínimo, siempre y cuando las demás variables consideradas en este estudio son constantes. Se puede decir que a mayor inversión mayor es la responsabilidad de estudiante en obtener un alto rendimiento académico.

Cuando un estudiante no trabaja o está realizando la práctica obligatoria del plan de estudios o trabaja para adquirir experiencia y recursos para gastos personales, obtiene valores más bajos en los puntajes de Lectura Crítica y Comunicación

Escrita que uno que trabaja como ayudante sin remuneración, siempre y cuando las demás variables consideradas en este estudio son constantes. Por tanto existe la posibilidad de que un estudiante que trabaja con esfuerzo y sin nada a cambio es mayor la dedicación y la motivación frente a sus intereses y obligaciones en este caso más responsabilidad en cumplir con una buena formación profesional. Por otro lado Osma, Mojica y Rivera⁶¹ afirman que el hecho de que el estudiante tenga o no trabajo no influye en los resultados de la prueba saber Pro. Por el contrario para Rodríguez⁶², si el estudiante trabaja su tiempo para el estudio disminuye y esto a su vez disminuye su rendimiento.

El tener servicio de televisión en el hogar influye negativamente en los puntajes de Comunicación Escrita e Inglés, es decir para aquel estudiante que no cuenta en su hogar con televisión tiene puntajes más altos en Comunicación Escrita e Inglés que uno que si cuenta con este medio, siempre y cuando las demás variables consideradas en este estudio son constantes. Es decir que existe la posibilidad de que este estudiante que no cuenta con TV dedique mucho más tiempo a la escritura, la lectura y a relacionarse con el medio que lo rodea.

El tener servicio de internet en el hogar influye negativamente en los puntajes de Inglés y positivamente en los puntajes de Comunicación Escrita. Según a Arias y Avila⁶³ el acceso a internet es una variable que afecta positivamente al

⁶¹ OSMA, W., MOJICA, A., y RIVERA, T. Factores asociados al rendimiento en las pruebas Saber Pro en estudiantes de Ingeniería Civil en universidades colombianas. Unidad de Santander. Innovaciencia. Facultad de ciencias. Exactas fis. Naturales. 2014; 2(1): 17 – 24.

⁶² RODRÍGUEZ, G. Op. cit., p. 29.

⁶³ ARIAS, I. y AVILA, C. (2014). Influencia de los padres en el rendimiento académico de los hijos: Una aproximación econométrica en el contexto de la educación media colombiana. Universidad Militar Nueva Granada. Bogotá, Colombia: ISSN 2011 – 5318.

rendimiento académico, dado que el poseer este servicio aumenta la probabilidad tener un puntaje alto en la prueba Saber Pro. Igualmente Escobar y Orduz⁶⁴ el tener acceso a internet en el hogar resulta ser un factor importante para el rendimiento académico dado que facilita y agiliza el aprendizaje del estudiante. Además de esto, el uso de internet en las instituciones educativas esta cada día más presente y va más allá de una herramienta tecnológica. Se trata de crear nuevas metodologías, más dinámicas y atractivas en el proceso de desarrollo del conocimiento.

A continuación se muestran los factores de **aspecto académico** más influyentes en los puntajes del componente genérico.

Tabla 34. Factores académicos que influyen en los puntajes del componente genérico.

VARIABLES ACADÉMICAS	COMPONENTE GENÉRICO			
	P.CE	P.I	P.LC	P.RC
Tomó algún curso de preparación para el examen (Si)	+	*	*	*
Máximo nivel educativo alcanzado por el padre (Ninguno)	*	*	*	+
Máximo nivel educativo alcanzado por el padre (Básica primaria)	*	-	*	*
Máximo nivel educativo alcanzado por el padre (Básica secundaria)	*	-	*	+
Máximo nivel educativo alcanzado por la madre (Ninguno)	*	*	+	-
Máximo nivel educativo alcanzado por la madre (Básica primaria)	*	*	+	*
Máximo nivel educativo alcanzado por la madre (Básica secundaria)	*	*	+	*

Al comparar dos estudiantes que tan sólo difieren en que uno cuyo padre alcanzó un nivel educativo de básica primaria o de básica secundaria y otro cuyo padre

⁶⁴ ESCOBAR, S. y ORDUZ, M. Determinantes de la calidad en la educación superior en Colombia. Facultad de Ciencias Económicas y Administrativas. Pontificia Universidad Javeriana. Bogotá, D.C., 2013.

alcanzó un nivel de educación superior, tiene un puntaje en Inglés más alto o le va mejor aquel estudiante cuyo padre de familia alcanzó un nivel de educación superior. Para Arias y Avila⁶⁵ la educación de los padres es considerada como un factor influyente en la determinación del rendimiento académico, así como generalmente a mayor nivel educativo de los padres mayor es el rendimiento académico de sus hijos. De igual forma Osma, Mojica y Rivera⁶⁶ afirman que el tipo de educación del padre, tipo de educación de la madre influyen en los resultados de la prueba Saber Pro.

Escobar y Orduz⁶⁷ también afirman que entre más alto sea el nivel educativo de la madre y del padre mejor será el desempeño en el puntaje de la Prueba Saber Pro del estudiante. El apoyo a los estudiantes desde las primeras etapas de vida es fundamental en el rendimiento académico. Este tipo de acompañamiento es más evidente en madres con mayores niveles de formación.

Para aquel estudiante que tomó algún curso de preparación para la prueba Saber Pro tiene un puntaje más alto en Comunicación Escrita respecto a otro estudiante que no tomó este curso, siempre y cuando las demás variables consideradas en este estudio son constantes. Según esto para tener un buen rendimiento en el puntaje de las pruebas Saber Pro es necesario de un curso de preparación ya que por medio de este se tiene en claro que se va evaluar y como se va evaluar, permite además poner en práctica y ampliar los conocimientos y capacidades.

En el caso de los desempeños, se puede afirmar que la situación del hogar actual es la única variable que explica los desempeños Comunicación Escrita e Inglés. En cuanto al factor socioeconómico el desempeño en Comunicación Escrita es

⁶⁵ *Ibíd.* Op. cit., p. 7

⁶⁶ OSMA, W., MOJICA, A., y RIVERA, T. Op. cit., p. 3

⁶⁷ ESCOBAR, S. y ORDUZ, M. Op. cit., p. 37

explicado según los resultados del modelo óptimo, por la ocupación del padre y el desempeño en Inglés es explicado por el valor de la matrícula y sí o no cuenta con servicio con lavadora. Respecto a factor académico el desempeño en Inglés es explicado por la educación de la madre y del padre del estudiante y sí o no tomo curso de preparación para la prueba.

Teniendo en cuenta algunos estudios investigativos y los resultados de esta investigación, a continuación se proponen, desde la perspectiva del autor, algunas estrategias encaminadas al mejoramiento institucional.

- Se debe diseñar una plataforma virtual gratuita, donde se brinden cursos y talleres de preparación y apoyo para las pruebas Saber Pro, esto con el fin de que el estudiante se capacite y ponga a prueba sus conocimientos.
- Ampliar los cursos y talleres de preparación, junto con capacitaciones, donde se exponga la importancia de la preparación de esta prueba (por qué y para que presentar esta prueba).
- Con la participación de docentes y estudiantes llevar a cabo una discusión, sobre los resultados de las pruebas Saber Pro.
- Incentivar a los estudiantes a que en el tiempo de ocio desarrollen dentro de la Universidad actividades culturales sobre diferentes temáticas donde se genere un ambiente participación y se lleve a cabo un aprendizaje significativo.
- Se propone que haya más financiación respecto a prácticas, seminarios y congresos para estudiantes de bajos recursos.
- Incentivar la participación de los estudiantes a los grupos de investigación que ofrece la Universidad y desarrollo permanente de proyectos de investigación.
- Se propone realizar charlas dentro de la Universidad para los estudiantes, sobre el cómo distribuir el tiempo para el desarrollo de sus actividades y el uso adecuado de medios tecnológicos.
- Se propone que la Universidad ofrezca cursos sobre métodos de estudio.

7. CONCLUSIONES

Los estudiantes del programa de Administración de Empresas Agropecuarias que presentaron la prueba Saber Pro entre el segundo semestre de 2011 y el segundo semestre de 2015 se identifican por tener una edad promedio de 25 años, el 63% son mujeres y el 37% son hombres, 87% son solteros y tan solo el 12% son cabeza de hogar. De los 222 estudiantes que se tuvieron en cuenta para este estudio el 70% tienen que pagar una matrícula de menos de 500 mil pesos, el 55% tienen varias formas de pago, dentro de estas incluyendo la ayuda de los padres, crédito, recursos propios y demás; la mayoría de los padres de familia de aquellos estudiantes que presentaron la prueba son empleados y trabajadores independientes, además mayoritariamente el 65% de las madres son amas de casa; el 27% de los estudiantes tienen ingresos familiares mensuales de menos un (1) SM y el 56% tienen ingresos familiares mensuales entre un (1) SM y menos de dos (2) SM; el 73% de los estudiantes no tomaron curso de preparación para esta prueba. Además, también se puede afirmar que el máximo nivel educativo alcanzado por el padre y la madre del estudiante se encuentra mayoritariamente en básica primaria con un 49% y un 46%.

Los factores demográficos que influyen en los resultados de Razonamiento Cuantitativo de la prueba Saber Pro en el programa de Administración de Empresas Agropecuarias de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, son la edad, el género y el ser cabeza de familia. En Lectura Crítica influyen la edad, el número de personas que conforma el hogar y número de personas de la que el estudiante se encuentra a cargo. En Comunicación Escrita influye solamente la edad. Ningún factor demográfico de los estudiados influye en los puntajes de Inglés.

