

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

**CASO: INGENIERÍA ELECTROMECAÁNICA DE LA UNIVERSIDAD PEDAGÓGICA Y
TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA**

Yeison Darío Acero Báez
Código: 201011254

Trabajo de grado presentado como requisito para optar al título de:
Licenciado en Matemáticas y Estadística

Directora:
M.Sc. Carmen Helena Cepeda Araque

Universidad Pedagógica y Tecnológica de Colombia
Facultad Seccional Duitama, Licenciatura en Matemáticas y Estadística
Duitama - Boyacá, Colombia
2017

Nota de aceptación

Firma jurado 1

Firma jurado 2

Firma presidente del jurado

Duitama, Octubre de 2017

A mi madre Luz Mery por enseñarme que la dedicación es la clave de los logros humanos, y sin ella, nos perderíamos.

“Porque no existe melancolía sin memoria, ni memoria sin melancolía, Will Rogers”.

Agradecimientos

Agradezco a:

Dios, Mis padres, mis hermanas y mi familia por su apoyo.

A mis amigos por las alegrías.

A mis compañeros por su colaboración.

A cada uno de mis profesores y en especial a mi profe Carmen por su infinita paciencia.

RESUMEN ANALÍTICO ESPECIALIZADO - RAE

1.Título	FACTORES DEMOGRÁFICOS, SOCIOECONÓMICOS Y ACADÉMICOS QUE INFLUYEN EN LOS RESULTADOS DEL COMPONENTE GENÉRICO DE LA PRUEBA SABER PRO CASO: INGENIERÍA ELECTROMECAÁNICA DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA
2.Autor	ACERO BÁEZ, Yeison Darío
3.Directora	CEPEDA ARAQUE, Carmen Helena
4.Publicación	Duitama. Universidad Pedagógica y Tecnológica de Colombia. 2.017.
5.Unidad Patrocinante	Universidad Pedagógica y Tecnológica de Colombia. Facultad Seccional Duitama. Escuela de Matemáticas y Estadística.
6.Palabras Claves	Investigación cualitativa, comparación, GAMLSS, Regresión Logística, Saber Pro, Modelos.
7.Objetivo	Determinar los factores demográficos, socioeconómicos y académicos que influyen en los resultados del componente genérico de las pruebas Saber Pro en el programa de Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama.
8.Descripción	En el documento se presenta un estudio que tiene como eje principal modelar el puntaje en las pruebas Saber Pro de los estudiantes de la carrera de Ingeniería Electromecánica vinculados a la UPTC Duitama, desde el segundo semestre de 2011, hasta el 2015, a través de los factores demográficos, académicos y socioeconómicos, teniendo como base los modelos aditivos generalizados para localización, escala y forma.
9.Fuentes	Para el desarrollo de este proyecto se consultaron 22 fuentes principales y varias secundarias que contenían temáticas afines al desarrollo del trabajo, entre ellas se menciona: estudios de las pruebas Saber Pro, Lineamientos internacionales y normativa para la ejecución de dichas pruebas, Bases de datos suministradas por el ICFES, Manual y diccionario de variables de la prueba, teoría sobre los GAMLSS, Regresión Logística Ordinal, entre otros.
10.Contenido	En el documento se presentan cinco capítulos, en el primero se presenta el marco general, el cual contiene el contexto que orienta la investigación, el segundo es la planeación y construcción de la teoría, en el tercero se presenta la metodología con la cual se trabajó, el cuarto consiste en la descripción de los datos de estudio, el quinto consiste en el análisis de la información basados en los modelos de regresión logística y Modelos

	Aditivos Generalizados para Localización, Escala y Forma (GAMLSS) y por ultimo las conclusiones y recomendaciones.
11. Metodología.	<p>Las faces que permitieron determinar la incidencia de los factores demográficos, socioeconómicos y académicos en el componente genérico para estudiantes del programa de Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama, fueron:</p> <ol style="list-style-type: none"> 1. Conformación del marco teórico el cual partió de la revisión documental sobre la Prueba Saber Pro, antecedentes investigativos, factores que inciden en el componente genérico y lo referente a modelos de regresión logística y Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS). 2. Definición del diseño metodológico, el cual, entre otros aspectos, implica la clasificación de las variables explicativas, además de las técnicas utilizadas. 3. Aplicación de instrumentos, tabulación y depuración de la base de datos. 4. Análisis e interpretación de la información y construcción de modelos que describen aquellos factores que influyen en el componente genérico de la prueba Saber Pro. 5. Conclusiones.
12. Conclusiones.	<p>Es importante recalcar que los resultados aquí presentados solo hacen referencia a los estudiantes de Ingeniería Electromecánica de la UPTC objeto de este estudio y que debido a las características de esta, no hace generalizaciones para la población colombiana.</p> <p>Se recomienda, que el análisis efectuado con base en los resultados obtenidos en las pruebas Saber Pro, aplicadas a los estudiantes de dicho programa sirva para desarrollar estrategias al interior del programa.</p> <p>A través de dicho estudio se logró hallar entre otras conclusiones que, tener estado civil de casado genera mejores puntajes en los componentes de inglés y desempeño en inglés, en comparación a estudiantes con otro tipo de estado civil, además contar con un hogar temporal también aumenta la posibilidad de tener puntajes altos en comparación a quienes cuentan con un hogar permanente en razonamiento cuantitativo, además obtener un tipo de bachillerato académico disminuye la posibilidad de lograr puntajes altos respecto a un bachiller de tipo técnico en lectura crítica. En cuanto a la conexión a internet, obtiene más puntos en comunicación escrita quien posea este servicio, así también como un estudiante que no trabaje.</p>
13. Autor del RAE.	Yeison D. Acero B.

TABLA DE CONTENIDO

1. PRESENTACIÓN DEL PROYECTO	11
2. REFERENTES TEÓRICOS	13
2.1. PRUEBAS SABER PRO.....	13
2.2. MODELOS LINEALES (LM).....	19
2.3. MODELOS LINEALES GENERALIZADOS (GLM).....	19
2.4. MODELOS ADITIVOS GENERALIZADOS (GAM).....	22
2.5. MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALAY FORMA (GAMLSS).....	23
2.6. REGRESIÓN LOGÍSTICA ORDINAL (RLO).....	31
3. DISEÑO METODOLÓGICO	35
3.1. VARIABLES DE ESTUDIO.....	35
3.2. MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS.....	40
4. DESCRIPCIÓN DE LOS DATOS	45
4.1. DESCRIPCIÓN UNIVARIADA POR COMPONENTE.....	41
4.2. RESULTADOS COMPETENCIAS GENÉRICAS AÑO 2016.....	46
4.3. DESCRIPCIÓN ESTUDIANTES DE INGENIERÍA ELECTROMECAÁNICA.....	48
4.5. DESCRIPCIÓN BIVARIADA.....	55
5. FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO	58
5.1. ESPECIFICACIÓN DE LOS MODELOS.....	58
5.2. FACTORES DEMOGRÁFICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO.....	63
5.3. FACTORES ACADÉMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO.....	76
5.4. FACTORES SOCIOECONÓMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO.....	89
6. DISCUSIÓN	106
7. CONCLUSIONES	112
REFERENCIAS BIBLIOGRÁFICAS	114
ANEXOS	106
A. RESULTADOS ESTADÍSTICOS ANÁLISIS BIVARIADO.....	116
B. ARTÍCULO.....	163

1 PRESENTACION DEL PROYECTO

El Ministerio de Educación Nacional (MEN) en su proyecto de mejoramiento de la calidad ha implementado en la evaluación de las pruebas Saber Pro un componente genérico que evalúa las competencias de los futuros profesionales de cada centro de educación superior del país, con el fin de establecer un referente de medición externa de la calidad de la educación superior. Es por ello que se hace necesario analizar los factores demográficos, académicos y socioeconómicos que intervienen en los puntajes de los estudiantes de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama.

Por lo tanto, en este trabajo monográfico, se buscó analizar los resultados de las pruebas Saber Pro, presentadas por los estudiantes del programa de Ingeniería Electromecánica, con el fin de identificar los factores que influyen en los resultados del componente genérico, teniendo como fuente las bases de datos suministradas por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES) desde el segundo semestre de 2011 a 2015, y establecer algunos aspectos a mantener o mejorar, para potenciar las competencias genéricas en sus estudiantes. La presentación de la investigación se hace en cinco capítulos, en el primero se presenta el marco general, el cual contiene el contexto que orienta la investigación, el segundo es la planeación y construcción de la teoría, en la cual se encuentra el marco de referencia realizado con base en las teorías de los autores que se tuvo en cuenta para este trabajo, en el tercero se presenta la metodología con la cual se trabajó, el cuarto, consiste en la descripción de los datos de estudio, y por último las conclusiones y recomendaciones.

En este orden de ideas, y para contextualizar al lector se hará una breve descripción en cuanto a la prueba Saber Pro, la cual se divide en dos grandes bloques: el primero es el módulo de competencias genéricas y el segundo corresponde a competencias específicas comunes. “Estas pruebas constituyen un instrumento estandarizado para la evaluación externa de la calidad de la educación superior, y tienen como principales estándares caracterizar el grado de calidad de la educación y la evaluación” (ICFES, 2009), además de medir el desarrollo de las competencias que tienen los futuros profesionales, y que son necesarias para afrontar los desafíos en el mercado laboral, haciendo obligatorio que todos los estudiantes que están próximos a graduarse de los diferentes programas de educación superior presenten la prueba.

Para la Organización para la Cooperación y el Desarrollo Económico (OCDE):

Incluir competencias genéricas de acuerdo con el proyecto: Evaluación de los Resultados del Aprendizaje en Educación Superior (AHELLO), implica estas competencias en pensamiento crítico, razonamiento analítico (la capacidad de generar nuevas ideas y la aplicación práctica de la teoría), resolución de problemas, facilidad de comunicación escrita, capacidad de liderazgo y capacidad de trabajo en equipo. A esta lista se puede añadir la competencia en una segunda lengua, sin embargo, aunque existe un acuerdo

general sobre la importancia de estas habilidades, las competencias genéricas son muy difíciles de medir, ya que existen pocos instrumentos para evaluarlas (MUNDIAL, 2012).

Asimismo la OCDE y el Banco Mundial, en la revisión realizada en 2012 sobre la educación superior, encontró que los estudiantes Colombianos están menos preparados en dichas pruebas en comparación con otros países, debido a que en el país son más jóvenes y sus estándares educativos son bajos, con excepción de estudiantes de altos ingresos económicos.

Mesa (2009) respecto a ello dice que:

La prueba Saber Pro tiene un alto impacto en las universidades, generando cambios en los currículos y su forma de evaluar. También en la forma como la universidad prepara a los estudiantes para presentar la prueba, fija parámetros y criterios para organizar el sistema de evaluación. Además constituye un referente de medición para la universidad en un ámbito local y nacional, posicionándola a través de dicha prueba, de modo que conocer aquellos factores que inciden en los resultados contribuye a mejorar la calidad académica y genera procesos de autoevaluación a fin de mejorar en competencias como razonamiento cuantitativo, lectura crítica, inglés y comunicación escrita

Al inicio del proceso de análisis de los datos, suministrados por el ICFES, se utilizó un modelo lineal generalizado (GLM), se obtuvo que la distribución de las variables dependientes no pertenecía a la familia exponencial de densidades. Para dar solución a este problema se emplearon los GAMLSS, los cuales incluyen algunas distribuciones que presentan una fuerte asimetría y/o curtosis. En cuanto a los desempeños se aplicó regresión logística ordinal.

Se encontró que aquellas variables de tipo académico como el nivel educativo de los padres, resulta influyente en las pruebas Saber Pro, concordando con los resultados de Arias y Ávila (2014), quienes señalan:

En general, a un mayor nivel educativo de los padres mayor será la probabilidad del estudiante de obtener un puntaje alto. Así mismo factores socioeconómicos como tener conexión a internet incide favorablemente en los resultados de los estudiantes en la prueba Saber Pro.

En cuanto a factores demográficos que afectan positivamente los puntajes en los resultados de las pruebas Saber Pro, se encontró que un estudiante soltero o en unión libre presenta mejores puntajes que a estudiantes casados. En esta investigación se logró además establecer que, el tipo de bachillerato que obtenían los estudiantes al salir del colegio incide en los resultados de estas pruebas, así mismo la variable nivel educativo de los padres mostró relación en los puntajes.

Dado el contexto de algunos factores de estudio, se trató de establecer acciones encaminadas al mejoramiento institucional, a través de unas recomendaciones generales, las cuales aparecen en el capítulo de discusión, al final del trabajo.

2 REFERENTES TEÓRICOS

A continuación se presentan las teorías que dan soporte a esta investigación en cuanto a la evaluación de la educación, las pruebas genéricas, las pruebas Saber Pro, las variables demográficas, socioeconómicas, académicas, además de algunos aspectos del modelos aditivos generalizados para localización, escala y forma (GAMLSS).

2.1 PRUEBAS Saber Pro

En el desarrollo del mejoramiento de la calidad de la educación en Colombia, el ICFES ha determinado las pruebas Saber Pro para la medición de competencias en la educación superior.

El Examen de Estado de Calidad de la Educación Superior, Saber Pro, 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 (ICFES, 2012).

Con lo anterior se especula que las pruebas saber pro, son en cierta medida un proceso continuo que tiene como finalidad establecer unas pautas y con base en ellas preponderar el mejoramiento de la educación en Colombia, pero esto no solo lo hace a nivel intrínseco, también lo hace en la medida de catalogar una serie de pautas principales que serán llamadas: COMPETENCIAS, de las cuales se evalúa el para qué. (MEN, 2015)

Con el Ministerio de Educación Nacional (MEN) se definieron los lineamientos para el diseño de los nuevos exámenes, de acuerdo con la política de formación por competencias, tanto en nivel universitario como tecnológico y técnico profesional. Estos módulos se han desarrollado con la participación permanente de las comunidades académicas, redes y asociaciones de facultades y programas.