Los factores socioeconómicos que influyen en los resultados de Comunicación Escrita de la prueba Saber son el valor anual de la matricula del año anterior, la ocupación del padre, la ocupación de la madre, el estrato socioeconómico, tener o no celular, contar o no en el hogar con conexión a internet, televisión, teléfono fijo y lavadora, los ingresos familiares mensuales familiares representado en salarios mínimos mensuales y el sí o no trabaja actualmente. En Inglés influye el valor anual de la matricula del año anterior, la ocupación del padre, la ocupación de la madre, tener o no celular, contar o no en el hogar con conexión a internet, televisión, teléfono fijo y los ingresos familiares mensuales familiares representado en salarios mínimos mensuales. En Lectura Crítica influyen el valor anual de la matricula del año anterior, los ingresos familiares mensuales familiares representado en salarios mínimos mensuales y el sí o no trabaja actualmente. En Razonamiento Cuantitativo influyen la ocupación de la madre y tener o no celular.

Los factores académicos que influyen en los resultados de Razonamiento Cuantitativo son el nivel educativo alcanzado por la madre y máximo nivel educativo alcanzado por el padre. El Lectura Crítica influye el máximo nivel educativo alcanzado por la madre. En Inglés influye el máximo nivel educativo alcanzado por el padre. En Comunicación Escrita Influye el sí o no tomo curso de preparación.

En el caso de los desempeños, en el aspecto demográfico, la situación del hogar actual es la única variable que explica los desempeños Comunicación Escrita e Inglés. En cuanto al factor socioeconómico el desempeño en Comunicación Escrita es explicado, por la ocupación del padre y el desempeño en Inglés es explicado por el valor de la matricula y sí o no cuenta con servicio con lavadora. Respecto a factor académico el desempeño en Inglés es explicado por la educación de la madre y del padre del estudiante y sí o no tomo curso de preparación para la prueba. Se considerada como un factor influyente en la

determinación del rendimiento académico, así como generalmente a mayor nivel educativo de los padres mayor es el rendimiento académico de sus hijos. De igual forma Osma, Mojica y Rivera⁶⁸ afirman que el tipo de educación del padre, tipo de educación de la madre influyen en los resultados de la prueba Saber Pro. De lo anterior se deduce que el nivel educativo de los padres es muy importante en la educación y formación académica del estudiante y de este depende del rendimiento y resultado de sus competencias, que se reflejan en la prueba Saber Pro.

Se propone, desde la perspectiva del autor, algunas estrategias que permitan potenciar las competencias genéricas de los estudiantes del Programa de Administración de Empresas Agropecuarias:

- Diseñar una plataforma virtual gratuita, donde se brinden cursos y talleres de preparación y apoyo para las pruebas Saber Pro, esto con el fin de que el estudiante se capacite y ponga a prueba sus conocimientos.
- Ampliar los cursos y talleres de preparación, junto con capacitaciones, donde se exponga la importancia de la preparación de esta prueba (por qué y para que presentar esta prueba).
- Incentivar a los estudiantes a que en el tiempo de ocio desarrollen dentro de la Universidad actividades culturales sobre diferentes temáticas donde se genere un ambiente participación y se lleve a cabo un aprendizaje significativo.
- Se propone ampliar los puntos de internet con el fin de que los estudiantes puedan acceder y navegar con más facilidad.
- Incentivar la participación de los estudiantes a los grupos de investigación que ofrece la Universidad.

⁶⁸ OSMA, W., MOJICA, A., y RIVERA, T. Op. cit., p. 3

- Se propone realizar charlas dentro de la Universidad para los estudiantes, sobre el cómo distribuir el tiempo para el desarrollo de sus actividades y el uso adecuado de medios tecnológicos.
- Propone que la Universidad ofrezca cursos sobre métodos de estudio.

REFERENCIAS BIBLIOGRÁFICAS

AGRESTI, A. (2002). *Categorical Data Analysis*. University of Florida, Gainesville, Florida. Second edition. Published simultaneously in Canada. ISBN 0-471-36093-7.

ABUCHAR, A. & SIMANCA, F. (2013). E-Learning en procesos de evaluación académica; pruebas Saber Pro. Universidad Distrital Francisco José de Caldas. Bogotá, Colombia. *Revista vínculos*.

BAHAMÓN, M., & REYES, R. L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro - año 2012. *Avances en Psicología Latinoamericana*, 32(3), 459-476.

CAÑADAS, L. (2013). *Regresión logística. Tratamiento computacional con R*. Universidad de Granada. España .2013

DÍAZ, L. & MORALES, M. (2012). *Análisis estadístico de datos categóricos*. Universidad Nacional de Colombia. Primera edición, 2009. Bogotá, Colombia. ISBN 978-958-719-186-8.

EGEA, J. & KESSLER, M. *Regresión lineal con R commander*, Universidad politécnica de Cartagena, departamento de matemática aplicada y estadística.

ESCOBAR, S. & ORDUZ, M. *Determinantes de la calidad en la educación superior en Colombia*. Facultad de Ciencias Económicas y Administrativas. Pontificia Universidad Javeriana. Bogotá, D.C., 2013.

GARZÓN, C. & ALBERTO, J. (2013). Predicción del resultado en la prueba Saber Pro para economía a partir de la información disponible en el proceso de admisión. Repository. Pontificia Universidad Javeriana.

GUZMÁN, C., SERNA, C. & HOYOS, D. (2012). Las pruebas ECAES en Colombia. una evaluación a la evaluación. Bogotá. Panorama.

HUAMANÍ, C., GUTIÉRREZ, C. & MEZONES, E. Correlación y concordancia entre el promedio ponderado universitario y la calificación del examen nacional de medicina total y por áreas, UNT 2008 – 2012. Perú.

ICFES. Exámenes de calidad de la educación superior en administración. Bogotá D.C., noviembre de 2004.

ICFES. Orientaciones para el examen de Estado de calidad de la educación superior.

MINISTERIO DE EDUCACIÓN NACIONAL. Decreto 1075 de 2015. 2015. Disponible en. <http://www.mineduacion.gov.co/1759/w3-article-351080.html>

MINISTERIO DE EDUCACIÓN NACIONAL. Decreto 1095 del 20 de abril de 2010. 2010. Disponible en. http://www.mineduacion.gov.co/1621/articles-229430_archivo_función_de_densidad_de_probabilidad_decreto1295.p

MINISTERIO DE EDUCACIÓN NACIONAL. Decreto 4216 del 30 de octubre de 2009. 2009. Disponible en. http://www.icfes.gov.co/exámenes/component/docman/doc_download/71-decreto-4216-de-2009-modificacion-decreto-3963-reglamentacion-examen-saber-pro [consultado 1 Jul 2012].

MISAS, G. (2004). La educación superior en Colombia. Análisis y estrategias para su desarrollo. Bogotá. Universidad Nacional de Colombia.

OROZCO, T.& JIMÉNEZ, R. (2009). Los ECAES y la cultura de evaluación en la educación superior. Universidad de Santo Tomas. RIIEP.

OSMA, W., MOJICA, A., & RIVERA, T. Factores asociados al rendimiento en las pruebas Saber Pro en estudiantes de Ingeniería Civil en universidades colombianas. Unidad de Santander. Innovaciencia. Facultad de ciencias. Exactas fis. Naturales. 2014; 2(1). 17 – 24.

PÉREZ, M. (2016). Modelos Aditivo Generalizados para Localización, Escala y Forma (GAMLSS). Trabajo para optar al título de máster. Universidad de Vigo. Pontevedra, España. 2016.

RAMÍREZ, C. (2014). Factores asociados al desempeño académico según nivel de formación pregrado y género de los estudiantes de educación superior Colombia. Revista Colombiana. Bogotá, Colombia.

RAMÍREZ, C. & TEICHLER, U. Factores socioeconómicos y educativos asociados con el desempeño académico, según nivel de formación y género de los estudiantes que presentaron la prueba Saber Pro 2009. ICFES. Saber Pro. Bogotá, D.C., Noviembre de 2014.

RODRÍGUEZ, G. Determinantes del desempeño académico universitario. el caso de la Región Caribe Colombiana. ICFES. Saber Pro. Bogotá, D.C., Noviembre de 2014.

SABER PRO (ECAES). Prueba de competencias comunes del área de educación. Bogotá D.C., Marzo de 2011.

STASINOPOULOS, M., RIGBY, B. & AKANTZILIOTOU, C. (2006). Instrucciones de cómo usar el paquete gamlss en R. Segunda edición. Centro de investigación de la Universidad Metropolitana de Londres. Londres, Inglaterra.

ANEXO 1
RESULTADOS ESTADÍSTICOS
ANÁLISIS BIVARIADO

RESULTADOS ESTADÍSTICOS ANÁLISIS BIVARIADO

A continuación se presentan los resultados del coeficiente de correlación de Spearman, el cual mide la correlación (la asociación o interdependencia) entre dos variables continuas.

Los siguientes resultados del coeficiente de correlación de Spearman obtenidos por el R estadístico, muestran si existe o no correlación entre los puntajes del componente genérico y algunas variables que identifican al estudiante, a un nivel de significancia del 5%.

Edad vs Puntaje en Razonamiento Cuantitativo

```
Spearman's rank correlation rho
data:  EDAD and MOD_RAZONA_CUANTITATIVO_PUNT
S = 2188779, p-value = 0.002713
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.2003364
```

Número de personas de las que se encuentra a cargo vs Puntaje en Razonamiento Cuantitativo

```
Spearman's rank correlation rho
data:  FAMI_NUM_PERS_CARGO and MOD_RAZONA_CUANTITATIVO_PUNT
S = 1844047, p-value = 0.7109
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.02507433
```

Número de personas que conforman el hogar vs Puntaje en Razonamiento Cuantitativo

```
Spearman's rank correlation rho
data:  FAMI_NUM_PERS_GRUP_FAM and MOD_RAZONA_CUANTITATIVO_PUNT
S = 1802863, p-value = 0.9743
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.002180801
```

Número de dormitorios de la residencia vs Puntaje en Razonamiento Cuantitativo

```
Spearman's rank correlation rho

data:  ESTU_DORMITORIOS and MOD_RAZONA_CUANTITATIVO_PUNT
S = 1828086, p-value = 0.8107
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.01620168
```

Según los resultados anteriores de la prueba de Spearman de correlación, se puede afirmar que, la edad del estudiante, el número de personas de las que se encuentra a cargo el estudiante, el número de personas que conforman el hogar, el número de dormitorios de la residencia, tienen una correlación negativa con el puntaje en Razonamiento Cuantitativo. Esto quiere decir que las dos variables se correlacionan en sentido inverso. A valores altos de una de ellas le suelen corresponder valores bajos de la otra o viceversa (Bárceñas, 2014).