Los módulos de evaluación son instrumentos que evalúan competencias consideradas fundamentales para los futuros egresados de programas de formación de educación superior. Algunos módulos evalúan competencias genéricas, es decir, competencias que deben desarrollar los estudiantes de cualquier programa de formación. Otros módulos evalúan competencias específicas, comunes a grupos de programas, ya sean de la misma o de distintas áreas de formación, es decir, haceres complejos que comparten distintas formaciones.

“Cada programa deberá seleccionar los módulos que responderán sus estudiantes; aún si son de la misma área, no se pretende que todos los programas tomen los mismos módulos” (ICFES, 2011).

Cabe resaltar que el ICFES cuenta con el respaldo del Ministerio de Educación Nacional y entre ellos han de complementar los lineamientos y pautas principales en la construcción y estructuración de formas de evaluación.

Según el ICFES (2011) el objetivo principal de las pruebas Saber Pro es “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.”

Pero inmersos en este tema se debe hablar con claridad de la existencia de una pauta que dice que presentar el examen Saber Pro se puede hacer pero con la condición de haber cursado el 75% de créditos de una carrera, lo cual es una realidad.

Así mismo se debe hablar de una estructura de la prueba Saber Pro que viene siendo modificada desde años atrás, pero que ha consolidado una base experimental a partir del segundo semestre del año 2011, que sería la siguiente:

Todos los estudiantes deberán presentar la prueba de competencias genéricas, sin importar el programa de formación que cursen. Para la aplicación de 2012-1 la prueba contiene 5 módulos: Escritura, razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés.

Pero para conocer a nivel más subjetivo estas pruebas, se debe mencionar los lineamientos de otras entidades como por ejemplo la OCDE (2012) que resalta:

En las modernas economías, los objetivos prioritarios de la política científica de los países son el desarrollo de la ciencia, la tecnología y la innovación, a través del fomento de la investigación y el desarrollo tecnológico y el fortalecimiento de la competitividad industrial. Para poder planificar, ejecutar y evaluar la actividad científica o técnica, se requiere necesariamente un trabajo estadístico previo de toma de datos básicos y posterior análisis de los mismos, para llegar a construir los necesarios indicadores de dicha actividad (Sancho, 2012).

La OCDE viene siendo una entidad europea que tiene como objetivo el análisis del avance económico de un país a través del desarrollo de manuales estadísticos tendentes a homogeneizar, a nivel internacional, los procedimientos para la selección y recogida de datos estadísticos.

Entorno a este macro sistema, las competencias (componente a analizar) serán determinadas por: las competencias como un conjunto de acciones que el sujeto realiza cuando interactúa significativamente en un contexto determinado, definición que se resume en: un saber hacer en contexto (ICFES, 2011).

La OCDE y el banco mundial dicen sobre las pruebas Saber Pro:

“Un instrumento más reciente para la evaluación de la calidad de la prestación de la educación superior son las pruebas Saber Pro, diseñadas y administradas por el ICFES, el cual está encargado de todas las pruebas y exámenes estandarizados nacionales. Las novedosas pruebas Saber Pro son presentadas por estudiantes de pregrado que terminaron satisfactoriamente por lo menos el 75% de los créditos académicos de su programa. Los

resultados de estas pruebas se pueden comparar directamente con los resultados de las pruebas Saber 11, que se presentan al terminar la educación media; esto permite evaluar el valor agregado por cada institución en particular y por el sistema en general, de la manera que lo han intentado otros países” (OCDE, 2013).

Además se puede determinar según el MEN que:

Los Exámenes de Calidad de Educación Superior (ECAES) son el principal mecanismo de evaluación externa que el Estado le brinda a las instituciones de educación superior, un insumo que les permite conocer sus propios procesos de formación y cotejarlos con el panorama nacional. La creación e implementación de una prueba general para la educación superior, que mida de manera objetiva el desempeño de los estudiantes de último año de formación en este nivel y de los egresados de las distintas áreas del conocimiento, es una búsqueda que se emprendió en nuestro país desde hace tres décadas, pero que solo hasta el año 2003 se concretó en un programa consolidado que ha venido creciendo desde entonces, convirtiéndose en una fuente de información tanto para el Ministerio de Educación Nacional, como para la propia comunidad académica (MEN, 2008).

Por lo tanto, con este panorama se entiende que las pruebas Saber Pro desempeñan el papel de herramienta a través de la cual un diseño evaluativo pueda implementar caracterizaciones que establezcan el desempeño de los estudiantes próximos a graduarse. De acuerdo con el Decreto 3963 de octubre de 2009, son objetivos de las Saber Pro los siguientes:

A: 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.

B: 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.

C: 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, que soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

El componente genérico de la prueba Saber Pro consta de:

PRUEBA DE COMUNICACIÓN ESCRITA: La prueba se enfoca en la competencia para comunicar ideas por escrito con base en el análisis de la información suministrada.

PRUEBA DE PENSAMIENTO CRÍTICO: La prueba de pensamiento crítico se centra en la habilidad de analizar y evaluar proposiciones y textos que presentan puntos de vista similares a los que los evaluados encontrarían en el mundo real.

LECTURA CRÍTICA: Este módulo evalúa competencias relacionadas con la capacidad para leer de manera analítica y reflexiva. Requiere comprender los planteamientos expuestos en un texto e identificar sus perspectivas y juicios de valor. Lo anterior exige que el lector identifique y recupere información presente en uno o varios textos, construya su sentido global, establezca relaciones entre enunciados y evalúe su intencionalidad.

RAZONAMIENTO CUANTITATIVO: Este módulo evalúa competencias relacionadas con las habilidades en la comprensión de conceptos básicos de las matemáticas para analizar, modelar y resolver problemas aplicando métodos y procedimientos cuantitativos y esquemáticos.

DESEMPEÑO INGLÉS: Este módulo evalúa la competencia para comunicarse efectivamente en inglés mediante preguntas de selección múltiple con única respuesta. Estas competencias, alineadas con el Marco Común Europeo, permiten clasificar a los examinados en seis niveles de desempeño A-, A1, A2, B1, B+, B2. No se presentan las categorías C1 o C2. Esto se debe a que solo una proporción bastante reducida de los estudiantes colombianos alcanza estos niveles de inglés.

A-: No alcanza el nivel A1.

A1: Es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato.

A2: Es capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes (información básica sobre sí mismo y su familia, compras, lugares de interés, ocupaciones, etc.)

B1: Es capaz de comprender los puntos principales de textos claros y en lengua estándar si tratan sobre cuestiones que le son conocidas, ya sean en situaciones de trabajo, de estudio o de ocio.

B+: Supera el nivel B1.

B2: El estudiante es capaz de entender las ideas principales de textos complejos que traten de temas concretos abstractos, incluso si son de carácter técnico, siempre que estén dentro de su especialización. Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad, de modo que la comunicáis se realice sin esfuerzo por ninguno de los interlocutores.

Puede producir textos claros y detallados en torno a temas diversos, así como defender un punto de vista sobre temas generales indicando los pros y los contras de las distintas opciones.

DESEMPEÑO COMUNICACIÓN ESCRITA: El nivel de desempeño es una descripción cualitativa de las competencias de los estudiantes en cada módulo. Se espera que una persona ubicada en un determinado nivel demuestre las competencias de ese nivel y de los niveles inferiores.

Niveles de desempeño en escritura:

Nivel 0: No contestó o el escrito no fue legible

Nivel 1: En el escrito se aborda la tarea propuesta, pero hay problemas en el manejo de las convenciones (serios errores de sintaxis, puntuación o escritura de las palabras) o un desarrollo insuficiente del tema (es tan breve que no se pueden aplicar los criterios de análisis).

Nivel 2: En el escrito se aprecian ideas, pero estas pueden ser incoherentes, o se presentan desarticuladamente. No hubo desarrollo organizado del tema o el escrito pudo ser innecesariamente largo o repetitivo.

Nivel 3: En el escrito se aprecia el esbozo de una intención comunicativa, es decir, se ve que quien escribe pretende alcanzar, por medio del escrito busca una reacción específica en el lector. Se encuentran problemas de manejo del lenguaje, pero estos no impiden la comprensión de los enunciados.

Nivel 4: En el texto se encuentra una idea central que se desarrolla de acuerdo con una intención comunicativa. El texto también posee una estructura básica, en otras palabras, es posible identificar 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).

Nivel 5: El texto alcanza unidad, por medio de la progresión temática, es decir, logra encadenar o relacionar efectivamente las ideas, dándole continuidad al escrito: incorpora información nueva vinculándola con la anterior, presentando la información en un orden cronológico, partiendo de temas generales para desglosar temas específicos, entre otras maneras de interrelacionar contenidos. Todo el texto se desarrolla en un mismo eje temático, aunque pueden presentarse fallas en el uso de conectores. Hay un buen uso del lenguaje, aunque pueden encontrarse errores en la aplicación de algunas reglas de ortografía y puntuación.

Nivel 6: En el texto se identifica la posición de quien escribe, se expresan con claridad las ideas y hay un uso correcto de las expresiones que permiten conectarlas. Hay un uso adecuado de distintos mecanismos que le dan coherencia y cohesión al texto (signos de puntuación, conectores, etc.). Lo anterior permite que el escrito sea fluido.

Nivel 7: En el texto se evidencia una planeación en la escritura en dos aspectos fundamentales. En el primero, el autor piensa en cómo expresar sus ideas de manera efectiva, aplicando diversos recursos textuales para evidenciar sus planteamientos. En el segundo, el autor adecúa su discurso para un público determinado, prevé los conocimientos previos de su lector y busca el lenguaje y los conceptos apropiados.

Nivel 8: El texto trasciende el estímulo dado, complejizando los planteamientos de manera efectiva tanto a nivel de pensamiento como de recursos lingüísticos. Incluye el problema planteado en un dialogo de ideas y posiciones, en una perspectiva más amplia e intertextual que construye y precisa su sentido. El texto de este nivel es de alto interés y motivación para el lector, debido a la eficacia del escrito.

Como se observó anteriormente los desempeños presentan una descripción cualitativa, en cuanto a los componentes de razonamiento cuantitativo, lectura crítica, comunicación escrita e inglés se muestra la siguiente tabla la cual ilustra la escala de interpretación de resultados de las pruebas Saber Pro de la siguiente forma:

PUNTAJE	RESULTADO
< 9	Muy bajo
9-9.5	Bajo
9.6-10.2	Regular
10.3-10.6	Bueno
10.7-11	Muy bueno
> 11	Excelente

*Sánchez, Juan de Jesús, Guía de interpretación de resultados de la prueba Saber Pro 2013.

2.2 MODELOS LINEALES (LM)

El modelamiento consiste en la aplicación de una serie de procesos con el objeto de conseguir una explicación apropiada del comportamiento de una variable respuesta (datos) a partir de una o más variables explicativas (modelo), la explicación del modelo en general no puede ser perfecta, existe un error o residual (Díaz, 2012).

El modelo más sencillo que podemos crear es un Modelo Lineal (LM), este modelo aunque sencillo logra cubrir un amplio aspecto de casos a pesar de sus estrictas restricciones como pueden ser el hecho que la variable respuesta debe adaptarse a una distribución normal o bien que la relación entre la variable, o las variables explicativas, y la variable respuesta deba asumirse lineal (Pérez, 2016).

Podemos expresar un modelo LM de la siguiente forma:

$$Y_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{rxri} + e_i$$

Donde $e_i \sim N(0, \sigma^2)$ para $i = 1, 2, 3, \dots, n$ para este modelo se asume que los errores e_i , son independientes y siguen una distribución normal de media cero y desviación constante. El modelo se expresa de la forma:

$$Y_i \sim N(0, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{rxri} + e_i$$

Donde X es una matriz np ($p = r+1$) que contiene todas las variables explicativas (además de una columna de 1 si se necesita una constante) y β es un vector desconocido de longitud p que será estimado a partir de las variables explicativas. A pesar de que la distribución es adecuada para numerosos e importantes ejemplos estudiados desde que se empezó a trabajar como concepto de regresión, también se ve que la distribución normal no puede abarcar todos los casos existentes. Los modelos lineales son un tipo de modelos de regresión univariante muestra que, entre otras, se dan dos asociaciones. La variable respuesta se considera que sigue una distribución normal y, además que la relación existente entre las variables explicativas y una variable respuesta tenga una forma lineal. Atendiendo la necesidad de que la variable respuesta de un modelo de regresión no siga una distribución normal aparecen los GLM, que son introducidos a continuación (Pérez, 2016).

2.3 MODELOS LINEALES GENERALIZADOS (GLM)

A principios de la década de los 70, Nelder y Wedderburn (1972), citados por Pérez (2016) propusieron los modelos lineales generalizados (GLM), que engloban varios modelos de

regresión ya existentes, proporcionando un marco unificador para aquellos modelos en los que la distribución de la variable respuesta pertenece a la familia exponencial como pueden ser por ejemplo, los modelos de regresión lineal o uno de regresión logística.

Para flexibilizar ligeramente un modelo LM aparece un Modelo Lineal Generalizado (GLM). Este modelo sigue una relación lineal entre la variable respuesta y las covariables del modelo. No obstante, permite que la distribución asumida por la variable respuesta no sea necesariamente una distribución normal, pudiendo ser esta una Poisson o una Binomial (Pérez, 2016),

Entre otras:

“Los Modelos lineales Generalizados (GLM) abarcan modelos de regresión ordinal con distribución no normal. Cuando se estudian varias mediciones simultáneamente, resulta útil conseguir una descripción parsimoniosa de los datos a través de un modelo matemático que explique, de alguna forma, las observaciones; a esto se le denomina Modelo óptimo, explicando la mayor variabilidad con el mínimo número de parámetros (principio de parsimonia)”, mencionado por Díaz (2012).