Edad vs Puntaje en Inglés

```
Spearman's rank correlation rho

data:  EDAD and MOD_INGLES_PUNT
S = 2053015, p-value = 0.06114
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1258831
```

Número de personas de las que se encuentra a cargo vs Puntaje en Inglés

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_CARGO and MOD_INGLES_PUNT
S = 1873935, p-value = 0.5376
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.04168871
```

Número de personas que conforman el hogar vs Puntaje en Inglés

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_GRUP_FAM and MOD_INGLES_PUNT
S = 1757616, p-value = 0.7342
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.02297114
```

Número de dormitorios de la residencia vs Puntaje en Inglés

```
Spearman's rank correlation rho

data:  ESTU_DORMITORIOS and MOD_INGLES_PUNT
S = 1590460, p-value = 0.08564
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1158905
```

Según los resultados anteriores de la prueba de Spearman de correlación, se puede afirmar que, el número de personas que conforman el hogar y el número de dormitorios de la residencia tienen una correlación positiva con el puntaje en Inglés. Esto quiere decir que las dos variables se correlacionan en sentido directo. A valores altos de una de ellas le suelen corresponder valores altos de la otra e igualmente que con los valores bajos (Bárceñas, 2014).

Edad vs Lectura Crítica

```
Spearman's rank correlation rho

data:  EDAD and MOD_LECTURA_CRITICA
S = 2115289, p-value = 0.01701
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1600341
```

Número de personas de las que se encuentra a cargo vs Lectura Crítica

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_CARGO and MOD_LECTURA_CRITICA
S = 1876382, p-value = 0.5244
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.04304853
```

Número de personas que conforman el hogar vs Lectura Crítica

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_GRUP_FAM and MOD_LECTURA_CRITICA
S = 1800311, p-value = 0.991
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.000761924
```

Número de dormitorios de la residencia vs Lectura Crítica

```
Spearman's rank correlation rho

data:  ESTU_DORMITORIOS and MOD_LECTURA_CRITICA
S = 1734081, p-value = 0.594
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.03605404
```

Según los resultados anteriores de la prueba de Spearman de correlación, se puede afirmar que, el número de dormitorios de la residencia tienen una correlación positiva con el puntaje en Lectura Crítica y las demás tienen una correlación negativa con este puntaje.

COMUNICACIÓN ESCRITA

Edad vs Comunicación Escrita

```
Spearman's rank correlation rho

data:  EDAD and MOD_COMUNICA_ESCRITA_PUNT
S = 2098002, p-value = 0.006727
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1822192
```

Número de personas de las que se encuentra a cargo vs Comunicación Escrita

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_CARGO and MOD_COMUNICA_ESCRITA_PUNT
S = 1892180, p-value = 0.2331
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.08091214
```

Número de personas que conforman el hogar vs Comunicación Escrita

```
Spearman's rank correlation rho

data:  FAMI_NUM_PERS_GRUP_FAM and MOD_COMUNICA_ESCRITA_PUNT
S = 1764223, p-value = 0.9084
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.00781616
```

Número de dormitorios de la residencia vs Comunicación Escrita

```
Spearman's rank correlation rho

data:  ESTU_DORMITORIOS and MOD_COMUNICA_ESCRITA_PUNT
S = 1671372, p-value = 0.5056
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.04522476
```

Según los resultados anteriores de la prueba de Spearman de correlación, solo el número de dormitorios tiene una correlación positiva con el puntaje en Comunicación Escrita.

A continuación se presentan pruebas de independencia, las cuales muestran si existe o no dependencia entre algunas variables con los desempeños de Inglés y Comunicación Escrita.

PRUEBAS DE INDEPENDENCIA CHI-CUADRADO DESEMPEÑO EN INGLÉS

Desempeño en Inglés vs Género

```

Frequency table:
      DESEMPEÑO EN INGLES
GENERO A-  A1  A2  B+  B1
  F  47  61  21   1   9
  M  28  34  18   0   3

Pearson's Chi-squared test

data:  .Table
X-squared = 2.7678, df = 4, p-value = 0.5974

```

Desempeño en Inglés vs Estado civil

```

Frequency table:
      DESEMPEÑO EN INGLES
ESTADO CIVIL A-  A1  A2  B+  B1
Casado          4   2   1   0   0
Soltero         67  84  34   1  12
Separado        0   1   0   0   0
Unión libre     4   6   4   0   0

Pearson's Chi-squared test

data:  .Table
X-squared = 5.2939, df = 12, p-value = 0.9474

```


Desempeño en Inglés vs Es cabeza de familia

Frequency table:

		DESEMPEÑO EN INGLES				
ES CABEZA FAMILIA	A-	A1	A2	B+	B1	
NO	65	83	35	1	11	
SI	10	11	4	0	1	

Pearson's Chi-squared test

data: .Table

X-squared = 0.533, df = 4, p-value = 0.9702

Desempeño en Inglés vs Situación del hogar actual

Frequency table:

		DESEMPEÑO EN INGLES				
SITUACIÓN DEL HOGAR DEL ACTUAL	A-	A1	A2	B+	B1	
Habitual o permanente	53	70	30	1	10	
Temporal	22	24	9	0	2	

Pearson's Chi-squared test

data: .Table

X-squared = 1.5146, df = 4, p-value = 0.824

Desempeño en Inglés vs Estrato

Frequency table:

		DESEMPEÑO EN INGLES				
ESTRATO	A-	A1	A2	B+	B1	
Nivel 1	10	11	5	0	0	
Nivel 2	55	64	27	1	10	
Nivel 3	10	16	6	0	2	
Nivel 4	0	2	1	0	0	

Pearson's Chi-squared test

data: .Table

X-squared = 4.8243, df = 12, p-value = 0.9636

Desempeño en Inglés vs Financiación de la matrícula

```
Frequency table:
                                DESEMPEÑO EN INGLES
FORMA PAGO MATRICULA  A-  A1  A2  B+  B1
Beca                    1   1   2   0   1
Padres                   7  13   2   0   1
Recursos propios        21  25  14   0   2
Varias formas pago     43  52  20   1   7

                                Pearson's Chi-squared test

data: .Table
X-squared = 8.9434, df = 12, p-value = 0.7078
```

Desempeño en Inglés vs Ocupación del padre

```
Frequency table:
                                DESEMPEÑO EN INGLES
OCUPACIÓN DEL PADRE  A-  A1  A2  B+  B1
EMPLEADO.D.C         38  27  16   0   3
HOGAR                 2   2   1   0   0
PENSIONADO.Y.A       6  21  13   0   4
TRABAJADOR.I. (PI.E) 29  44   9   1   5

                                Pearson's Chi-squared test

data: .Table
X-squared = 21.9536, df = 12, p-value = 0.03804
```

Desempeño en Inglés vs Ocupación de la madre

```
Frequency table:
                                DESEMPEÑO EN INGLES
OCUPACIÓN DE LA MADRE  A-  A1  A2  B+  B1
EMPLEADO.D.C           8  22   8   0   0
HOGAR                  55  59  22   0   9
PENSIONADO.Y.A         2   4   5   0   0
TRABAJADOR.I. (PI.E)  10   9   4   1   3

                                Pearson's Chi-squared test

data: .Table
X-squared = 23.5889, df = 12, p-value = 0.02312
```

Desempeño en Inglés vs El hogar cuenta con celular

```
Frequency table:
                DESEMPEÑO EN INGLES
TIENE CELULAR A- A1 A2 B+ B1
NO           7  6  6  0  2
SI          68 88 33  1 10

                Pearson's Chi-squared test

data: .Table
X-squared = 3.4566, df = 4, p-value = 0.4845
```

Desempeño en Inglés vs El hogar cuenta con conexión a internet

```
Frequency table:
                DESEMPEÑO EN INGLES
TIENE INTERNET A- A1 A2 B+ B1
NO           56 67 24  0 10
SI           19 27 15  1  2

                Pearson's Chi-squared test

data: .Table
X-squared = 5.5281, df = 4, p-value = 0.2373
```

Desempeño en Inglés vs El hogar cuenta con servicio cerrado de televisión

```
Frequency table:
                DESEMPEÑO EN INGLES
TIENE SERVICIO TV A- A1 A2 B+ B1
NO           44 51 17  0  6
SI           31 43 22  1  6

                Pearson's Chi-squared test

data: .Table
X-squared = 3.5736, df = 4, p-value = 0.4668
```

Desempeño en Inglés vs El hogar cuenta con servicio de teléfono fijo

Frequency table:

	DESEMPEÑO EN INGLES				
TIENE TELÉFONO	A-	A1	A2	B+	B1
NO	65	81	33	1	9
SI	10	13	6	0	3

Pearson's Chi-squared test

data: .Table

X-squared = 1.3793, df = 4, p-value = 0.8478

Desempeño en Inglés vs El hogar cuenta con lavadora

Frequency table:

	DESEMPEÑO EN INGLES				
TIENE LAVADORA	A-	A1	A2	B+	B1
NO	42	40	10	1	4
SI	33	54	29	0	8

Pearson's Chi-squared test

data: .Table

X-squared = 11.6298, df = 4, p-value = 0.02033

Desempeño en Inglés vs Tiene computador en su hogar

Frequency table:

	DESEMPEÑO EN INGLES				
TIENE COMPUTADOR	A-	A1	A2	B+	B1
NO	20	28	4	0	2
SI	55	66	35	1	10

Pearson's Chi-squared test

data: .Table

X-squared = 6.6225, df = 4, p-value = 0.1572

Desempeño en Inglés vs Ingresos familiares mensuales representado en salarios mínimos mensuales

```

Frequency table:
                DESEMPEÑO EN INGLES
INGRESOS MENSUALES      A-  A1  A2  B+  B1
Menos de 1 SM          28  21  9   0   2
Entre 1 y Menos de 2 SM 36  61  18  1   8
Entre 2 y Menos de 3 SM  7   9  10  0   0
Entre 3 y Menos de 7 SM  4   3   2  0   2

                Pearson's Chi-squared test

data: .Table
X-squared = 20.7435, df = 12, p-value = 0.05427

```

Desempeño en Inglés vs Trabaja actualmente

```

Frequency table:
                DESEMPEÑO EN INGLES
TRABAJA ACTUALMENTE      A-  A1  A2  B+  B1
Ayudante sin remuneración 42  35  15  0   5
Experiencia y gastos personales 8  12  5   0   0
No                          24  47  19  1   7
Práctica obligatoria       1   0   0  0   0

                Pearson's Chi-squared test

data: .Table
X-squared = 12.5196, df = 12, p-value = 0.4049

```

Desempeño en Inglés vs Tomó algún curso de preparación para el examen

```

Frequency table:
                DESEMPEÑO EN INGLES
TOMO CURSO A-  A1  A2  B+  B1
NO          52  68  31  1   9
SI          1   0   0  0   0

                Pearson's Chi-squared test

data: .Table
X-squared = 2.0694, df = 4, p-value = 0.723

```

Desempeño en Inglés vs Máximo nivel educativo alcanzado por el padre

```

Frequency table:
                                DESEMPEÑO EN INGLES
EDUCACION DEL PADRE           A- A1 A2 B+ B1
Nivel de educación superior  11 14 12  0  3
Ninguno                       2  5  2  0  0
Primaria                       45 45 17  0  2
Secundaria                     17 30  8  1  7

Pearson's Chi-squared test

data: .Table
X-squared = 19.1505, df = 12, p-value = 0.08496
    
```

Desempeño en Inglés vs Máximo nivel educativo alcanzado por la madre

```

Frequency table:
                                DESEMPEÑO EN INGLES
EDUCACION DEL MADRE           A- A1 A2 B+ B1
Nivel de educación superior   8 15  9  0  0
Ninguno                        2  1  0  0  0
Primaria                       45 40 12  1  3
Secundaria                     20 38 18  0  9

Pearson's Chi-squared test

data: .Table
X-squared = 21.6699, df = 12, p-value = 0.04139
    
```

Desempeño en Inglés vs Valor anual de la matricula del año anterior

```

Frequency table:
                                DESEMPEÑO EN INGLES
VLR.MATRICULA.ANT             A- A1 A2 B+ B1
No pago matricula             0  3  3  1  0  1
Menos de 500 mil              1 53 64 30  1  7
Entre 500 mil y Menos de 1 millón 2 19 24  7  0  1
Entre 1 millón y 3 millones   3  0  3  1  0  3

Pearson's Chi-squared test

data: .Table
X-squared = 24.3788, df = 12, p-value = 0.01806
    
```

DESEMPEÑO EN COMUNICACIÓN ESCRITA

Desempeño en Comunicación Escrita vs Género

```
Frequency table:
DESEMPEÑO EN COMUNICACIÓN ESCRITA          GENERO
                                                F  M
    Nivel 1                                0  0
    Nivel 2                                 1  1
    Nivel 3                                12  8
    Nivel 4                                54 33
    Nivel 5                                50 28
    Nivel 6                                 18 10
    Nivel 7                                 3  2
    Nivel 8                                 0  0

    Pearson's Chi-squared test

data: .Table
X-squared = 0.3264, df = 5, p-value = 0.9971
```

Desempeño en Comunicación Escrita vs Estado civil

```
Frequency table:
DESEM.COMUNICA.ESCRITA          ESTADO CIVIL
                                Casado Soltero Separado Unión.Libre
    Nivel 1                      0      0      0      0
    Nivel 2                      0      2      0      0
    Nivel 3                      1     19      0      0
    Nivel 4                      2     77      1      5
    Nivel 5                      3     69      0      6
    Nivel 6                      1     24      0      3
    Nivel 7                      0      5      0      0
    Nivel 8                      0      0      0      0

    Pearson's Chi-squared test

data: .Table
X-squared = 5.2791, df = 15, p-value = 0.9895
```

Desempeño en Comunicación Escrita vs Es cabeza de familia

```

Frequency table:
                ES CABEZA DE FAMILIA
DESEMPEÑO EN COMUNICACIÓN ESCRITA NO SI
Nivel 1         0 0
Nivel 2         2 0
Nivel 3        17 3
Nivel 4        75 11
Nivel 5        67 11
Nivel 6        27 1
Nivel 7         5 0
Nivel 8         0 0

Pearson's Chi-squared test

data: .Table
X-squared = 3.4142, df = 5, p-value = 0.6364

```

Desempeño en Comunicación Escrita vs Situación del hogar actual

```

Frequency table:
                ESTU_HOGAR_ACTUAL
MOD_COMUNICA_ESCRITA_DESEM Hb o Pr Temporal
Nivel 1                     0      0
Nivel 2                     2      0
Nivel 3                    15      5
Nivel 4                    67     19
Nivel 5                    56     22
Nivel 6                    19      9
Nivel 7                     3      2
Nivel 8                     0      0

Pearson's Chi-squared test

data: .Table
X-squared = 2.6491, df = 5, p-value = 0.7539

```

Desempeño en Comunicación Escrita vs Estrato

```

Frequency table:
                ESTRATO
DESEMPEÑO.COMUNI.ESCRITA Nivel 1 Nivel 2 Nivel 3 Nivel 4
Nivel 1                   0      0      0      0
Nivel 2                   0      2      0      0
Nivel 3                   1     15      4      0
Nivel 4                   9     62     12      2
Nivel 5                  10     51     16      1
Nivel 6                   5     22      1      0
Nivel 7                   1      4      0      0
Nivel 8                   0      0      0      0

Pearson's Chi-squared test

data: .Table
X-squared = 10.0064, df = 15, p-value = 0.8193

```

Desempeño en Comunicación Escrita vs Financiación de la matrícula

Frequency table:

COMUNICA_ESCRITA_DESEM	FORMA_PAGO_MATRICULA			
	Beca	Padres	Recursos.P	VFP
Nivel 1	0	0	0	0
Nivel 2	0	0	1	1
Nivel 3	0	5	2	13
Nivel 4	2	10	24	47
Nivel 5	2	3	26	45
Nivel 6	1	4	7	13
Nivel 7	0	1	1	3
Nivel 8	0	0	0	0

Pearson's Chi-squared test

data: .Table

X-squared = 13.316, df = 15, p-value = 0.5779

Desempeño en Comunicación Escrita vs Ocupación del padre

Frequency table:

FAMI_COD_OCUP_PADRE	MOD_COMUNICA_ESCRITA_DESEM							
	N1	N2	N3	N4	N5	N6	N7	N8
EMPLEADO.D.C	0	0	9	35	27	12	0	0
HOGAR	0	1	1	1	1	0	1	0
PENSIONADO.Y.A	0	1	2	19	17	3	1	0
TRABAJADOR.I.(PI.E)	0	0	8	31	33	13	3	0

Pearson's Chi-squared test

data: .Table

X-squared = 37.9091, df = 15, p-value = 0.0009306

Desempeño en Comunicación Escrita vs Ocupación del madre

Frequency table:

OCUP.MADRE	DESM.COMUNICA.ESCRITA							
	N1	N2	N3	N4	N5	N6	N7	N8
EMPLEADO.D.C	0	0	3	15	11	7	1	0
HOGAR	0	2	14	58	52	17	2	0
PENSIONADO.Y.A	0	0	2	2	5	1	0	0
TRABAJADOR.I.(PI.E)	0	0	1	11	10	3	2	0

Pearson's Chi-squared test

data: .Table

X-squared = 10.4424, df = 15, p-value = 0.791

Desempeño en Comunicación Escrita vs El hogar cuenta con celular

Frequency table:

DESEM.COMUN.ESCRITA	TIENE CELULAR	
	NO	SI
Nivel 1	0	0
Nivel 2	0	2
Nivel 3	1	19
Nivel 4	11	75
Nivel 5	7	71
Nivel 6	2	26
Nivel 7	0	5
Nivel 8	0	0

Pearson's Chi-squared test

data: .Table

X-squared = 2.4723, df = 5, p-value = 0.7807

Desempeño en Comunicación Escrita vs El hogar cuenta con conexión a internet

Frequency table:

DESEM.COMUNICA.ESCRITA	TIENE SERVICIO DE INTERNET	
	NO	SI
Nivel 1	0	0
Nivel 2	2	0
Nivel 3	15	5
Nivel 4	68	18
Nivel 5	48	30
Nivel 6	19	9
Nivel 7	5	0
Nivel 8	0	0

Pearson's Chi-squared test

data: .Table

X-squared = 9.3431, df = 5, p-value = 0.09614

Desempeño en Comunicación Escrita vs El hogar cuenta con servicio cerrado de televisión

Frequency table:

DESEM.COMUNICA.ESCRITA	TIENE SERVICIO TV	
	NO	SI
Nivel 1	0	0
Nivel 2	2	0
Nivel 3	8	12
Nivel 4	49	37
Nivel 5	41	37
Nivel 6	14	14
Nivel 7	4	1
Nivel 8	0	0

Pearson's Chi-squared test

data: .Table

X-squared = 5.1912, df = 5, p-value = 0.393

Desempeño en Comunicación Escrita vs El hogar cuenta con servicio de teléfono fijo.