Las etapas del modelamiento estadístico se pueden resumir en:

Especificación: Se determinan los supuestos del componente aleatorio, se establece la función del componente sistemático y se determina como los dos componentes son combinación en el modelo mediante la función enlace.

Selección: Se estiman los parámetros del componente sistemático y se valora la precisión de las estimaciones, calculando la discrepancia entre pares de modelos, lo anterior con el objetivo de seleccionar el modelo óptimo.

Evaluación: Donde se analiza si el modelo representa la realidad más aproximada posible. En ello se analiza a través de una prueba, si dicho modelo cumple con las especificaciones.

Una variable respuesta $\{Y_i\}$ con alguna distribución de probabilidad (componente aleatorio) $i = 1, 2, \dots, n$, la cual pertenezca a la familia exponencial de densidades.

Un conjunto de variables explicativas $\{x_i\}$ (componente sistemático) y un vector de parámetros β .

Una función de enlace $g(.)$ entre el componente aleatorio y el componente sistemático la cual describe como se relaciona $x'\beta$ con el valor esperado de $\{Y_i\}$. Implica que $f(\beta_0, \dots, \beta_p, x_1, \dots, x_p)$ es una combinación lineal de parámetros desconocidos es decir:

$$g(\mu_i) = x' \beta = \beta_0 + \beta_1 x_i + \dots + \beta_p x_p = \sum_{j=1}^p \beta_j g_j(x_j)$$

Se puede indicar que las mejoras más importantes introducidas por los modelos GLM fueron los siguientes:

El supuesto de que la variable respuesta sigue una distribución normal fue sustituido por el de que la variable respuesta sigue una distribución de la familia exponencial, pudiendo incluir las distribuciones como la Gamma o una Poisson.

Se comenzó a utilizar una función de enlace monótona, denotada por $g(\cdot)$, para modelar la relación existente entre las variables explicativas y la media de la distribución escogida para la variable respuesta.

Atendiendo a estas consideraciones se puede escribir, utilizando una notación matricial, el modelo GLM de la siguiente forma:

$$Y \sim \text{ExpF}(\mu, \Phi),$$

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

donde la denominación *ExpF* se refiere a la familia exponencial.

Para cada observación, la función de densidad de probabilidad de la familia exponencial puede escribirse como sigue:

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

donde $E(Y) = \mu = b'(\theta)$ y $\text{Var}(Y) = \phi v$ siendo $v(\mu) = b''[\theta(\mu)]$.

Con la introducción de los modelos GLM se consigue suavizar considerablemente la restricción de que la distribución de la variable respuesta tuviese que ser necesariamente normal. No obstante, o de hecho de que la relación entre la variable respuesta y las variables explicativas no fuese lineal seguiría pendiente de modelación. Trabajando con modelos GLM a los que se le introdujeron, por ejemplo, polinomios. Por lo tanto, la relación de linealidad fue la principal motivación para la aparición de los modelos aditivos generalizados (GAM) que se expone en el siguiente apartado (Pérez, 2016).

2.4 MODELOS ADITIVOS GENERALIZADOS (GAM)

“Los modelos GAM fueron populares en la década de los 80. Los primeros en introducir estos modelos fueron Hastie e Tibshirani (1990) y, posteriormente, fueron extendidos por Wood en el año 2006” (Pérez, 2016).

Dichos modelos se pueden considerar una extensión no paramétrica de los GLM; la idea de los modelos GAM es permitir que sean los propios datos los que determinen la relación entre un predictor lineal η y a las variables explicativas. Un modelo GAM se puede escribir de la siguiente forma:

$$Y \sim \text{ExpF}(\mu, \Phi)$$

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j$$

donde h son funciones de suavizado no paramétricas que se aplican sobre aquellas variables explicativas continuas que no presentan una relación lineal con respecto a la variable respuesta.

Como lo explica Pérez (2016):

Un siguiente paso, antes de la creación de los Modelos Aditivos Generalizados (GAM), fue intentar construir relaciones no lineales entre la variable respuesta y las covariables del modelo utilizado, por ejemplo los polinomios. Aun así, las suavizaciones de hecho puntualmente sobre los GLM no se aproximaban suficiente y dieron lugar a la de los modelos GAM. Estos últimos permiten la utilización de funciones suavizadoras para describir la relación entre la variable respuesta y las variables explicativas en un sentido no paramétrico. No obstante, y a pesar de las modificaciones de hecho sobre los GAM, todos estos modelos quedan cortos para aquellas distribuciones que tienen más de un parámetro (que son la gran mayoría) porque todos los modelos mencionados hasta ahora solo son capaces de modelar un parámetro de localización directamente a partir de las covariables. En algunas ocasiones también se logra modelar un parámetro de escala, pero siempre a través de una relación con el parámetro de localización; por esta razón aparecieron los Modelos Aditivos de Localización, Escala y Forma (GAMLSS). Dichos modelos permiten la modelación del parámetro de escala, e incluso de los parámetros de forma (asimetría y curtosis) de la distribución utilizada directamente a partir de las variables explicativas.

Con la introducción de los GAM quedan flexibilizadas las dos condiciones más restrictivas de los modelos LM que eran la imposibilidad de considerar una familia distinta a la normal y de considerar una relación no lineal entre la variable respuesta y las distintas variables explicativas. Aun así, todo es mejorable y con la idea de introducir

nuevas distribuciones adicionales a las de la familia exponencial y de poder modelar un parámetro de escala y los parámetros de forma (dependiendo de la distribución) directamente a partir de las variables explicativas, sin necesidad de utilizar la relación existente entre un parámetro de localización y los restantes parámetros, apareciendo los modelos GAMLSS (Pérez, 2016).

2.5 MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALA Y FORMA (GAMLSS)

Así mismo, Pérez (2016), indica que:

Estos modelos hacen referencia a un grupo de modelos estadísticos realizados por Rigby y Stasinopoulos (2002), permitiendo modelar la media o mediana de las variables dependientes y además, la variabilidad y la asimetría, en relación con variables independientes. Los modelos aditivos generalizados para la localización, escala y forma son unos modelos de regresión univariante que se definen como objetivo de superar varias limitaciones de los modelos ya existentes en este campo, como pueden ser los Modelos Lineales Generalizados (GLM) o los Modelos Aditivos Generalizados (GAM). Los GAMLSS permiten llevar a cabo diversas mejoras, entre las más importantes están poder abandonar la familia exponencial para escoger la distribución de los datos de entre una amplia gama de opciones, incluida algunas que presentan una fuerte asimetría y/o curtosis. Otra de las ventajas desde modelos es la posibilidad de modelar todos los parámetros de distribución directamente a partir de variables explicativa.

Un modelo lineal generalizado (GLM) y un modelo aditivo generalizado (GAM) ocupan un lugar importante en el campo de las técnicas estadísticas para la regresión univariante, ver Nelder e Wedderburn (1972) y Hastie y Tibshirani (1990) respectivamente quienes son mencionados por Pérez. Estos dos modelos consideran, para la variable respuesta, una distribución de la familia exponencial. La media es modelada a partir de las variables explicativas mientras que la varianza está dada por la siguiente expresión $\text{Var}(Y) = \phi \nu(\mu)$, donde ϕ es un parámetro de dispersión constante y, por lo tanto, depende de la media. Además si se considera una distribución de familia exponencial, tanto la asimetría como la curtosis están en función de la media y del parámetro ϕ . Por lo tanto, si se trabaja con modelos GLM o GAM ni la varianza, ni la asimetría, ni la curtosis es modelada explícitamente a partir de las variables explicativas sino a través de la relación de las variables independientes con la media y, la de esta última, con el resto de los parámetros.

Los GAMLSS pueden considerarse modelos de regresión semi paramétricos. Son paramétricos en sentido no que necesitan una distribución paramétrica para la variable respuesta y semi en sentido no que lo modelado de los parámetros de la distribución en función de las variables explicativas puede implicar el uso de funciones suavizadoras no paramétricas.

En los modelos GAMLSS la suposición de que la variables respuesta “Y” pertenece a la familia exponencial y flexibilizada, permitiendo así trabajar con distribuciones con mayor asimetría o curtosis. Por lo tanto, la distribución de la variable respuesta “Y” puede ser seleccionada de entre una amplia gama de distribuciones, incluida aquellas con una alta asimetría o curtosis, tanto continuas como discretas. Además, la parte sistemática del modelo se amplía permitiendo modelar no solo la media (parámetro de localización) sino también la varianza (parámetro de escala), la asimetría o la curtosis (parámetro de forma). Incluye relaciones lineales como no lineales entre los parámetros de las distintas variables explicativas.

Pérez (2016) continúa:

Recordando la forma de los GAM diciendo que un modelo GAMLSS puede escribirse de la siguiente forma:

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j$$

Y teniendo en cuenta que el modelo GAMLSS modela no solo la media sino todos los parámetros de la distribución a partir de las variables explicativas, siguiendo el modelo:

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

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

$$g_3(v) = X_3\beta_3 + \sum_{j=1}^{J_3} h_{j3} X_{j3}$$

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

Ya que en la mayoría de los casos prácticos se tienen como mucho 4 parámetros, siendo estos la media, varianza, asimetría y curtosis, estos 4 parámetros se denotaran por μ , σ , v , τ respectivamente. Si se quiere generalizar un poco la notación utilizada se puede escribir:

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

Donde θ_k es el vector de parámetros de la distribución.

El modelo GAMLSS queda descrito de la siguiente forma:

$$y \sim D(\theta_k)$$

$$g_k(\theta_k) = X_k \beta_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk} \quad ,$$

donde D es la distribución de la variable respuesta y Z_{jk} y γ_{jk} para $k = 1, 2, 3, 4$ y $j = 1, \dots, J_k$ son matrices de diseño para los términos lineales y suavizados respectivamente. β_{jk} es un vector de parámetros de longitud J_k y γ_{jk} es una variable aleatoria q_{jk} dimensional. En GAMLSS se lleva a cabo maximizando la siguiente función de verosimilitud:

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

donde $\sum_{i=1}^n \log f(y_i | \theta_i)$ es un algoritmo de verosimilitud, λ_{jk} son los parámetros y G_{jk} es una matriz simétrica que depende de los parámetros λ_{jk} . Por lo tanto, se debe estimar β , λ y γ .

Siendo $M = D, G, T, \Lambda$ un modelo GAMLSS donde:

D = especifica la distribución de la variable respuesta

G = especifica el conjunto de funciones enlace (g_1, g_2, \dots, g_p) para los parámetros $(1, \dots, p)$

T = engloba a las variables utilizadas en los predictores correspondientes a μ , σ , v y τ

Λ = engloba los distintos parámetros de suavizado para las funciones de suavizado para un conjunto de datos en concreto, el proceso de selección consiste en comparar muchos modelos construidos combinando los distintos componentes de M , hablando ahora de cada una de estas componentes por separado.

Componente D : La selección de una distribución apropiada cuenta con 2 pasos. El primer paso será ajustar diferentes modelos utilizando diferentes distribuciones y permaneciendo con aquella que proporcione un modelo con menor AIC. El siguiente paso será validar la distribución escogida a través de un proceso de diagnóstico como puede ser la gráfica de gusano.

Componente G: La selección de la función de enlace viene determinada generalmente por el rango 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 aparece usando el criterio deviance (se quedará con el que presente la 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 requerirá de su correspondiente parámetro de suavizado. Este puede ser previamente fijado o estimado a partir de los datos.

Para este trabajo no se llevarán a cabo suavizadores.

Una vez tenido en cuenta todos los componentes del modelo M y construidos varios modelos se debe tratar de escoger uno entre todos los propuestos. A la hora de escoger el modelo más acertado se puede utilizar el criterio AIC.

Una vez ajustado el modelo GAMLSS, se utilizarán los residuales cuantiles aleatorizados descritos en Dunn y Smyth (1996), mencionados por Pérez (2016), para comprobar la adecuación del modelo y, más específicamente, la distribución elegida para la variable respuesta “Y”.

Los residuos cuantiles aleatorizados vienen dados por la expresión $r_i = -1 (\mu_i)$ donde -1 es la función inversa de la distribución acumulada 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 $[F((y_i - 1)|\hat{\theta}_i), F(y_i|\hat{\theta}_i)]$ si y_i es discreta.

La principal ventaja de estos nuevos residuos es que sea cual sea la distribución de la variable respuesta, estos siempre tienden a una distribución normal estándar, siempre y cuando el modelo sea correcto (Pérez, 2016).

La función `plot.gamlss()` representa gráficamente si aquellos residuales aleatorizados presentan un buen ajuste, la función `plot()` tiene como salida 4 gráficos que representan lo siguiente:

1: residuos frente a los valores ajustados para a media

2: residuos frente a las variables explicativas del modelo.

3: estimación Kernel de densidad de los residuos.

4: QQ-plot de los residuos.