```

Frequency table:
                TIENE SERVICIO DE TELEFONÍA
DESEM.COMUNICA.ESCRITA  NO SI
Nivel 1                0  0
Nivel 2                 2  0
Nivel 3                15  5
Nivel 4                76 10
Nivel 5                68 10
Nivel 6                21  7
Nivel 7                 5  0
Nivel 8                 0  0

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 6.1638, df = 5, p-value = 0.2906

```

Desempeño en Comunicación Escrita vs El hogar cuenta con lavadora

```

Frequency table:
                TIENE SERVICIO DE LAVADORA
DESEMPEÑO EN COMUNICACIÓN ESCRITA  NO SI
Nivel 1                            0  0
Nivel 2                             1  1
Nivel 3                            10 10
Nivel 4                             37 49
Nivel 5                             35 43
Nivel 6                             11 17
Nivel 7                              3  2
Nivel 8                              0  0

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 1.1416, df = 5, p-value = 0.9504

```

Desempeño en Comunicación Escrita vs Tiene computador en su hogar

```

Frequency table:
                TIENE COMPUTADOR
DESEMPEÑO EN COMUNICACIÓN ESCRITA  NO SI
Nivel 1                            0  0
Nivel 2                             0  2
Nivel 3                             5 15
Nivel 4                             25 61
Nivel 5                             19 59
Nivel 6                              5 23
Nivel 7                              0  5
Nivel 8                              0  0

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 3.8941, df = 5, p-value = 0.5648

```

Desempeño en Comunicación Escrita vs Ingresos familiares mensuales representado en salarios mínimos

```

Frequency table:
                DESEMPEÑO EN COMUNICACIÓN ESCRITA
INGRESOS MENSUALES      N1  N2  N3  N4  N5  N6  N7
Menos de 1 SM           0   1   7  21  24   6   1
Entre 1 y Menos de 2 SM 0   1   9  51  41  17   4
Entre 2 y Menos de 3 SM 0   0   3   7  12   4   0
Entre 3 y Menos de 7 SM 0   0   1   7   1   1   0

Pearson's Chi-squared test

data: .Table
X-squared = 10.8471, df = 15, p-value = 0.7634

```

Desempeño en Comunicación Escrita vs Trabaja actualmente

```

Frequency table:
                DESEMPEÑO COMUNICACION ESCRITA
TRABAJA              N1  N2  N3  N4  N5  N6  N7
Ayudante sin remuneración  0  1  4  44  35  9  3
Experiencia y gastos personales 0  1  1  5  12  4  1
No                          0  0  14  37  31  15  1
Práctica obligatoria        0  1  0  0  0  0  0

Pearson's Chi-squared test

data: .Table
X-squared = 27.7701, df = 15, p-value = 0.02305

```

Desempeño en Comunicación Escrita vs Tomó algún curso de preparación para el examen

```

Frequency table:
                TOMO CURSO
DESEMPEÑO EN COMUNICACIÓN ESCRITA NO SI
Nivel 1         0  0
Nivel 2         2  0
Nivel 3        13  0
Nivel 4        64  0
Nivel 5        60  1
Nivel 6        19  0
Nivel 7         2  0
Nivel 8         0  0

Pearson's Chi-squared test

data: .Table
X-squared = 1.6496, df = 5, p-value = 0.8952

```

Desempeño en Comunicación Escrita vs Máximo nivel educativo alcanzado por el padre

```

Frequency table:
                                EDUCACIÓN DEL PADRE
DESEM.COMUNICA.ESCRITA Edu.Superior Ninguno Primaria Secundaria
    Nivel 1                0         0         0         0
    Nivel 2                0         0         2         0
    Nivel 3                4         1         8         7
    Nivel 4               17         5        44        20
    Nivel 5               11         1        39        27
    Nivel 6                7         2        10         9
    Nivel 7                0         0         5         0
    Nivel 8                0         0         0         0
    Pearson's Chi-squared test

data: .Table
X-squared = 15.3248, df = 15, p-value = 0.4283

```

Desempeño en Comunicación Escrita vs Máximo nivel educativo alcanzado por la madre

```

Frequency table:
                                EDUCACIÓN DE LA MADRE
DESEM.COMUNICA.ESCRITA Edu.Superior Ninguno Primaria Secundaria
    Nivel 1                0         0         0         0
    Nivel 2                0         0         1         1
    Nivel 3                2         0         8        10
    Nivel 4               15         2        43        26
    Nivel 5               10         0        32        36
    Nivel 6                3         1        13        11
    Nivel 7                1         0         3         1
    Nivel 8                0         0         0         0
    Pearson's Chi-squared test

data: .Table
X-squared = 9.6191, df = 15, p-value = 0.843

```

Desempeño en Comunicación Escrita vs anual de la matricula del año anterior

```

Frequency table:
                                DESEMPEÑO EN COMUNICACIÓN ESCRITA
VALOR DE LA MATRICULA ANTERIOR  N1 N2 N3 N4 N5 N6 N7 N8
No pago matricula              0 0 0 3 2 3 0 0
Menos de 500 mil               0 2 18 57 54 19 4 0
Entre 500 mil y Menos de 1 millón 0 0 2 22 20 5 1 0
Entre 1 millón y 3 millones    0 0 0 4 2 1 0 0
    Pearson's Chi-squared test

data: .Table
X-squared = 11.0403, df = 15, p-value = 0.7497

```

A continuación se muestran pruebas de normalidad para los puntajes del componente genérico.

PRUEBAS DE NORMALIDAD

PUNTAJE RAZONAMIENTO CUANTITATIVO

```
Shapiro-Wilk normality test  
data: MOD_RAZONA_CUANTITATIVO_PUNT  
W = 0.974, p-value = 0.0004045
```

PUNTAJE EN INGLÉS

```
Shapiro-Wilk normality test  
data: MOD_INGLES_PUNT  
W = 0.9387, p-value = 4.903e-08
```

LECTURA CRÍTICA

```
Shapiro-Wilk normality test  
data: MOD_LECTURA_CRITICA  
W = 0.9838, p-value = 0.01208
```

COMUNICACIÓN ESCRITA

```
Shapiro-Wilk normality test  
data: MOD_COMUNICA_ESCRITA_PUNT  
W = 0.9936, p-value = 0.4594
```

Según los resultados de la prueba de normalidad de Shapiro Wilk, el puntaje en Comunicación Escrita es la única variable de respuesta que sigue una distribución

normal ya que su p-valor es mayor que cualquier nivel de significancia considerado.

PRUEBAS DE KRUSCAL WALLIS –TEST NO PARAMÉTRICOS

Como ninguno de los puntajes a excepción del puntaje en Comunicación Escrita, siguen una distribución normal se realizó el test no paramétrico de Kruskal Wallis, con el fin de ver si dos variables (una independiente y otra dependiente) están relacionadas o tienen alguna asociación, en base a si las medianas de la variable dependiente (puntajes componente genérico) son diferentes en las categorías o grupos de la variable independiente (variable cualitativa del estudiante). Es decir, señala si las medianas entre dos o más grupos son similares o diferentes.

Puntaje en Razonamiento Cuantitativo vs El hogar cuenta con celular

```
> with(spl, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_CELULAR, median,
+ na.rm=TRUE))
NO SI
9.2 9.7

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_CELULAR, data=spl)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_CELULAR
Kruskal-Wallis chi-squared = 2.5379, df = 1, p-value = 0.1111
```

Puntaje en Razonamiento Cuantitativo vs Tiene computador en su hogar

```
> with(spl, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_COMPUTADOR, media:
+ na.rm=TRUE))
NO SI
9.7 9.7

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_COMPUTADOR, data=spl)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_COMPUTADOR
Kruskal-Wallis chi-squared = 0.1434, df = 1, p-value = 0.7049
```

Puntaje en Razonamiento Cuantitativo vs El hogar cuenta con conexión a internet

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_INTERNET, median,
+ na.rm=TRUE))
NO SI
9.7 9.8

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_INTERNET, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_INTERNET
Kruskal-Wallis chi-squared = 0.1718, df = 1, p-value = 0.6785

```

Puntaje en Razonamiento Cuantitativo vs Género

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_GENERO, median,
+ na.rm=TRUE))
F M
9.7 9.8

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_GENERO, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_GENERO
Kruskal-Wallis chi-squared = 4.3141, df = 1, p-value = 0.0378

```

Puntaje en Razonamiento Cuantitativo vs Situación del hogar actual

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_HOGAR_ACTUAL, median,
+ na.rm=TRUE))
HB.O.PR TEMPORAL
9.7 9.6

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_HOGAR_ACTUAL, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_HOGAR_ACTUAL
Kruskal-Wallis chi-squared = 0.0061, df = 1, p-value = 0.9376

```

Puntaje en Razonamiento Cuantitativo vs Es cabeza de familia

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_SN_CABEZA_FMLIA,
+ median, na.rm=TRUE))
NO SI
9.70 9.65

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_SN_CABEZA_FMLIA, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_SN_CABEZA_FMLIA
Kruskal-Wallis chi-squared = 5e-04, df = 1, p-value = 0.983

```

Puntaje en Razonamiento Cuantitativo vs Título


```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_TITULO_BTO, median,
+   na.rm=TRUE))
  A   N   T
9.65 9.70 9.80

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_TITULO_BTO, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_TITULO_BTO
Kruskal-Wallis chi-squared = 0.6527, df = 2, p-value = 0.7215

```

Puntaje en Razonamiento Cuantitativo vs Tomó algún curso de preparación para el examen

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_TOMOCURSO, median,
+   na.rm=TRUE))
  NO  SI
9.7  9.0

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_TOMOCURSO, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_TOMOCURSO
Kruskal-Wallis chi-squared = 1.6238, df = 1, p-value = 0.2026

```

Puntaje en Razonamiento Cuantitativo vs Trabaja actualmente

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_TRABAJA, median, na.rm=TRUE))
  AYUDANTE.SIN.R EXPERIENCIA.GASTOS.P      NO      PRACTICA.O
  9.7             9.9             9.7             9.3

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_TRABAJA, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_TRABAJA
Kruskal-Wallis chi-squared = 1.5514, df = 3, p-value = 0.6705

```

Puntaje en Razonamiento Cuantitativo vs El hogar cuenta con lavadora

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_LAVADORA, median,
+   na.rm=TRUE))
  NO  SI
9.7  9.7

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_LAVADORA, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_LAVADORA
Kruskal-Wallis chi-squared = 0.433, df = 1, p-value = 0.5105

```

Puntaje en Razonamiento Cuantitativo vs El hogar cuenta con Servicio cerrado de televisión

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_SERVICIO_TV, median,
+ na.rm=TRUE))
  NO  SI
9.75 9.70

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_SERVICIO_TV, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_SERVICIO_TV
Kruskal-Wallis chi-squared = 1.1214, df = 1, p-value = 0.2896

```

Puntaje en Razonamiento Cuantitativo vs El hogar cuenta con servicio de teléfono fijo

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ECON_SN_TELEFONIA, median,
+ na.rm=TRUE))
  NO  SI
9.7  9.8

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ECON_SN_TELEFONIA, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ECON_SN_TELEFONIA
Kruskal-Wallis chi-squared = 0.0887, df = 1, p-value = 0.7658

```

Puntaje en Razonamiento Cuantitativo vs Estado civil

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_ESTADO_CIVIL, median,
+ na.rm=TRUE))
  C  SLT  SP  UL
9.2 9.7 9.8 9.6

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_ESTADO_CIVIL, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_ESTADO_CIVIL
Kruskal-Wallis chi-squared = 3.0014, df = 3, p-value = 0.3914

```

Puntaje en Razonamiento Cuantitativo vs Educación de la madre

```

> with(sp1, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, FAMI_COD_EDUCA_MADRE,
+ median, na.rm=TRUE))
EDU.SUPERIOR  NINGUNO  PRIMARIA  SECUNDARIA
      9.85      10.20      9.50      9.70

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ FAMI_COD_EDUCA_MADRE,
+ data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by FAMI_COD_EDUCA_MADRE
Kruskal-Wallis chi-squared = 5.0347, df = 3, p-value = 0.1693

```

Puntaje en Razonamiento Cuantitativo vs Ocupación de la madre

```

> with(spl, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, FAMI_COD_OCUP_MADRE,
+ median, na.rm=TRUE))
      EMPLEADO.D.C          HOGAR          PENSIONADO.Y.A
TRABAJADOR.I.(PI.E)
      9.9              9.6              9.4
      9.7

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ FAMI_COD_OCUP_MADRE, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by FAMI_COD_OCUP_MADRE
Kruskal-Wallis chi-squared = 6.1148, df = 3, p-value = 0.1062

```

Puntaje en Razonamiento Cuantitativo vs Ocupación del padre

```

> with(spl, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, FAMI_COD_OCUP_PADRE,
+ median, na.rm=TRUE))
      EMPLEADO.D.C          HOGAR          PENSIONADO.Y.A
TRABAJADOR.I.(PI.E)
      9.65              9.70              9.75
      9.70

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ FAMI_COD_OCUP_PADRE, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by FAMI_COD_OCUP_PADRE
Kruskal-Wallis chi-squared = 1.6247, df = 3, p-value = 0.6538

```

Puntaje en Razonamiento Cuantitativo vs Forma de pago de la matricula

```

> with(spl, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, FORMA_PAGO_MATRICULA,
+ median, na.rm=TRUE))
      B  PD  RP  VFP
10.5  9.7  9.7  9.7

> kruskal.test(MOD_RAZONA_CUANTITATIVO_PUNT ~ FORMA_PAGO_MATRICULA,
+ data=spl)

Kruskal-Wallis rank sum test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by FORMA_PAGO_MATRICULA
Kruskal-Wallis chi-squared = 1.7555, df = 3, p-value = 0.6247

```

PUNTAJE EN INGLÉS

Puntaje en Inglés vs El hogar cuenta con celular

```

> with(spl, tapply(MOD_INGLES_PUNT, ECON_SN_CELULAR, median, na.rm=TRUE))
      NO  SI
9.8  9.6

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_CELULAR, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ECON_SN_CELULAR
Kruskal-Wallis chi-squared = 0.4305, df = 1, p-value = 0.5118

```

Puntaje en Inglés vs Tiene computador en su hogar

```
> with(sp1, tapply(MOD_INGLES_PUNT, ECON_SN_COMPUTADOR, median, na.rm=TRUE))
NO SI
9.6 9.6

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_COMPUTADOR, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by ECON_SN_COMPUTADOR
Kruskal-Wallis chi-squared = 0.9971, df = 1, p-value = 0.318
```

Puntaje en Inglés vs El hogar cuenta con conexión a internet

```
> with(sp1, tapply(MOD_INGLES_PUNT, ECON_SN_INTERNET, median, na.rm=TRUE))
NO SI
9.6 9.6

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_INTERNET, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by ECON_SN_INTERNET
Kruskal-Wallis chi-squared = 0.3638, df = 1, p-value = 0.5464
```

Puntaje en Inglés vs Género

```
> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_GENERO, median, na.rm=TRUE))
F M
9.6 9.7

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_GENERO, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by ESTU_GENERO
Kruskal-Wallis chi-squared = 0.2435, df = 1, p-value = 0.6217
```

Puntaje en Inglés vs Situación del hogar actual

```

> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_HOGAR_ACTUAL, median, na.rm=TRUE))
HB.O.PR TEMPORAL
  9.6      9.6

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_HOGAR_ACTUAL, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_HOGAR_ACTUAL
Kruskal-Wallis chi-squared = 0.78, df = 1, p-value = 0.3771

```

Puntaje en Inglés vs Es cabeza de familia

```

> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_SN_CABEZA_FMLIA, median,
+ na.rm=TRUE))
NO SI
9.6 9.5

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_SN_CABEZA_FMLIA, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_SN_CABEZA_FMLIA
Kruskal-Wallis chi-squared = 2.2656, df = 1, p-value = 0.1323

```

Puntaje en Inglés vs Título

```

> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_TITULO_BTO, median, na.rm=TRUE))
A N T
9.60 9.73 9.60

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_TITULO_BTO, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_TITULO_BTO
Kruskal-Wallis chi-squared = 0.7945, df = 2, p-value = 0.6722

```

Puntaje en Inglés vs Tomó algún curso de preparación para el examen

```

> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_TOMOCURSO, median, na.rm=TRUE))
NO SI
9.61 9.03

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_TOMOCURSO, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_TOMOCURSO
Kruskal-Wallis chi-squared = 2.4123, df = 1, p-value = 0.1204

```

Puntaje en Inglés vs Trabaja actualmente

```

> with(sp1, tapply(MOD_INGLES_PUNT, ESTU_TRABAJA, median, na.rm=TRUE))
      AYUDANTE.SIN.R EXPERIENCIA.GASTOS.P          NO
      9.5              9.6              9.7
      PRACTICA.O
      9.1

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_TRABAJA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_TRABAJA
Kruskal-Wallis chi-squared = 8.2005, df = 3, p-value = 0.04204

```

Puntaje en Inglés vs El hogar cuenta con lavadora

```

> with(sp1, tapply(MOD_INGLES_PUNT, ECON_SN_LAVADORA, median, na.rm=TRUE))
      NO SI
      9.6 9.7

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_LAVADORA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ECON_SN_LAVADORA
Kruskal-Wallis chi-squared = 4.2991, df = 1, p-value = 0.03813

```

Puntaje en Inglés vs El hogar cuenta con Servicio cerrado de televisión

```

> with(sp1, tapply(MOD_INGLES_PUNT, ECON_SN_SERVICIO_TV, median,
+ na.rm=TRUE))
      NO SI
      9.6 9.6

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_SERVICIO_TV, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ECON_SN_SERVICIO_TV
Kruskal-Wallis chi-squared = 0.7362, df = 1, p-value = 0.3909

```

Puntaje en Inglés vs El hogar cuenta con servicio de teléfono fijo

```

> with(sp1, tapply(MOD_INGLES_PUNT, ECON_SN_TELEFONIA, median, na.rm=TRUE))
      NO SI
      9.60 9.65

> kruskal.test(MOD_INGLES_PUNT ~ ECON_SN_TELEFONIA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ECON_SN_TELEFONIA
Kruskal-Wallis chi-squared = 0.3939, df = 1, p-value = 0.5302

```

Puntaje en Inglés vs Estado civil

```

> with(spl, tapply(MOD_INGLES_PUNT, ESTU_ESTADO_CIVIL, median, na.rm=TRUE))
  C  SLT  SP  UL
9.20 9.60 9.60 9.65

> kruskal.test(MOD_INGLES_PUNT ~ ESTU_ESTADO_CIVIL, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by ESTU_ESTADO_CIVIL
Kruskal-Wallis chi-squared = 2.3331, df = 3, p-value = 0.5062

```

Puntaje en Inglés vs Educación de la madre

```

> with(spl, tapply(MOD_INGLES_PUNT, FAMI_COD_EDUCA_MADRE, median,
+ na.rm=TRUE))
EDU.SUPERIOR      NINGUNO      PRIMARIA      SECUNDARIA
      9.765          9.400          9.500          9.700

> kruskal.test(MOD_INGLES_PUNT ~ FAMI_COD_EDUCA_MADRE, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by FAMI_COD_EDUCA_MADRE
Kruskal-Wallis chi-squared = 10.6237, df = 3, p-value = 0.01394

```

Puntaje en Inglés vs Educación del padre

```

> with(spl, tapply(MOD_INGLES_PUNT, FAMI_COD_EDUCA_PADRE, median,
+ na.rm=TRUE))
EDU.SUPERIOR      NINGUNO      PRIMARIA      SECUNDARIA
      9.8          9.8          9.6          9.6