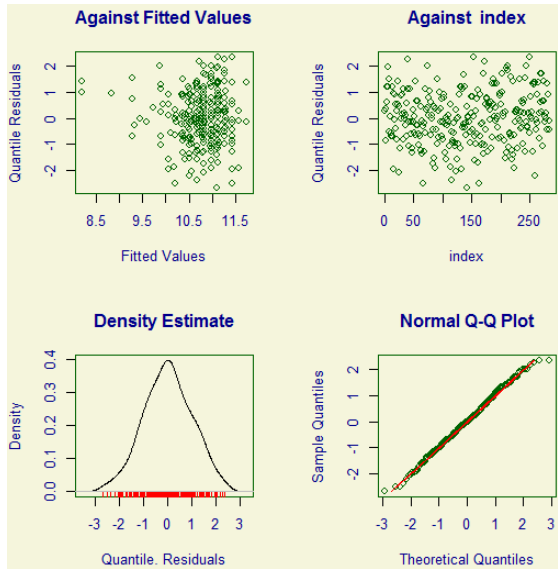


Figura 2-1: Función `plot.gamlss()`

Otra función que tiene la intención de identificar las relaciones de una variable explicativa donde el modelo no se ajusta adecuadamente a los datos, es la función `wp()` worm plot introducidos por van Buuren e Fredrils (2001) mencionados por Pérez (2016) y tiene como salida un único gráfico el cual representa lo siguiente:

1: Los puntos representan los residuos y la línea discontinua horizontal sus valores esperados, por lo tanto observando esto se puede ver que tan lejos están los unos de los otros.

2: Las líneas discontinuas negras marcan el intervalo de confianza del 95%, por lo tanto para que se pueda considerar que el modelo es correcto, sólo un 5% de los puntos podrían quedar fuera de este intervalo, si esto no es así, debe considerarse el modelo como inadecuado para explicar la variable respuesta.

3: La curva continua horizontal es un ajuste cúbico de los puntos y puede indicar diferentes problemas en el modelo (Stasinopoulos, 2015).

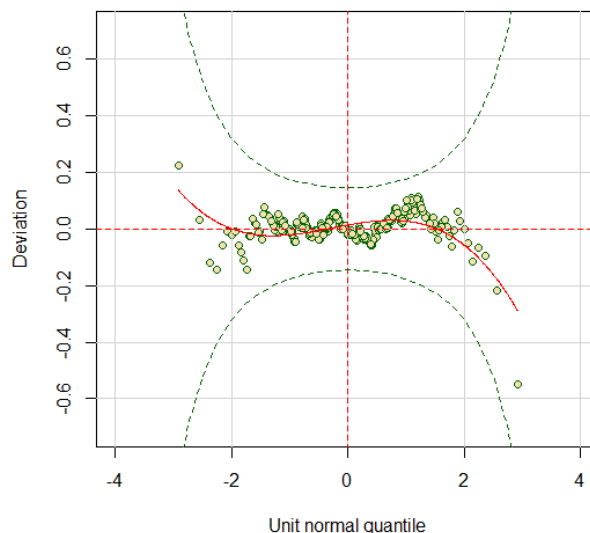


Figura 2-2: Función $w_p()$

Cuando se dispone a trabajar con modelos GAMLSS son muchas las funciones que están disponibles en R. Para esta monografía se tendrá en cuenta el paquete `gamlss()`, y las funciones `histDist()` y `fitDist()` las cuales ajustan los datos a una determinada distribución de entre una amplia gama de distribuciones. En este caso el argumento más relevante que indica el tipo de distribución que se quiere ajustar a los datos es la función `realAll`, esta opción recoge todas las distribuciones continuas que están definidas en toda la recta real. Esta función juega un papel importante a la hora de escoger la distribución más acertada para un conjunto de datos.

SELECCION DEL MODELO

Una vez ajustado el modelo debemos hacer una buena selección, una de las funciones de ayuda para una buena selección es `stepGAICAll.A()`, la estrategia utilizada por esta función para una distribución dada es:

1. Utilizar el criterio GAIC hacia adelante para seleccionar un modelo apropiado para μ , fijando δ , v y τ .
2. Dado el modelo para μ obtenido en 1 y para v y τ fijas, usa un proceso de selección hacia adelante para escoger un modelo apropiado para δ .
3. Dados los modelos para μ y δ obtenidos en 1 y 2 respectivamente y con τ fija, se utiliza un proceso de selección hacia adelante para escoger un modelo apropiado para v .

4. Dados los modelos para μ , δ y ν obtenidos en 1, 2 y 3 respectivamente, se usa un proceso de selección hacia adelante para escoger un modelo apropiado para τ .
5. Dados los modelos para μ , δ y τ obtenidos en 1, 2 y 4 respectivamente, es utilizado un proceso de selección hacia atrás para escoger un modelo apropiado para ν .
6. Dados los modelos para μ , ν y τ obtenidos en 1, 5 y 4 respectivamente, se usa un proceso de selección hacia atrás para escoger un modelo apropiado para δ .
7. Dados los modelos para δ , ν y τ obtenidos en 6, 5 y 4 respectivamente, se utiliza un proceso de selección hacia atrás para escoger un modelo apropiado para μ y finalizar de esta manera con el proceso.

ALGUNAS DISTRIBUCIONES CONTINUAS PARA LOS GAMLSS

“Una vez realizado el ajuste de los datos a una determinada distribución para todas las variables respuesta, se obtuvo como resultado la siguiente tabla, la cual muestra la distribución para cada variable respuesta y sus correspondientes funciones de enlace” (Pérez, 2016);

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	α	ν	τ
Skew t type 3 (ST3)	Razonamiento cuantitativo Lectura crítica	identity	log	log	log
Logistic (LO)	Comunicación escrita	identity	log	---	---
Skew t type 2 (ST2)	Inglés	identity	log	identity	log

Distribución Skew t type 3 (ST3)

La distribución Skew t type 3, denotada por $ST3(\mu, \delta, \nu, \tau)$, definida por:

$$f_Y(y|\mu, \delta, \nu, \tau) = \frac{c}{\delta} \left\{ 1 + \frac{z^2}{\tau} \left[\nu^2 I(y < \mu) + \frac{1}{\nu^2} I(y \geq \mu) \right] \right\} \quad (2-21)$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$ y $\delta > 0$, $\nu > 0$ y $\tau > 0$, y cuando $z = \frac{(y-\mu)}{\delta}$. La media y la varianza estan dadas por: $E(Y) = \mu + \delta E(Z)$ y $Var(Y) = \delta^2 V(Z)$. Fernandez y Steel (1998), p360, eqn. (5). Con función de enlace $\mu = X\beta - \delta E(Z)$, $\delta^2 = \frac{e^{X\beta}}{V(Z)}$, $\nu = e^{X\beta}$ y $\tau = e^{X\beta}$

Distribución Logistic (LO)

La parametrización de la distribución Logistic, denotada aqui como $LO(\mu, \delta)$, es dada por:

$$f_Y(y|\mu, \delta) = \frac{1}{\delta} \left\{ \exp\left[-\left(\frac{y-\mu}{\delta}\right)\right] \right\} \left\{ 1 + \exp\left[-\left(\frac{y-\mu}{\delta}\right)\right] \right\}^{-2} \quad .18)$$

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

Con función de enlace $\mu = X\beta$ y $\delta^2 = \frac{3e^{X\beta}}{\pi^2}$

Distribución Skew t type 2 (ST2)

La distribución Skew t type 2, denotada por $ST2(\mu, \delta, \nu)$, definida por:

$$f_Y(y|\mu, \delta, \nu, \tau) = \frac{c}{\delta} \left\{ 1 + \frac{z^2}{\tau} \left[\nu^2 I(y < \mu) + \frac{1}{\nu^2} I(y \geq \mu) \right] \right\}$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$ y $\delta > 0$, $\nu > 0$ y $\tau > 0$, y cuando $z = \frac{(y-\mu)}{\delta}$. La media y la varianza estan dadas por: $E(Y) = \mu + \delta E(Z)$ y $Var(Y) = \delta^2 V(Z)$. Fernandez y Steel (1998), p360, eqn. (5). Con función de enlace $\mu = X\beta - \delta E(Z)$, $\delta^2 = \frac{e^{X\beta}}{V(Z)}$, $\nu = e^{X\beta}$ y $\tau = e^{X\beta}$

2.6 REGRESION LOGÍSTICA ORDINAL (RLO)

La Regresión Logística es una técnica estadística multivariante que nos permite estimar la relación existente entre una variable dependiente no Paramétrica, en particular dicotómica y un conjunto de variables independientes Paramétricas o no Paramétricas. Para este caso la variable dependiente indica la ocurrencia o no del suceso. El objetivo de la regresión logística es predecir la probabilidad de un evento de interés en una investigación, así como identificar las variables predictoras útiles para la predicción (Díaz, 2012).

En el análisis de datos es frecuente encontrarse con variables dicotómicas (sí/no, presencia /Ausencia), o variables medidas en escala ordinal. Una práctica usual, es tratar este tipo de variables como si fueran continuas, asignándoles una puntuación arbitraria basada en la codificación de las distintas categorías de respuesta, esta práctica, si bien pudiera considerarse correcta en el caso de variables ordinales, no lo es si las variables son simplemente nominales (Cañadas, 2013).

Los modelos de regresión logística son una herramienta que permite explicar el comportamiento de una variable respuesta discreta (binaria o con más de dos categorías) a través de una o varias variables independientes explicativas de naturaleza cuantitativa y/o cualitativa. Según el tipo de variable respuesta estaremos hablando de regresión logística binaria (variable dependiente con 2 categorías), o de regresión logística multinomial (variable dependiente con más de 2 categorías), pudiendo ser esta última de respuesta nominal u ordinal. Los modelos de respuesta discreta son un caso particular de los modelos lineales generalizados formulados por Nelder y Wedderburn en 1972, (Iglesias, 2013).

$$P(Y = 1) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 X_1)}}$$

En el caso particular de la RLO se utiliza una función de enlace para relacionar de forma lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la categoría i de la variable ordinal, y la probabilidad que la variable tome un valor mayor que la categoría i .

En la RLO las funciones de enlace más empleadas son la Logit y la Cloglog. Plantea que el enlace con Logit es 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 (Agresti, 2002).

Para el caso de este estudio, los valores de la variable ordinal representan los puntajes en cuanto al desempeño en inglés y comunicación escrita, cuya complejidad es valorada por los educandos como de Muy bajo a Excelente, por lo tanto no ocurre que los mayores valores (mejores puntajes)

sean los más probables, siendo plausible considerar la función de enlace Logit como la más satisfactoria para este caso.

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:

$$O_i = P(\text{valor sea } \leq \text{categoría } i \mid \text{valores de } X) / (1 - P(\text{valor sea } \leq \text{categoría } i \mid \text{valores de } X)) \quad *2$$

Con el término valor en la ecuación (*2) se hace referencia a cualquier valor de la variable dependiente. Como se observa, las probabilidades de la ecuación (*2) es condicional, 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 variable independientes, existen p coeficientes, y βX se reemplaza por la combinación lineal entre $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$. Estos coeficientes cuantifican el efecto de las variables independientes sobre el logaritmo de la razón de probabilidad.

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 esta 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 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 tanto existen ecuaciones $k-1$ con los mismos coeficientes acompañando a las variables explicativas y que solo se diferencian en el valor del intercepto (Agresti, 2002).

“Para estimar los coeficientes de la ecuación de regresión se utilizan diversos procedimientos, siendo la estimación de máxima verosimilitud el más empleado” (Agresti, 2002). Agresti continúa con su explicación:

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:

$$G = -2 \ln \left(\frac{\text{probabilidad de la muestra sin variables}}{\text{probabilidad de la muestra con variables}} \right)$$

Este estadístico sigue distribución ji-cuadrado, con tantos grados de libertad como variables independientes existan, dicho por McCullagh, quien es citado por Agresti, quien recalca que: “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 estas, 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 variable. El estadístico utilizado es el siguiente:

$$Z_{\text{Wald}} = \frac{b_j}{\text{ES}(b_j)} \quad ,$$

donde b_j es el coeficiente de regresión estimado para la variable independiente j. Bajo la hipótesis de que el coeficiente poblacional b_j para la variable j, la razón entre la estimación de este coeficiente (b_j) y el error estándar de esta estimación $[(\text{ES } b_j)]$, sigue una distribución normal estándar. Mayores valores de este estadígrafo indican que el coeficiente b_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”. También, Agresti plantea que para comparar estas frecuencias generalmente se calcula el estadístico ji cuadrado de Pearson según la fórmula:

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

en la expresión se considera que la variable dependiente tiene k categorías y que se forman m combinaciones de valores con las variables explicativas, de manera que: **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.

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 X^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 que:

$$P(\text{valor sea} \leq \text{categoría } i \mid \text{valores de } X) - P(\text{valor sea} > \text{categoría } i \mid \text{valores de } X) = \ln(O_i) = \alpha_i + \beta x$$

De esta ecuación se deriva que:

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

La expresión es de suma utilidad, pues 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 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 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 $x = x_1$. Numéricamente sería:

$$\text{odds-ratio} = \frac{P(Y \leq ij \mid X = X_2)}{P(Y \geq ij \mid X = X_2)} \frac{P(Y \leq ij \mid X = X_1)}{P(Y \geq ij \mid X = X_1)}$$

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 cual es utilizado para interpretar el efecto de las variables explicativas sobre la variable objeto de estudio. Si este 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 dependiente.

3 DISEÑO METODOLÓGICO

El tipo de investigación aplicado tiene un perfil descriptivo e inferencial de tipo cuantitativo, además se establecieron ciertas fases que permitieron determinar la incidencia de los factores demográficos, socioeconómicos y académicos en el componente genérico para estudiantes del programa de Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama, las cuales fueron:

1. Conformación del marco teórico, el cual partió de la revisión documental sobre la Prueba Saber Pro, antecedentes investigativos, factores que inciden en el componente genérico y lo referente a modelos de regresión logística y Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS).
2. Definición del diseño metodológico, el cual, entre otros aspectos, implica la clasificación de las variables explicativas, además de las técnicas utilizadas.
3. Aplicación de instrumentos, tabulación y depuración de la base de datos.
4. Análisis e interpretación de la información y construcción de modelos que describen aquellos factores que influyen en el componente genérico de la prueba Saber Pro.
5. Conclusiones.

La población objeto de estudio se centró en la Universidad Pedagógica y Tecnológica de Colombia, Seccional Facultad Duitama. Para los resultados de la prueba Saber Pro, además se contó con el registro de 287 estudiantes del programa de Ingeniería Electromecánica de la UPTC que presentaron las pruebas Saber Pro entre el segundo semestre de 2011 a 2015.

Este trabajo monográfico se ejecutó a partir de un diseño bibliográfico; se tomó una base de datos suministradas por el ICFES, a través del taller preparatorio para convocatorias 2016 orientado por el asesor de Gestión de Proyectos de Investigación Luis Eduardo Jaramillo, tomando como referencia la guía (diccionarios) de interpretación de resultados suministradas por el ICFES para entender la codificación de las variables, esta base de datos tiene 287 filas por 58 columnas.

3.1 VARIABLES DE ESTUDIO

Una de las principales utilidades de la estadística en investigación es la caracterización de poblaciones o grupos poblacionales. Toda investigación está orientada por un determinado objetivo, y por tanto, cada una de ellas limita características de la población objeto de estudio y estas deben estar bien definidas. Para ello hacemos uso del concepto de variable.

En estadística podemos decir que una variable es una característica que al ser medida en diferentes individuos (o unidades o sucesos) es susceptible de adoptar diferentes valores. Teniendo en cuenta lo anterior podemos decir que, una variable es un conjunto de valores que clasifica a la población objeto de análisis en distintos grupos (a través de distintos factores); para ello cada variable debe generarse a partir de un único y coherente criterio clasificatorio, es así que para este problema de investigación se ha determinado la siguiente clasificación:

Demográficas:

A esta categoría pertenecen aquellas variables que se refieran a características de la población tales como: sexo, edad, situación del hogar, estado civil, número de personas en el hogar, número de personas a cargo, es cabeza de familia.

Académicas:

A esta categoría pertenecen aquellas variables asociadas al rendimiento académico previo y al esfuerzo mostrado por el estudiante, así como al currículo que sigue el mismo. Así, abarcan aquellos factores educativos que rodean directamente la intervención pedagógica o docente, como: tipo de bachillerato, nivel educativo de la madre y nivel educativo del padre.

Socioeconómicas:

A esta categoría pertenecen aquellas variables asociadas al ingreso económico como: Estrato socioeconómico según recibo de energía, valor anual de la matrícula, tener celular, tener internet, tener servicio de tv, tener computador, tener lavadora, tener teléfono fijo en la residencia, salarios mensuales del grupo familiar, número de dormitorios, tener trabajo, número de horas que trabaja en la semana, ocupación del padre y ocupación de la madre.

A continuación se exponen las variables que se consideraron en este estudio:

Tabla 1. Variables demográficas

VARIABLE		CATEGORÍAS	ESCALA DE MEDICIÓN
Sexo del inscrito	Femenino	F	Cualitativa Nominal
	Masculino	M	
Edad del estudiante		rango:[20-39]	Cuantitativa Razón
Estado civil del estudiante	Soltero	1	Cualitativa Nominal
	Casado	2	
	Viudo	3	
	Separado	4	
	Unión libre	5	
Situación de su hogar actual		Valores posibles	Cualitativa Nominal
Es habitual o permanente		1	
Es temporal por razones de estudio u otra razón		2	
Número de personas que conforman el hogar		rango: [1,12]	Cuantitativa Razón
Es cabeza de familia	No	0	Cualitativa Nominal
	Si	1	
Número de personas de las que usted se encuentra a cargo		rango: [0,5]	Cuantitativa Razón
Estrato de la residencia según factura de energía	Estrato 1	1	Cualitativo Ordinal
	Estrato 2	2	
	Estrato 3	3	
	Estrato 4	4	
	Estrato 5	5	
	Estrato 6	6	
	Zona rural donde no hay estratificación económica		

Tabla 2. Variables socioeconómicas

VARIABLES		CATEGORÍAS	ESCALA DE MEDICIÓN
Ocupación del padre, Ocupación de la madre		Valores posibles	Cualitativas Nominal
Empresarios		1	
Administradores o gerentes		2	
Profesionales independientes		3	
Profesionales empleados		4	
Trabajadores independientes		5	
Trabajadores empleados		6	
Rentistas		7	
Obreros		8	
Jubilados		9	
Hogar		10	
Estudiantes		11	
No devengan ingreso o buscan trabajo		12	
Empresario		13	
Pequeño empresario		14	
Empleado con cargo como director o gerente		15	
Empleado de nivel directivo		16	
Empleado de nivel técnico o profesional		17	
Empleado de nivel auxiliar o administrativo		18	
Empleado obrero u operario		19	
Profesional Independiente		20	
Trabajador por cuenta propia		21	
Hogar		22	
Pensionado		23	
Rentista		24	
Estudiante		25	
Otra actividad u ocupación		26	
No sabe		99	
El hogar cuenta con celular		Valores posibles	Cualitativa Nominal
No		0	
Si		1	
El hogar cuenta con conexión a internet		Valores posibles	Cualitativa Nominal
No		0	
Si		1	
El hogar cuenta con servicio cerrado de televisión		Valores posibles	Cualitativa Nominal
No		0	
Si		1	
El hogar cuenta con lavadora		Valores posibles	Cualitativa Nominal
No		0	
Si		1	

Continuación variables socioeconómicas:

VARIABLE		CATEGORÍAS	ESCALA DE MEDICIÓN
El hogar cuenta con servicio de teléfono fijo	No	Valores posibles	Cualitativa Nominal
	Si	0	
		1	
Trabaja actualmente	No	Valores posibles	Cualitativa Nominal
	Si, con remuneración en dinero o en especie	0	
	Si, como ayudante sin remuneración	1	
	Si, por ser práctica obligatoria del plan de estudios	3	
	Si, por experiencia o recursos en gastos personales	4	
		5	
Horas que trabaja a la semana	No trabaja	Valores posibles	Cualitativa Ordinal
	De 1 a 15 horas	1	
	De 6 a 10 horas	2	
	De 11 a 15 horas	3	
	De 16 a 20 horas	4	
		5	
Número de dormitorios de la residencia		rango: [1,10]	Cuantitativa Razón
Ingresos mensuales en salarios mínimos mensuales	Menos de 1 SM	Valores posibles	Cualitativa Ordinal
	Entre 1 y Menos de 2 SM	1	
	Entre 2 y Menos de 3 SM	2	
	Entre 3 y Menos de 5 SM	3	
	Entre 5 y Menos de 7 SM	4	
	Entre 7 y Menos de 10 SM	5	
	10 o ms SM	6	
		7	
Forma pagó matrícula	pagó por padres	valores posibles	Cualitativo Nominal
	Varias formas de pagó	0	
	No cancelo matrícula	1	
	pagó por beca	2	
	pagó por credito	3	
	pagó propio	4	
		5	
Valor anual de la matrícula del año anterior	No pagó matrícula	valores posibles	Cualitativo Ordinal
	Menos de 500 mil	0	
	Entre 500 mil y menos de 1 millón	1	
	Entre 1 millón y 3 millones	2	
	Entre 3 millones y 5 millones	3	
	Más de 5 millones	4	
		5	

Tabla 3. Variables Académicas

VARIABLE	CATEGORÍAS	ESCALA DE MEDICIÓN
Máximo nivel educativo alcanzado por del padre; madre	Valores posibles	Cualitativa Ordinal
Ninguno	0	
No tuvo escuela	1	
Preescolar	2	
Básica primaria	3	
Básica secundaria	4	
Media vocacional	5	
Tecnológico o técnico	6	
Universitario	7	
Postgrado	8	
Primaria incompleta	9	
Primaria completa	10	
Secundaria (bachillerato) incompleta	11	
Secundaria (bachillerato) completa	12	
Educación técnica o tecnológica incompleta	13	
Educación técnica o tecnológica completa	14	
Educación profesional incompleta	15	
Educación profesional completa	16	
Postgrado	17	
No sabe	99	
Tomó algún curso de preparación para el examen	valores posibles	Cualitativa Nominal
No	0	
Si	1	
Bachillerato del que se graduó	valores posibles	Cualitativo Nominal
Académico	A	
Técnico	N	
Normalista Superior	T	
Desconocido	D	

3.2 MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS

Para este trabajo se utilizó técnicas de tipo descriptivo, pruebas de diferencia de medias, pruebas de correlación, tablas de contingencia, pruebas de bondad de ajuste y técnicas de modelamiento como: GAMLSS y modelo multinomial ordinal. Para el procesamiento de los datos se utilizó el software libre R como procesador estadístico versión 3.4.1, como uno de los lenguajes más utilizados en investigación por la comunidad estadística, con sus paquetes: gamlss, MASS, gamlss.dist, gamlss.data, splines, nlme, parallel y Rcmdr.

4 DESCRIPCIÓN DE LOS DATOS

A continuación se describe el desempeño de los estudiantes del Programa de Ingeniería Electromecánica en las Pruebas Saber Pro. En primer lugar, se describe los resultados, entre los años 2011 (segundo semestre) y el año 2015, discriminados por competencias. Posteriormente se presentan los resultados del año 2016. La información tomada para desarrollar las secciones 1.1 y 1.2 fue tomada de la página <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados> asociada al ICFES. La información que sustenta la sección 1.3 se presenta en el Anexo 1- Resultados IE año 2016- y la información que sustenta la sección 1.4 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 debe desarrollar todo estudiante, independiente de su formación profesional, en ellas se evalúa inglés, razonamiento cuantitativo, lectura crítica, comunicación escrita y competencias ciudadanas.

La siguiente gráfica muestra el resultado promedio por año para las competencias de razonamiento cuantitativo, inglés, comunicación escrita y lectura crítica.

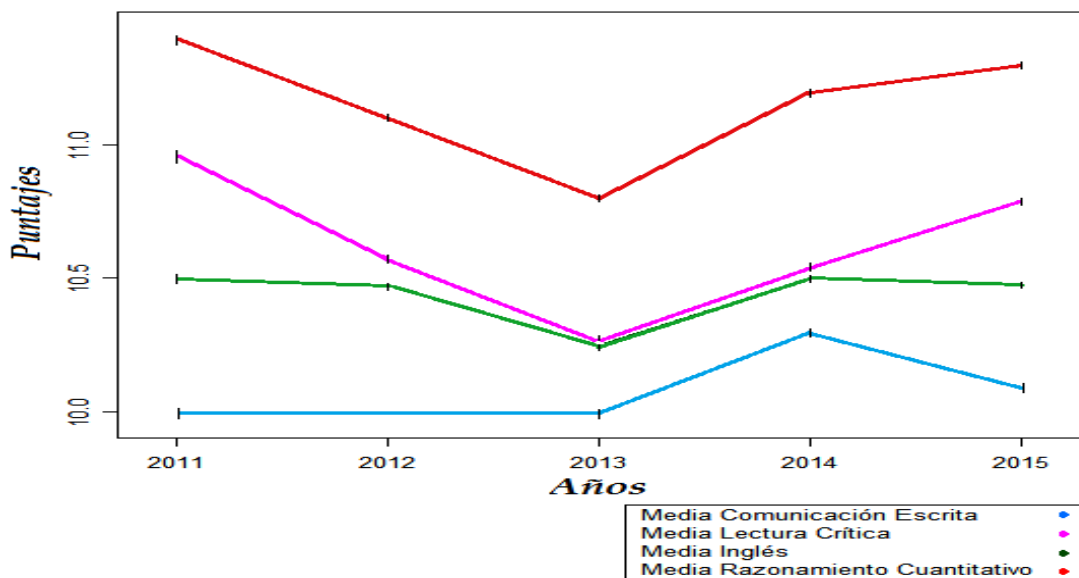


Figura 4-1: Resultados promedio por año y competencia

Podemos observar que los puntajes de los estudiantes de Ingeniería Electromecánica en los niveles del componente genérico han mostrado un descenso a partir del año 2011 y hasta el 2013, para en los dos años siguientes marcar una paulatina mejora, con excepción de comunicación escrita que demuestra un nivel constante aunque siempre en el rango de regular.

Comunicación escrita obtuvo en estos años los puntajes promedio más bajos, indicando que los estudiantes presentan dificultades a la hora de comunicar ideas por escrito con base en el análisis de información suministrada, mientras que razonamiento cuantitativo es el componente que en promedio demuestra un nivel de muy bueno y excelente a través de este periodo de tiempo.

Por otra parte el desempeño en Lectura crítica permanece en un rango de puntajes entre bueno y muy bueno, mientras inglés se sitúa en el rango de bueno, es decir que los estudiantes comprenden y utilizan expresiones cotidianas de uso frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato.

Los resultados de cada módulo referentes a la UPTC se presentan en una tabla, la cual incluye de los resultados correspondientes al Programa, siendo n , el número de estudiantes que presentaron la prueba, \bar{x} : puntaje promedio individual obtenido, DE: desviación estándar, CA: coeficiente de asimetría, CC: coeficiente de curtosis, CV: coeficiente de variación, Q_1 , Q_2 y Q_3 representan los cuartiles uno, dos y tres respectivamente y NA se refiere al número de datos faltantes.