> kruskal.test(MOD_INGLES_PUNT ~ FAMI_COD_EDUCA_PADRE, data=spl)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by FAMI_COD_EDUCA_PADRE
Kruskal-Wallis chi-squared = 7.0018, df = 3, p-value = 0.07184

```

Puntaje en Inglés vs Ocupación de la madre

```

> with(sp1, tapply(MOD_INGLES_PUNT, FAMI_COD_OCUP_MADRE, median,
+   na.rm=TRUE))
      EMPLEADO.D.C          HOGAR      PENSIONADO.Y.A
TRABAJADOR.I. (PI.E)
      9.67          9.60          9.80
      9.70

> kruskal.test(MOD_INGLES_PUNT ~ FAMI_COD_OCUP_MADRE, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by FAMI_COD_OCUP_MADRE
Kruskal-Wallis chi-squared = 2.9797, df = 3, p-value = 0.3948

```

Puntaje en Inglés vs Ocupación del padre

```

> with(sp1, tapply(MOD_INGLES_PUNT, FAMI_COD_OCUP_PADRE, median,
+   na.rm=TRUE))
      EMPLEADO.D.C          HOGAR      PENSIONADO.Y.A
TRABAJADOR.I. (PI.E)
      9.50          9.48          9.88
      9.60

> kruskal.test(MOD_INGLES_PUNT ~ FAMI_COD_OCUP_PADRE, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by FAMI_COD_OCUP_PADRE
Kruskal-Wallis chi-squared = 13.0005, df = 3, p-value = 0.004635

```

Puntaje en Inglés vs Forma de pago de la matrícula

```

> with(sp1, tapply(MOD_INGLES_PUNT, FORMA_PAGO_MATRICULA, median,
+   na.rm=TRUE))
      B   PD   PP  VFP
10.1  9.5  9.7  9.6

> kruskal.test(MOD_INGLES_PUNT ~ FORMA_PAGO_MATRICULA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by FORMA_PAGO_MATRICULA
Kruskal-Wallis chi-squared = 1.2894, df = 3, p-value = 0.7317

```

LECTURA CRÍTICA

Puntaje Lectura Crítica vs El hogar cuenta con celular


```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_CELULAR, median,
+ na.rm=TRUE))
NO SI
9.7 9.7

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_CELULAR, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ECON_SN_CELULAR
Kruskal-Wallis chi-squared = 0.0545, df = 1, p-value = 0.8155

```

Puntaje Lectura Crítica vs Tiene computador en su hogar

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_COMPUTADOR, median,
+ na.rm=TRUE))
NO SI
9.65 9.70

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_COMPUTADOR, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ECON_SN_COMPUTADOR
Kruskal-Wallis chi-squared = 0.2986, df = 1, p-value = 0.5848

```

Puntaje Lectura Crítica vs El hogar cuenta con conexión a internet

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_INTERNET, median,
+ na.rm=TRUE))
NO SI
9.70 9.85

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_INTERNET, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ECON_SN_INTERNET
Kruskal-Wallis chi-squared = 1.6362, df = 1, p-value = 0.2009

```

Puntaje Lectura Crítica vs Género

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_GENERO, median, na.rm=TRUE))
  F  M
9.8 9.7

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_GENERO, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ESTU_GENERO
Kruskal-Wallis chi-squared = 0.9191, df = 1, p-value = 0.3377

```

Puntaje Lectura Crítica vs Situación del hogar actual

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_HOGAR_ACTUAL, median,
+ na.rm=TRUE))
  HB.O.PR TEMPORAL
  9.7      9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_HOGAR_ACTUAL, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ESTU_HOGAR_ACTUAL
Kruskal-Wallis chi-squared = 3e-04, df = 1, p-value = 0.9856

```

Puntaje Lectura Crítica vs Es cabeza de familia

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_SN_CABEZA_FMLIA, median,
+ na.rm=TRUE))
  NO  SI
9.70 9.65

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_SN_CABEZA_FMLIA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ESTU_SN_CABEZA_FMLIA
Kruskal-Wallis chi-squared = 0.404, df = 1, p-value = 0.525

```

Puntaje Lectura Crítica vs Título

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_TITULO_BTO, median,
+ na.rm=TRUE))
  A  N  T
9.7 8.9 9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_TITULO_BTO, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ESTU_TITULO_BTO
Kruskal-Wallis chi-squared = 5.0408, df = 2, p-value = 0.08043

```

Puntaje Lectura Crítica vs Tomó algún curso de preparación para el examen

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_TOMOCURSO, median, na.rm=TRUE))
NO SI
9.7 9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_TOMOCURSO, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ESTU_TOMOCURSO
Kruskal-Wallis chi-squared = 0.0139, df = 1, p-value = 0.9063

```

Puntaje Lectura Crítica vs Trabaja actualmente

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_TRABAJA, median, na.rm=TRUE))
AYUDANTE.SIN.R EXPERIENCIA.GASTOS.P NO
9.80 9.70 9.65
PRACTICA.O
8.30

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_TRABAJA, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ESTU_TRABAJA
Kruskal-Wallis chi-squared = 5.3864, df = 3, p-value = 0.1456

```

Puntaje Lectura Crítica vs El hogar cuenta con lavadora

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_LAVADORA, median,
+ na.rm=TRUE))
NO SI
9.6 9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_LAVADORA, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ECON_SN_LAVADORA
Kruskal-Wallis chi-squared = 0.6382, df = 1, p-value = 0.4244

```

Puntaje Lectura Crítica vs El hogar cuenta con Servicio cerrado de televisión

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_SERVICIO_TV, median,
+ na.rm=TRUE))
NO SI
9.7 9.7

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_SERVICIO_TV, data=sp1)

Kruskal-Wallis rank sum test

data: MOD_LECTURA_CRITICA by ECON_SN_SERVICIO_TV
Kruskal-Wallis chi-squared = 0.1452, df = 1, p-value = 0.7032

```

Puntaje Lectura Crítica vs El hogar cuenta con servicio de teléfono fijo

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ECON_SN_TELEFONIA, median,
+ na.rm=TRUE))
  NO  SI
9.70 9.75

> kruskal.test(MOD_LECTURA_CRITICA ~ ECON_SN_TELEFONIA, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ECON_SN_TELEFONIA
Kruskal-Wallis chi-squared = 0.5776, df = 1, p-value = 0.4472

```

Puntaje Lectura Crítica vs Estado civil

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, ESTU_ESTADO_CIVIL, median,
+ na.rm=TRUE))
  C  SLT  SP  UL
9.20 9.75 9.90 9.60

> kruskal.test(MOD_LECTURA_CRITICA ~ ESTU_ESTADO_CIVIL, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by ESTU_ESTADO_CIVIL
Kruskal-Wallis chi-squared = 6.1581, df = 3, p-value = 0.1042

```

Puntaje Lectura Crítica vs Educación de la madre

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, FAMI_COD_EDUCA_MADRE, median,
+ na.rm=TRUE))
EDU.SUPERIOR  NINGUNO  PRIMARIA  SECUNDARIA
          9.6         10.5         9.6         9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ FAMI_COD_EDUCA_MADRE, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by FAMI_COD_EDUCA_MADRE
Kruskal-Wallis chi-squared = 7.5215, df = 3, p-value = 0.05701

```

Puntaje Lectura Crítica vs Educación del padre

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, FAMI_COD_EDUCA_PADRE, median,
+ na.rm=TRUE))
EDU.SUPERIOR  NINGUNO  PRIMARIA  SECUNDARIA
          9.5         10.3         9.7         9.8

> kruskal.test(MOD_LECTURA_CRITICA ~ FAMI_COD_EDUCA_PADRE, data=sp1)

Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by FAMI_COD_EDUCA_PADRE
Kruskal-Wallis chi-squared = 3.8941, df = 3, p-value = 0.2731

```

Puntaje Lectura Crítica vs Ocupación de la madre

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, FAMI_COD_OCUP_MADRE, median,
+ na.rm=TRUE))
      EMPLEADO.D.C          HOGAR      PENSIONADO.Y.A
      9.85             9.70             9.30
TRABAJADOR.I. (PI.E)
      9.50

> kruskal.test(MOD_LECTURA_CRITICA ~ FAMI_COD_OCUP_MADRE, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by FAMI_COD_OCUP_MADRE
Kruskal-Wallis chi-squared = 1.4127, df = 3, p-value = 0.7026

```

Puntaje Lectura Crítica vs Ocupación del padre

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, FAMI_COD_OCUP_PADRE, median,
+ na.rm=TRUE))
      EMPLEADO.D.C          HOGAR      PENSIONADO.Y.A
      9.80             9.90             9.95
TRABAJADOR.I. (PI.E)
      9.45

> kruskal.test(MOD_LECTURA_CRITICA ~ FAMI_COD_OCUP_PADRE, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by FAMI_COD_OCUP_PADRE
Kruskal-Wallis chi-squared = 3.19, df = 3, p-value = 0.3632

```

Puntaje Lectura Crítica vs Forma de pago de la matrícula

```

> with(sp1, tapply(MOD_LECTURA_CRITICA, FORMA_PAGO_MATRICULA, median,
+ na.rm=TRUE))
      B      PD      PP      VFP
10.20  9.60  9.75  9.70

> kruskal.test(MOD_LECTURA_CRITICA ~ FORMA_PAGO_MATRICULA, data=sp1)

      Kruskal-Wallis rank sum test

data:  MOD_LECTURA_CRITICA by FORMA_PAGO_MATRICULA
Kruskal-Wallis chi-squared = 2.4526, df = 3, p-value = 0.4839

```

Como el puntaje en Comunicación Escrita se distribuye como una normal, se realizó una prueba ANOVA con el fin de ver si dos variables (una independiente y otra dependiente) están relacionadas en base a si las medias de la variable dependiente son diferentes en las categorías o grupos de la variable independiente (variable cualitativa). Es decir, señala si las medias entre dos o más grupos son similares o diferentes.

LECTURA CRÍTICA

Puntaje Comunicación Escrita vs El hogar cuenta con celular