Tabla 4. Resultados Comunicación Escrita

Año	n	\bar{x}	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	35	10.0	0.24	0.02	0.08	9.4	9.8	10.4	0
2012	66	10.0	-0.7	1.4	0.11	9.6	10.0	10.7	0
2013	67	10.0	-1.01	3.0	0.08	9.5	9.9	10.5	1
2014	60	10.3	0.01	-0.61	0.08	9.7	10.2	11.3	1
2015	57	10.1	-4.28	26.9	0.16	9.5	10.1	10.9	0

Se observa que, para cada uno de los años que ha presentado la prueba, los estudiantes de Ingeniería Electromecánica presentan puntajes similares (coeficiente de variación del orden del 11%), con excepción del año 2015. A juzgar por los valores de los coeficientes de curtosis los puntajes en **COMUNICACIÓN ESCRITA** indican que no hay estudiantes que hayan obtenido puntajes muy diferentes entre los años 2011 y 2014, es decir los resultados se concentran alrededor de su media para cada año, exceptuando el año 2015 donde se indica la presencia de datos atípicos; por lo anterior y teniendo en cuenta que el promedio es de 10.08 para esta prueba, podemos concluir que a través de los años el promedio de puntajes se ubica en la categoría regular, lo cual apunta a que los estudiantes de Ingeniería Electromecánica presentan dificultad

para comunicar ideas por escrito con base en el análisis de la información suministrada. Los resultados anteriores se pueden complementar con los resultados cualitativos en la prueba de **DESEMPEÑO EN COMUNICACIÓN ESCRITA**, Tabla 2.

Tabla 5. Resultados en Comunicación Escrita

<i>Año</i>	<i>Nivel</i> 1	<i>Nivel</i> 2	<i>Nivel</i> 3	<i>Nivel</i> 4	<i>Nivel</i> 5	<i>Nivel</i> 6	<i>Nivel</i> 7	<i>Nivel</i> 8	<i>n</i>	<i>NA</i>
2011	0%	2.8%	17.1%	48.6%	20%	8.6%	2.8%	0%	35	0
2012	4.5%	1.5%	10.6%	39.4%	24.2%	16.7%	3.03%	0%	66	0
2013	2.9%	1.5%	1.5%	52.2%	32.8%	9%	0%	0%	67	0
2014	0%	0%	8.3%	40%	25%	25%	1.7%	0%	60	1
2015	0%	0%	7.1%	44.6%	35.7%	7.1%	5.4%	0%	56	1

Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes (44.72%) se ubican en el nivel 4 de desempeño, indicando que los estudiantes encuentran la idea central que se desarrolla de acuerdo con una intención comunicativa, hallando también la estructura básica, en otras palabras, identifican la introducción al tema que se abordará, un desarrollo y una conclusión. Sin embargo, como el texto no incluye toda la información necesaria (progresión temática), la organización no es completamente efectiva, o rompen 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 28.1% de los estudiantes se ubican en el nivel 5 y pocos (13.3%) en el nivel 6.

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 estudiantes de Ingeniería Electromecánica han sido heterogéneos respecto a los puntajes (los coeficientes de variación, oscilan entre el 8% y el 16%). Además los coeficientes de curtosis indican que existen estudiantes que obtuvieron puntajes muy diferentes a los del grupo con los que presentaron la prueba, exceptuando el año 2014.

Por lo tanto sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 10.43, se puede indicar que los puntajes a través de los años los ubica en la categoría de bueno.

Tabla 6. Resultados Módulo Inglés

<i>Año</i>	<i>n</i>	\bar{x}	DE	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	35	10.5	1.08	2.2	7.15	0.10	9.8	10.3	10.9	0
2012	66	10.47	1.14	1.6	3.12	0.11	9.9	10.1	10.8	0
2013	68	10.25	1.20	1.9	4.3	0.11	9.5	10	10.5	0
2014	61	10.5	0.93	0.8	-0.04	0.08	9.9	10.2	11.1	0
2015	57	10.47	1.77	-3.4	21.9	0.16	9.9	10.3	11.2	0

Ahora bien, sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes se ubican en el nivel A2 de **DESEMPEÑO EN INGLÉS**, indicando que los estudiante son capaces de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes como información básica sobre ellos mismos y su familia, compras, lugares de interés, ocupaciones, etc., como se muestra en la Tabla 7.

Tabla 7. Resultados de Desempeño en Inglés

DESEMPEÑO EN INGLÉS	A-	A1	A2	B+	B1	<i>n</i>	NA
2011	11.4%	25.7%	28.6%	5.7%	28.57%	35	0
2012	10.6%	30.3%	33.3%	9.1%	16.7%	66	0
2013	22.1%	29.4%	26.5%	7.3%	14.7%	68	0
2014	4.9%	31.1%	27.9%	9.8%	26.2%	61	0
2015	5.3%	24.6%	29.8%	7.0%	33.3%	57	0

La Tabla 8 muestra la descripción por año de los resultados obtenidos por los estudiantes de Ingeniería Electromecánica para el componente en **LECTURA CRÍTICA**, en la cual se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Ingeniería Electromecánica son muy homogéneos (coeficientes de variación, son inferiores al 10%); excepto en el año 2015, donde se observa heterogeneidad en los puntajes. 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. Los valores indican que entre el año 2011 y el 2015 el puntaje promedio de los estudiantes ha estado entre 10.26 y 10.96, lo cual los ubica en las categorías de bueno y muy bueno en la escala, por lo tanto, los estudiantes de Ingeniería Electromecánica

tienen la capacidad para leer de manera analítica y reflexiva, además de comprender los planteamientos expuestos en un texto exigiendo que el lector identifique y recupere información presente en uno o varios textos, construir un sentido global, establecer relaciones entre enunciados y evaluar su intencionalidad.

Tabla 8. Resultados de Lectura Crítica

<i>Año</i>	<i>n</i>	\bar{x}	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	<i>NA</i>
2011	35	10.96	0.91	1.41	0.07	10.5	10.8	11.4	0
2012	66	10.57	0.16	-0.47	0.06	10.2	10.6	10.9	0
2013	68	10.26	0.61	0.72	0.07	9.8	10.2	10.7	0
2014	61	10.54	1.22	3.54	0.08	9.8	10.4	11.0	0
2015	57	10.79	-5.52	37.5	0.14	10.5	10.9	11.4	0

La Tabla 9 muestra la descripción por año de los resultados obtenidos por los estudiantes de Ingeniería Electromecánica para el componente en **RAZONAMIENTO CUANTITATIVO**, en la cual se tiene que, para el periodo comprendido entre los años 2011 y 2014, los puntajes de los estudiantes de Ingeniería Electromecánica son muy homogéneos (coeficientes de variación, por debajo del 10%), además al juzgar los valores de los coeficientes de curtosis los puntajes en **RAZONAMIENTO CUANTITATIVO** indican que no hay estudiantes que hayan obtenido puntajes muy diferentes, es decir los resultados se concentran alrededor de su media para cada año. Ahora bien, para el año 2015 se presentó un alto coeficiente de variación y de curtosis, indicando que existieron puntajes que superaron la media (cerca del 50%), confirmado por los cuartiles. Por lo anterior se tiene que el promedio general de la prueba es de 11.1 concluyendo que a través de los años se ubica en la categoría de excelente, lo cual apunta a que los estudiantes desarrollan en buena medida competencias relacionadas con las habilidades en la comprensión de conceptos básicos de las matemáticas para analizar, modelar y resolver problemas aplicando métodos y procedimientos cuantitativos y esquemáticos.

Tabla 9. Resultados Razonamiento Cuantitativo

<i>Año</i>	<i>n</i>	\bar{x}	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	<i>NA</i>
2011	35	11.4	-0.05	-0.27	0.06	10.8	11.4	12.1	0
2012	66	11.1	0.38	0.8	0.08	10.5	10.9	11.6	0
2013	68	10.8	0.41	0.16	0.08	10.2	10.8	11.5	0
2014	61	11.2	0.57	-0.6	0.08	10.6	10.9	11.9	0
2015	57	11.3	-4.4	28.3	0.16	10.8	11.4	11.7	1

Ahora bien, para el año 2016 la prueba Saber Pro cambió la forma de evaluación respecto a los puntajes, es decir, se califica en un rango de 0 a 300 (cero a trescientos), razón por la cual se describen los resultados de la siguiente manera:

4.2 RESULTADOS EN COMPETENCIAS GENÉRICAS AÑO 2016

Para el año 2016, el puntaje global del programa fue de 164 puntos con una desviación estándar de 16 puntos, esto indica que los estudiantes de Ingeniería Electromecánica fueron homogéneos (coeficiente de variación 9.91%) en la prueba y que al menos el 75% de ellos obtuvo entre 136 y 178 puntos. La media en el grupo de referencia de Ingeniería del País fue de 156 puntos con una desviación de 22. Es decir, al menos el 75% de los estudiantes obtuvo entre 115 y 205 puntos. Lo cual implica que Ingeniería Electromecánica tiene un comportamiento similar al grupo de referencia del País.

En competencias genéricas se tiene que en **COMUNICACIÓN ESCRITA** el puntaje del Programa (150 (ds=25)) estuvo por debajo de los puntajes de la Institución, la Sede y el grupo de referencia de Ingenierías del País. La mayoría de estudiantes del Programa (48.48%) se ubica en nivel de **DESEMPEÑO** tres 3, lo que indica que superan las preguntas de cierta complejidad del examen y los ubica en un uso bueno del lenguaje, identificando un planteamiento o posición personal que va más allá de una opinión aislada sobre el tema, sin embargo el texto presentado presenta algunas contradicciones, digresiones y repeticiones.

En **RAZONAMIENTO CUANTITATIVO**, el Programa fue muy superior (189 (ds=19)) respecto al grupo de referencia de Ingenierías del país, a la Institución y comparado a los programas de la Sede. La mayoría de estudiantes del Programa (69.69%) se ubica en nivel de **DESEMPEÑO** tres 3, seguido de un grupo de estudiantes (24.24%) que se encuentra en un nivel de desempeño 4, lo que indica un desempeño muy bueno en las competencias exigibles para el examen.

Es conveniente mencionar que los estudiantes de Ingeniería Electromecánica en su gran mayoría extraen información implícita contenida en representaciones no usuales asociadas a una misma situación y provenientes de una única fuente de información, argumentan la validez de procedimientos, y resuelven problemas utilizando modelos que combinan procedimientos aritméticos, algebraicos, variacionales y aleatorios.

En **LECTURA CRÍTICA** y **COMPETENCIAS CIUDADANAS** el rendimiento del programa fue similar con los de la Institución, la Sede y las licenciaturas del País. **EN LECTURA CRÍTICA** la mayoría de estudiantes del Programa (63.63%) se ubica en nivel de **DESEMPEÑO** 3, lo que indica que superan las preguntas de exigencia media del examen, mostrando que los

estudiantes reconocen la macro estructura del texto e identifican la tipología textual, las estrategias discursivas, y reconocen las funciones del lenguaje para comprender el sentido del texto.

En **COMPETENCIAS CIUDADANAS** la mayoría de estudiantes del Programa (54.54%) se ubica en nivel de **DESEMPEÑO 2**, es decir, muestran un desempeño aceptable en las competencias exigibles para el examen. Esto muestra que los estudiantes no reconocen argumentos implícitos en un enunciado o afirmación sin establecer una relación entre la afirmación y la cosmovisión, y casi no analizan y comparan propuestas de solución a un problema.

En **INGLÉS** la mayoría de estudiantes del Programa (66%) se ubica entre los niveles de **DESEMPEÑO A1** y **A2**, es decir, los estudiantes apenas son capaces de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. Le cuesta presentarse él mismo y ante otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce. Además se le dificulta relacionarse de forma elemental con un interlocutor.

Así mismo en **DESEMPEÑO EN COMUNICACIÓN ESCRITA** el programa estuvo muy por debajo del promedio respecto al grupo de referencia, la institución y comparado con los demás programas de la Sede. La mayoría de estudiantes del Programa (48%) se ubica en nivel de **DESEMPEÑO 3**, lo que indica que superan las preguntas de mediana complejidad del examen y los ubica en un buen uso del lenguaje, identificando un planteamiento o posición personal que va más allá de una opinión aislada sobre el tema.

En cuanto a competencias específicas, el módulo de profundización en ingenierías lo conforman tres aspectos: **Diseño de sistemas mecánicos, Formulación de proyectos de ingeniería y Pensamiento científico**. A continuación se presentan los resultados obtenidos por los estudiantes de Ingeniería Electromecánica:

Los puntajes para los módulos de **Diseño de sistemas mecánicos** y **formulación de proyectos** de Ingeniería, fueron de 158 con desviación estándar de 24 puntos y 155 con desviación estándar de 31 puntos respectivamente, lo cual implica que los estudiantes de Ingeniería fueron homogéneos (Coeficientes de variación inferiores al 10%) en la prueba. Al comparar estos promedios con el módulo de Ingenierías en el país, los de la Institución y la Sede se tienen que los promedios están por debajo. Por otra parte el módulo de **pensamiento científico** cuyo puntaje fue 169 con desviación estándar de 27, estuvo por encima del promedio del módulo de ingenierías del país, la institución y la sede.

4.3 DESCRIPCIÓN DE LOS ESTUDIANTES DE INGENIERÍA ELECTROMECAÁNICA

A continuación se presenta la descripción de las variables explicativas que se tuvieron en cuenta para este trabajo monográfico y su respectiva clasificación.

FACTORES DE ESTUDIO DEMOGRÁFICOS					
DESCRIPCIÓN		RANGO		RESUMEN	
Sexo del inscrito	<i>Femenino</i>	Valores Posibles: 2	<i>F</i>	<i>F=10.45%</i>	<i>M=89.55%</i>
	<i>Masculino</i>		<i>M</i>		
Edad del estudiante		Rango:[20-43]		<i>n=287</i>	
				<i>$\bar{x}=24.54$</i>	
				<i>De=3.12</i>	
				<i>Cv=0.12</i>	
				<i>Cu=8.75</i>	
				<i>Q₁=23</i>	
				<i>Q₂=24</i>	
				<i>Q₃=26</i>	
Estado civil del estudiante	<i>Soltero</i>	Valores posibles: 5	<i>1</i>	<i>1=95.12%</i>	<i>2=0.70%</i>
	<i>Casado</i>		<i>2</i>		
	<i>Viudo</i>		<i>3</i>		
	<i>Separado</i>		<i>4</i>		
	<i>Unión libre</i>		<i>5</i>		
Situación de su hogar actual		Valores posibles:2		<i>1=70.73%</i>	
<i>Es habitual o permanente</i>		<i>1</i>		<i>2=29.27%</i>	
<i>Es temporal por razones de estudio</i>		<i>2</i>			

De los 287 estudiantes que presentaron la prueba la gran mayoría son hombres, entre ellos predomina ser solteros, con una edad promedio de 25 años y tan solo 84 personas (29%) residen en un hogar temporal por razones de estudio u otras razones.

Continuación de los factores de estudio demográficos.

FACTORES DE ESTUDIO DEMOGRÁFICOS										
DESCRIPCIÓN	RANGO	RESUMEN								
Número de personas que conforman el hogar	Rango:[1,10]	<table border="1"> <tr> <td>$n=287$</td> </tr> <tr> <td>$\bar{x} = 4.46$</td> </tr> <tr> <td>$De=1.53$</td> </tr> </table>	$n=287$	$\bar{x} = 4.46$	$De=1.53$					
$n=287$										
$\bar{x} = 4.46$										
$De=1.53$										
<table border="1"> <tr> <td>Es cabeza de familia</td> </tr> <tr> <td><i>No</i></td> </tr> <tr> <td><i>Si</i></td> </tr> </table>	Es cabeza de familia	<i>No</i>	<i>Si</i>	<table border="1"> <tr> <td>Valores Posibles:2</td> </tr> <tr> <td><i>0</i></td> </tr> <tr> <td><i>1</i></td> </tr> </table>	Valores Posibles:2	<i>0</i>	<i>1</i>	<table border="1"> <tr> <td>$0=90.24\%$</td> </tr> <tr> <td>$1=9.76\%$</td> </tr> </table>	$0=90.24\%$	$1=9.76\%$
Es cabeza de familia										
<i>No</i>										
<i>Si</i>										
Valores Posibles:2										
<i>0</i>										
<i>1</i>										
$0=90.24\%$										
$1=9.76\%$										
Número de personas que usted tiene a cargo	Valores posibles: [0,4]	<table border="1"> <tr> <td>$n=287$</td> </tr> <tr> <td>$\bar{x} = 0.25$</td> </tr> <tr> <td>$De= 0.70$</td> </tr> </table>	$n=287$	$\bar{x} = 0.25$	$De= 0.70$					
$n=287$										
$\bar{x} = 0.25$										
$De= 0.70$										

En promedio, el número de personas que conforman el núcleo familiar de los estudiantes de Ingeniería Electromecánica es de 4 personas, además solo un pequeño porcentaje de los estudiantes son cabeza de familia (9.76%), por lo tanto se deduce que la mayor parte de ellos no tiene ninguna persona a cargo.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																									
DESCRIPCIÓN	RANGO	RESUMEN																							
<table border="1"> <tr> <td>Estrato residencia según factura de energía</td> </tr> <tr> <td><i>Estrato 1</i></td> </tr> <tr> <td><i>Estrato 2</i></td> </tr> <tr> <td><i>Estrato 3</i></td> </tr> <tr> <td><i>Estrato 4</i></td> </tr> <tr> <td><i>Estrato 5</i></td> </tr> <tr> <td><i>Estrato 6</i></td> </tr> <tr> <td><i>Zona rural sin estratificación</i></td> </tr> </table>	Estrato residencia según factura de energía	<i>Estrato 1</i>	<i>Estrato 2</i>	<i>Estrato 3</i>	<i>Estrato 4</i>	<i>Estrato 5</i>	<i>Estrato 6</i>	<i>Zona rural sin estratificación</i>	<table border="1"> <tr> <td>Rango:[1,7]</td> </tr> <tr> <td><i>1</i></td> </tr> <tr> <td><i>2</i></td> </tr> <tr> <td><i>3</i></td> </tr> <tr> <td><i>4</i></td> </tr> <tr> <td><i>5</i></td> </tr> <tr> <td><i>6</i></td> </tr> <tr> <td><i>7</i></td> </tr> </table>	Rango:[1,7]	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<table border="1"> <tr> <td><i>1=9.76%</i></td> </tr> <tr> <td><i>2=65.51%</i></td> </tr> <tr> <td><i>3=23%</i></td> </tr> <tr> <td><i>4=1.74%</i></td> </tr> <tr> <td><i>5=0.0%</i></td> </tr> <tr> <td><i>6=0.0%</i></td> </tr> <tr> <td><i>7=0.0%</i></td> </tr> </table>	<i>1=9.76%</i>	<i>2=65.51%</i>	<i>3=23%</i>	<i>4=1.74%</i>	<i>5=0.0%</i>	<i>6=0.0%</i>	<i>7=0.0%</i>
Estrato residencia según factura de energía																									
<i>Estrato 1</i>																									
<i>Estrato 2</i>																									
<i>Estrato 3</i>																									
<i>Estrato 4</i>																									
<i>Estrato 5</i>																									
<i>Estrato 6</i>																									
<i>Zona rural sin estratificación</i>																									
Rango:[1,7]																									
<i>1</i>																									
<i>2</i>																									
<i>3</i>																									
<i>4</i>																									
<i>5</i>																									
<i>6</i>																									
<i>7</i>																									
<i>1=9.76%</i>																									
<i>2=65.51%</i>																									
<i>3=23%</i>																									
<i>4=1.74%</i>																									
<i>5=0.0%</i>																									
<i>6=0.0%</i>																									
<i>7=0.0%</i>																									
<table border="1"> <tr> <td>Valor anual de la matrícula del año anterior</td> </tr> <tr> <td><i>No pagó matrícula</i></td> </tr> <tr> <td><i>Menos de 500 mil</i></td> </tr> <tr> <td><i>Entre 500 mil y menos de 1 millón</i></td> </tr> <tr> <td><i>Entre 1 millón y 3 millones</i></td> </tr> <tr> <td><i>Entre 3 millones y 5 millones</i></td> </tr> <tr> <td><i>Más de 5 millones</i></td> </tr> </table>	Valor anual de la matrícula del año anterior	<i>No pagó matrícula</i>	<i>Menos de 500 mil</i>	<i>Entre 500 mil y menos de 1 millón</i>	<i>Entre 1 millón y 3 millones</i>	<i>Entre 3 millones y 5 millones</i>	<i>Más de 5 millones</i>	<table border="1"> <tr> <td>Valores posibles: 6</td> </tr> <tr> <td><i>0</i></td> </tr> <tr> <td><i>1</i></td> </tr> <tr> <td><i>2</i></td> </tr> <tr> <td><i>3</i></td> </tr> <tr> <td><i>4</i></td> </tr> <tr> <td><i>5</i></td> </tr> </table>	Valores posibles: 6	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<table border="1"> <tr> <td><i>0=1.05%</i></td> </tr> <tr> <td><i>1=59.23%</i></td> </tr> <tr> <td><i>2=32.75%</i></td> </tr> <tr> <td><i>3=6.62%</i></td> </tr> <tr> <td><i>4=0.35%</i></td> </tr> <tr> <td><i>5=0%</i></td> </tr> </table>	<i>0=1.05%</i>	<i>1=59.23%</i>	<i>2=32.75%</i>	<i>3=6.62%</i>	<i>4=0.35%</i>	<i>5=0%</i>			
Valor anual de la matrícula del año anterior																									
<i>No pagó matrícula</i>																									
<i>Menos de 500 mil</i>																									
<i>Entre 500 mil y menos de 1 millón</i>																									
<i>Entre 1 millón y 3 millones</i>																									
<i>Entre 3 millones y 5 millones</i>																									
<i>Más de 5 millones</i>																									
Valores posibles: 6																									
<i>0</i>																									
<i>1</i>																									
<i>2</i>																									
<i>3</i>																									
<i>4</i>																									
<i>5</i>																									
<i>0=1.05%</i>																									
<i>1=59.23%</i>																									
<i>2=32.75%</i>																									
<i>3=6.62%</i>																									
<i>4=0.35%</i>																									
<i>5=0%</i>																									
<table border="1"> <tr> <td>Forma de pago de matrícula</td> </tr> <tr> <td><i>Pago por padres</i></td> </tr> <tr> <td><i>Varias formas de pago</i></td> </tr> <tr> <td><i>No canceló matrícula</i></td> </tr> <tr> <td><i>Pago por beca</i></td> </tr> <tr> <td><i>Pago por crédito</i></td> </tr> <tr> <td><i>Pago propio</i></td> </tr> </table>	Forma de pago de matrícula	<i>Pago por padres</i>	<i>Varias formas de pago</i>	<i>No canceló matrícula</i>	<i>Pago por beca</i>	<i>Pago por crédito</i>	<i>Pago propio</i>	<table border="1"> <tr> <td>Valores posibles: 6</td> </tr> <tr> <td><i>0</i></td> </tr> <tr> <td><i>1</i></td> </tr> <tr> <td><i>2</i></td> </tr> <tr> <td><i>3</i></td> </tr> <tr> <td><i>4</i></td> </tr> <tr> <td><i>5</i></td> </tr> </table>	Valores posibles: 6	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<table border="1"> <tr> <td><i>0=58%</i></td> </tr> <tr> <td><i>1=6%</i></td> </tr> <tr> <td><i>2=2%</i></td> </tr> <tr> <td><i>3=4%</i></td> </tr> <tr> <td><i>4=3%</i></td> </tr> <tr> <td><i>5=31%</i></td> </tr> </table>	<i>0=58%</i>	<i>1=6%</i>	<i>2=2%</i>	<i>3=4%</i>	<i>4=3%</i>	<i>5=31%</i>			
Forma de pago de matrícula																									
<i>Pago por padres</i>																									
<i>Varias formas de pago</i>																									
<i>No canceló matrícula</i>																									
<i>Pago por beca</i>																									
<i>Pago por crédito</i>																									
<i>Pago propio</i>																									
Valores posibles: 6																									
<i>0</i>																									
<i>1</i>																									
<i>2</i>																									
<i>3</i>																									
<i>4</i>																									
<i>5</i>																									
<i>0=58%</i>																									
<i>1=6%</i>																									
<i>2=2%</i>																									
<i>3=4%</i>																									
<i>4=3%</i>																									
<i>5=31%</i>																									

Según la factura de energía que llega a la residencia familiar de los estudiantes se clasifica el estrato socioeconómico de los mismos, aquí se observa que la mayoría pertenecen al estrato 2. Respecto al valor anual de la matrícula, los estudiantes (170 personas) pagan menos de 500 mil pesos y la forma de pago es a través de los padres (Es importante mencionar que en la variable forma de pago existen varios datos faltantes).

Continuación de los factores de estudio socioeconómicos.

FACTORES DE ESTUDIO SOCIOECONÓMICOS			
DESCRIPCIÓN	RANGO	RESUMEN	
Ocupación del padre; madre	Rango:[1,26]		
<i>Empresarios</i>	1	1=0%	1=0%
<i>Administradores o gerentes</i>	2	2=0%	2=0%
<i>Profesionales independientes</i>	3	3=0%	3=0%
<i>Profesionales empleados</i>	4	4=0%	4=0%
<i>Trabajadores independientes</i>	5	5=0%	5=0%
<i>Trabajadores empleados</i>	6	6=0%	6=0%
<i>Rentistas</i>	7	7=0%	7=0%
<i>Obreros</i>	8	8=0%	8=0%
<i>Jubilados</i>	9	9=0%	9=0%
<i>Hogar</i>	10	10=0%	10=0%
<i>Estudiantes</i>	11	11=0%	11=0%
<i>No devengan ingreso o buscan trabajo</i>	12	12=0%	12=0%
<i>Empresario</i>	13	13=0.69%	13=0%
<i>Pequeño empresario</i>	14	14=9.06%	14=4.88%
<i>Empleado con cargo como empleador o gerente</i>	15	15=0.35%	15=0%
<i>Empleado de nivel directivo</i>	16	16=2.44%	16=1.40%
<i>Empleado de nivel técnico o profesional</i>	17	17=9.40%	17=3.83%
<i>Empleado de nivel auxiliar o administrativo</i>	18	18=1.74%	18=3.83%
<i>Empleado obrero u operario</i>	19	19=15.33%	19=7.66%
<i>Profesional independiente</i>	20	20=4.18%	20=3.13%
<i>Trabajador por cuenta propia</i>	21	21=34.14%	21=9.04%
<i>Hogar</i>	22	22=0.69%	22=55.05%
<i>Pensionado</i>	23	23=10.80%	23=2.44%
<i>Rentista</i>	24	24=11.14%	24=0%
<i>Estudiante</i>	25	25=0%	25=0%
<i>Otra actividad u ocupación</i>	26	26=0%	26=5.57%
<i>No sabe</i>	99	99=0%	99=0%