```
> AnovaModel.24 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_CELULAR,
+ data=sp1)

> summary(AnovaModel.24)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_CELULAR  1    0.0  0.0021   0.003  0.955
Residuals      217  148.1  0.6826
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ECON_SN_CELULAR,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.23333 0.7080490    21     0
SI 10.24394 0.8372293   198     2
```

Puntaje Comunicación Escrita vs Tiene computador en su hogar

```
> AnovaModel.23 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_COMPUTADOR,
+ data=sp1)

> summary(AnovaModel.23)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_COMPUTADOR  1    0.59  0.5944   0.874  0.351
Residuals          217 147.52  0.6798
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ECON_SN_COMPUTADOR,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.15185 0.7319503    54     0
SI 10.27273 0.8522827   165     2
```

Puntaje Comunicación Escrita vs El hogar cuenta con conexión a internet

```
> AnovaModel.22 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_INTERNET,
+ data=sp1)

> summary(AnovaModel.22)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_INTERNET  1  0.33  0.3315  0.487  0.486
Residuals      217 147.78  0.6810
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ECON_SN_INTERNET,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.21847 0.8630463   157      0
SI 10.30484 0.7196145    62      2
```

Puntaje Comunicación Escrita vs Género

```
> AnovaModel.21 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_GENERO, data=sp1)

> summary(AnovaModel.21)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_GENERO  1  0.54  0.5396  0.792  0.374
Residuals   218 148.46  0.6810
2 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_GENERO,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
F 10.27681 0.8040309   138      1
M 10.17439 0.8599183    82      1
```

Puntaje Comunicación Escrita vs Situación del hogar actual

```

> AnovaModel.20 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_HOGAR_ACTUAL,
+ data=sp1)

> summary(AnovaModel.20)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_HOGAR_ACTUAL  1  2.07  2.075  3.083 0.0805 .
Residuals        217 146.04  0.673
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_HOGAR_ACTUAL,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
HB.O.PR 10.18519 0.8132614  162     2
TEMPORAL 10.40702 0.8404632   57     0

```

Puntaje Comunicación Escrita vs Es cabeza de familia

```

> AnovaModel.19 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_SN_CABEZA_FMLIA,
+ data=sp1)

> summary(AnovaModel.19)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_SN_CABEZA_FMLIA  1  0.85  0.8511  1.254 0.264
Residuals            217 147.27  0.6786
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=ESTU_SN_CABEZA_FMLIA, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.26580 0.8393149  193     2
SI 10.07308 0.6931422   26     0

```

Puntaje Comunicación Escrita vs Titulo

```

> AnovaModel.18 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_TITULO_BTO,
+ data=sp1)

> summary(AnovaModel.18)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TITULO_BTO  2  0.44  0.2222  0.325 0.723
Residuals        216 147.67  0.6837
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_TITULO_BTO,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
A 10.255652 0.7899802  115     1
N  9.866667 0.4509250    3     0
T 10.239604 0.8724768  101     1

```

Puntaje Comunicación Escrita vs Tomó algún curso de preparación para el examen


```

> AnovaModel.17 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_TOMOCURSO, data=sp1)

> summary(AnovaModel.17)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TOMOCURSO  1    0.8  0.7994    1.23  0.269
Residuals     159  103.3  0.6500
61 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_TOMOCURSO,
+   statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.20312 0.8062197   160      1
SI 11.10000      NA     1      0

```

Puntaje Comunicación Escrita vs Trabaja actualmente

```

> AnovaModel.16 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_TRABAJA, data=sp1)

> summary(AnovaModel.16)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TRABAJA  3    4.04  1.3472    2.01  0.113
Residuals    215  144.08  0.6701
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_TRABAJA,
+   statistics=c("mean", "sd")))
              mean      sd data:n data:NA
AYUDANTE.SIN.R  10.25417 0.7837718   96      1
EXPERIENCIA.GASTOS.P 10.45833 0.9699470   24      1
NO               10.19592 0.8127741   98      0
PRACTICA.O       8.60000      NA     1      0

```

Puntaje Comunicación Escrita vs El hogar cuenta con lavadora

```

> AnovaModel.15 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_LAVADORA,
+   data=sp1)

> summary(AnovaModel.15)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_LAVADORA  1    0.0  0.0010    0.002  0.969
Residuals       217  148.1  0.6826
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ECON_SN_LAVADORA,
+   statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.24536 0.8701078   97      0
SI 10.24098 0.7895761  122      2

```

Puntaje Comunicación Escrita vs El hogar cuenta con Servicio cerrado de televisión

```

> AnovaModel.14 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_SERVICIO_TV,
+ data=sp1)

> summary(AnovaModel.14)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_SERVICIO_TV  1   0.65  0.6473   0.952  0.33
Residuals          217 147.47  0.6796
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=ECON_SN_SERVICIO_TV, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.29322 0.8366323   118      0
SI 10.18416 0.8097818   101      2

```

Puntaje Comunicación Escrita vs El hogar cuenta con servicio de teléfono fijo

```

> AnovaModel.13 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ECON_SN_TELEFONIA,
+ data=sp1)

> summary(AnovaModel.13)
              Df Sum Sq Mean Sq F value Pr(>F)
ECON_SN_TELEFONIA  1   0.06  0.0644   0.094  0.759
Residuals          217 148.05  0.6823
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ECON_SN_TELEFONIA,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
NO 10.23583 0.8136270   187      2
SI 10.28438 0.8966279    32      0

```

Puntaje Comunicación Escrita vs Estado civil

```

> AnovaModel.12 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_ESTADO_CIVIL,
+ data=sp1)

> summary(AnovaModel.12)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_ESTADO_CIVIL  3   1.45  0.4831   0.705  0.55
Residuals          214 146.65  0.6853
4 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT, groups=ESTU_ESTADO_CIVIL,
+ statistics=c("mean", "sd")))
      mean      sd data:n data:NA
C   10.20000 0.8602325     7      0
SLT 10.22551 0.8362077   196      2
SP   9.90000      NA     1      0
UL  10.54286 0.6710661   14      0

```

Puntaje Comunicación Escrita vs Educación de la madre

```

> AnovaModel.11 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ FAMI_COD_EDUCA_MADRE,
+ data=sp1)

> summary(AnovaModel.11)
              Df Sum Sq Mean Sq F value Pr(>F)
FAMI_COD_EDUCA_MADRE  3  0.05  0.0154  0.022  0.995
Residuals            215 148.07  0.6887
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=FAMI_COD_EDUCA_MADRE, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
EDU.SUPERIOR 10.20968 0.7760390   31     1
NINGUNO      10.26667 1.2503333    3     0
PRIMARIA     10.25300 0.8245544  100     1
SECUNDARIA   10.24235 0.8421029   85     0

```

Puntaje Comunicación Escrita vs Educación del padre

```

> AnovaModel.10 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ FAMI_COD_EDUCA_PADRE,
+ data=sp1)

> summary(AnovaModel.10)
              Df Sum Sq Mean Sq F value Pr(>F)
FAMI_COD_EDUCA_PADRE  3  0.02  0.0072  0.01  0.999
Residuals            215 148.09  0.6888
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=FAMI_COD_EDUCA_PADRE, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
EDU.SUPERIOR 10.23590 0.7998313   39     1
NINGUNO      10.21111 0.9061518    9     0
PRIMARIA     10.24074 0.8417868  108     1
SECUNDARIA   10.25556 0.8171109   63     0

```

Puntaje Comunicación Escrita vs Ocupación de la madre

```

> AnovaModel.9 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ FAMI_COD_OCUP_MADRE,
+ data=sp1)

> summary(AnovaModel.9)
              Df Sum Sq Mean Sq F value Pr(>F)
FAMI_COD_OCUP_MADRE  3  0.76  0.2519  0.368  0.777
Residuals            215 147.36  0.6854
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=FAMI_COD_OCUP_MADRE, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
EMPLEADO.D.C  10.28378 0.9075344   37     1
HOGAR         10.22207 0.7886758  145     0
PENSIONADO.Y.A 10.08000 0.9307106   10     1
TRABAJADOR.I.(PI.E) 10.35926 0.8845943   27     0

```

Puntaje Comunicación Escrita vs Ocupación del padre

```

> AnovaModel.8 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ FAMI_COD_OCUP_PADRE,
+ data=sp1)

> summary(AnovaModel.8)
              Df Sum Sq Mean Sq F value Pr(>F)
FAMI_COD_OCUP_PADRE  3   1.89  0.6303   0.927  0.429
Residuals          215 146.23  0.6801
3 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=FAMI_COD_OCUP_PADRE, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
EMPLEADO.D.C    10.20482 0.8037873    83     1
HOGAR           9.82000 1.6177144     5     0
PENSIONADO.Y.A  10.18140 0.6887725    43     1
TRABAJADOR.I.(PI.E) 10.33295 0.8499793    88     0

```

Puntaje Comunicación Escrita vs Forma de pago de la matricula

```

> AnovaModel.7 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ FORMA_PAGO_MATRICULA

> summary(AnovaModel.7)
              Df Sum Sq Mean Sq F value Pr(>F)
FORMA_PAGO_MATRICULA  3   1.61  0.5351   0.792  0.5
Residuals          207 139.84  0.6755
11 observations deleted due to missingness

> with(sp1, numSummary(MOD_COMUNICA_ESCRITA_PUNT,
+ groups=FORMA_PAGO_MATRICULA, statistics=c("mean", "sd")))
      mean      sd data:n data:NA
B    10.52000 0.6572671     5     0
PD   10.12174 0.9481832    23     0
PP   10.33279 0.8366045    61     1
VFP  10.18279 0.7942659   122     1

```