Al analizar la ocupación de los padres de los estudiantes de ingeniería electromecánica que presentaron la prueba saber pro entre el año 2011-2 y el 2015 se tiene que, en su mayoría son trabajadores por cuenta propia, es decir que tienen un ingreso asociado a alguna actividad comercial de compra y/o venta de servicios. En el caso de la ocupación de la madre, la mayoría están vinculadas a la labor doméstica en sus hogares.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																
DESCRIPCIÓN	RANGO	RESUMEN														
<table border="1"> <tr><td>El estudiante tiene celular</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El estudiante tiene celular	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=5.23%</td></tr> <tr><td><i>1</i>=94.77%</td></tr> </table>	<i>0</i> =5.23%	<i>1</i> =94.77%						
El estudiante tiene celular																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =5.23%																
<i>1</i> =94.77%																
<table border="1"> <tr><td>El hogar cuenta con servicio de internet</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio de internet	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=61.32%</td></tr> <tr><td><i>1</i>=38.68%</td></tr> </table>	<i>0</i> =61.32%	<i>1</i> =38.68%						
El hogar cuenta con servicio de internet																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =61.32%																
<i>1</i> =38.68%																
<table border="1"> <tr><td>El hogar cuenta con servicio cerrado de televisión</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio cerrado de televisión	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=55.75%</td></tr> <tr><td><i>1</i>=44.25%</td></tr> </table>	<i>0</i> =55.75%	<i>1</i> =44.25%						
El hogar cuenta con servicio cerrado de televisión																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =55.75%																
<i>1</i> =44.25%																
<table border="1"> <tr><td>El hogar cuenta con servicio de teléfono fijo</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio de teléfono fijo	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=68.29%</td></tr> <tr><td><i>1</i>=31.71%</td></tr> </table>	<i>0</i> =68.29%	<i>1</i> =31.71%						
El hogar cuenta con servicio de teléfono fijo																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =68.29%																
<i>1</i> =31.71%																
<table border="1"> <tr><td>El hogar cuenta con lavadora</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con lavadora	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=35.54%</td></tr> <tr><td><i>1</i>=64.46%</td></tr> </table>	<i>0</i> =35.54%	<i>1</i> =64.46%						
El hogar cuenta con lavadora																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =35.54%																
<i>1</i> =64.46%																
<table border="1"> <tr><td>Trabaja actualmente</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si, con remuneración</i></td></tr> <tr><td><i>Si, sin remuneración</i></td></tr> <tr><td><i>Si, por ser práctica de estudios</i></td></tr> </table>	Trabaja actualmente	<i>No</i>	<i>Si, con remuneración</i>	<i>Si, sin remuneración</i>	<i>Si, por ser práctica de estudios</i>	<table border="1"> <tr><td>Valores Posibles: 4</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> <tr><td><i>2</i></td></tr> <tr><td><i>3</i></td></tr> </table>	Valores Posibles: 4	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<table border="1"> <tr><td><i>0</i>=51.22%</td></tr> <tr><td><i>1</i>=9.41%</td></tr> <tr><td><i>2</i>=39.02%</td></tr> <tr><td><i>3</i>=0.35%</td></tr> </table>	<i>0</i> =51.22%	<i>1</i> =9.41%	<i>2</i> =39.02%	<i>3</i> =0.35%
Trabaja actualmente																
<i>No</i>																
<i>Si, con remuneración</i>																
<i>Si, sin remuneración</i>																
<i>Si, por ser práctica de estudios</i>																
Valores Posibles: 4																
<i>0</i>																
<i>1</i>																
<i>2</i>																
<i>3</i>																
<i>0</i> =51.22%																
<i>1</i> =9.41%																
<i>2</i> =39.02%																
<i>3</i> =0.35%																

Se puede establecer que la mayor parte de los estudiantes tienen celular y lavadora, no obstante pocos cuentan con servicio de internet en sus hogares y no tienen teléfono fijo. Respecto a su situación laboral, un poco más de la mitad de los estudiantes no trabaja, y si lo hacen es por tener experiencia.

Continuación de los factores de estudio socioeconómicos.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																									
DESCRIPCIÓN	RANGO	RESUMEN																							
<table border="1"> <tr> <td>Horas que trabaja a la semana</td> </tr> <tr> <td><i>No trabaja</i></td> </tr> <tr> <td><i>De 1 a 5 horas</i></td> </tr> <tr> <td><i>De 6 a 10 horas</i></td> </tr> <tr> <td><i>De 11 a 15 horas</i></td> </tr> <tr> <td><i>De 16 a 20 horas</i></td> </tr> </table>	Horas que trabaja a la semana	<i>No trabaja</i>	<i>De 1 a 5 horas</i>	<i>De 6 a 10 horas</i>	<i>De 11 a 15 horas</i>	<i>De 16 a 20 horas</i>	<table border="1"> <tr> <td>Valores Posibles: 5</td> </tr> <tr> <td>1</td> </tr> <tr> <td>2</td> </tr> <tr> <td>3</td> </tr> <tr> <td>4</td> </tr> <tr> <td>5</td> </tr> </table>	Valores Posibles: 5	1	2	3	4	5	<table border="1"> <tr> <td>1=51.57%</td> </tr> <tr> <td>2=2.44%</td> </tr> <tr> <td>3=10.80%</td> </tr> <tr> <td>4=8.01%</td> </tr> <tr> <td>5=27.18%</td> </tr> </table>	1=51.57%	2=2.44%	3=10.80%	4=8.01%	5=27.18%						
Horas que trabaja a la semana																									
<i>No trabaja</i>																									
<i>De 1 a 5 horas</i>																									
<i>De 6 a 10 horas</i>																									
<i>De 11 a 15 horas</i>																									
<i>De 16 a 20 horas</i>																									
Valores Posibles: 5																									
1																									
2																									
3																									
4																									
5																									
1=51.57%																									
2=2.44%																									
3=10.80%																									
4=8.01%																									
5=27.18%																									
<p>Número de dormitorios de la residencia</p>	<p>Rango: [1,10]</p>	<table border="1"> <tr> <td>1=5.92%</td> </tr> <tr> <td>2=24.74%</td> </tr> <tr> <td>3=44.25%</td> </tr> <tr> <td>4=18.47%</td> </tr> <tr> <td>5=4.53%</td> </tr> <tr> <td>6=1.74%</td> </tr> <tr> <td>7=1.14%</td> </tr> <tr> <td>8=0%</td> </tr> <tr> <td>9=0%</td> </tr> <tr> <td>10=0%</td> </tr> </table>	1=5.92%	2=24.74%	3=44.25%	4=18.47%	5=4.53%	6=1.74%	7=1.14%	8=0%	9=0%	10=0%													
1=5.92%																									
2=24.74%																									
3=44.25%																									
4=18.47%																									
5=4.53%																									
6=1.74%																									
7=1.14%																									
8=0%																									
9=0%																									
10=0%																									
<table border="1"> <tr> <td>Ingresos mensuales en salarios mínimos legales vigentes</td> </tr> <tr> <td><i>Menos de 1 SM</i></td> </tr> <tr> <td><i>Entre 1 y Menos de 2 SM</i></td> </tr> <tr> <td><i>Entre 2 y Menos de 3 SM</i></td> </tr> <tr> <td><i>Entre 3 y Menos de 5 SM</i></td> </tr> <tr> <td><i>Entre 5 y Menos de 7 SM</i></td> </tr> <tr> <td><i>Entre 7 y Menos de 10 SM</i></td> </tr> <tr> <td><i>10 o más SM</i></td> </tr> </table>	Ingresos mensuales en salarios mínimos legales vigentes	<i>Menos de 1 SM</i>	<i>Entre 1 y Menos de 2 SM</i>	<i>Entre 2 y Menos de 3 SM</i>	<i>Entre 3 y Menos de 5 SM</i>	<i>Entre 5 y Menos de 7 SM</i>	<i>Entre 7 y Menos de 10 SM</i>	<i>10 o más SM</i>	<table border="1"> <tr> <td>Valores Posibles: 7</td> </tr> <tr> <td>1</td> </tr> <tr> <td>2</td> </tr> <tr> <td>3</td> </tr> <tr> <td>4</td> </tr> <tr> <td>5</td> </tr> <tr> <td>6</td> </tr> <tr> <td>7</td> </tr> </table>	Valores Posibles: 7	1	2	3	4	5	6	7	<table border="1"> <tr> <td>1=20.56%</td> </tr> <tr> <td>2=52.61%</td> </tr> <tr> <td>3=18.47%</td> </tr> <tr> <td>4=5.23%</td> </tr> <tr> <td>5=2.79%</td> </tr> <tr> <td>6=0%</td> </tr> <tr> <td>7=0.35%</td> </tr> </table>	1=20.56%	2=52.61%	3=18.47%	4=5.23%	5=2.79%	6=0%	7=0.35%
Ingresos mensuales en salarios mínimos legales vigentes																									
<i>Menos de 1 SM</i>																									
<i>Entre 1 y Menos de 2 SM</i>																									
<i>Entre 2 y Menos de 3 SM</i>																									
<i>Entre 3 y Menos de 5 SM</i>																									
<i>Entre 5 y Menos de 7 SM</i>																									
<i>Entre 7 y Menos de 10 SM</i>																									
<i>10 o más SM</i>																									
Valores Posibles: 7																									
1																									
2																									
3																									
4																									
5																									
6																									
7																									
1=20.56%																									
2=52.61%																									
3=18.47%																									
4=5.23%																									
5=2.79%																									
6=0%																									
7=0.35%																									

Se observa que los estudiantes que trabajan en su mayoría lo hacen entre 16 y 20 horas a la semana, con base en el dato anterior se puede afirmar que los estudiantes de Ingeniería Electromecánica trabajan entre 3 y 4 horas diarias por ganar experiencia o ayudar a sus padres. Por otra parte se observa que los hogares cuentan con tres dormitorios por residencia en su gran mayoría contrastando con la información que en los hogares el promedio de personas en el hogar es de 4.

En cuanto a los ingresos familiares mensuales se observa que en gran parte de los hogares llegan entre 1 y menos de 2 salarios, acorde a un hogar de estrato 2, el cual es al que la mayoría de estudiantes de Ingeniería Electromecánica pertenecen.

FACTORES DE ESTUDIO ACADÉMICOS																	
DESCRIPCIÓN	RANGO	RESUMEN															
<table border="1"> <tr> <td>Tipo de bachillerato obtenido</td> </tr> <tr> <td><i>Académico</i></td> </tr> <tr> <td><i>Técnico</i></td> </tr> <tr> <td><i>Normalista</i></td> </tr> <tr> <td><i>Desconocido</i></td> </tr> </table>	Tipo de bachillerato obtenido	<i>Académico</i>	<i>Técnico</i>	<i>Normalista</i>	<i>Desconocido</i>	<table border="1"> <tr> <td>Valores Posibles: 4</td> </tr> <tr> <td><i>A</i></td> </tr> <tr> <td><i>T</i></td> </tr> <tr> <td><i>N</i></td> </tr> <tr> <td><i>D</i></td> </tr> </table>	Valores Posibles: 4	<i>A</i>	<i>T</i>	<i>N</i>	<i>D</i>	<table border="1"> <tr> <td><i>A=43.9%</i></td> </tr> <tr> <td><i>T=55.4%</i></td> </tr> <tr> <td><i>N=0.7%</i></td> </tr> <tr> <td><i>D=0%</i></td> </tr> </table>		<i>A=43.9%</i>	<i>T=55.4%</i>	<i>N=0.7%</i>	<i>D=0%</i>
Tipo de bachillerato obtenido																	
<i>Académico</i>																	
<i>Técnico</i>																	
<i>Normalista</i>																	
<i>Desconocido</i>																	
Valores Posibles: 4																	
<i>A</i>																	
<i>T</i>																	
<i>N</i>																	
<i>D</i>																	
<i>A=43.9%</i>																	
<i>T=55.4%</i>																	
<i>N=0.7%</i>																	
<i>D=0%</i>																	
Nivel educativo del padre; madre	Rango:[0,99]																
<i>Ninguno</i>	0	0=4.53%	0=1.39%														
<i>No tuvo escuela</i>	1	1=0%	1=0%														
<i>Preescolar</i>	2	2=0%	2=0%														
<i>Básica primaria</i>	3	3=0%	3=0%														
<i>Básica secundaria</i>	4	4=0%	4=0%														
<i>Media vocacional</i>	5	5=0%	5=0%														
<i>Tecnológico o técnico</i>	6	6=0%	6=0%														
<i>Universitario</i>	7	7=0%	7=0%														
<i>Postgrado</i>	8	8=0%	8=0%														
<i>Primaria incompleta</i>	9	9=17.07%	9=14.98%														
<i>Primaria completa</i>	10	10=21.95%	10=16.02%														
<i>Secundaria (bachillerato) incompleto</i>	11	11=15.67%	11=25.08%														
<i>Secundaria (bachillerato) completo</i>	12	12=19.5%	12=21.6%														
<i>Educación técnica o tecnológica incompleta</i>	13	13=2.09%	13=1.04%														
<i>Educación técnica o tecnológica completa</i>	14	14=7.31%	14=5.92%														
<i>Educación profesional incompleta</i>	15	15=2.09%	15=3.48%														
<i>Educación profesional completa</i>	16	16=6.27%	16=6.62%														
<i>Postgrado</i>	17	17=3.48%	17=3.83%														
<i>No sabe</i>	99	99=0%	99=0%														

La mayor parte de los estudiantes de Ingeniería electromecánica se graduó del colegio con un tipo de bachillerato técnico. En cuanto a la educación promedio alcanzada por el padre el mayor nivel se sitúa en primaria completa, seguido por bachillerato completo. Ahora al observar el máximo nivel de educación logrado por la madre que fue bachillerato incompleto, seguido por bachillerato completo al igual que el del padre, se puede determinar que:

- El nivel educativo de padres y madres es bajo, al considerar que apenas logran superar el nivel de bachillerato intermedio, es decir hasta la básica (novenno grado).
- Comparando el nivel de los padres, se alcanza a notar cierta diferencia a favor del entorno académico de la madre, es decir que alcanzan un nivel educativo más alto.

4.4 DESCRIPCIÓN BIVARIADA

A continuación se presentan las conclusiones al relacionar los puntajes obtenidos en los módulos del componente genérico y algunas características de los estudiantes. Los soportes de los resultados logrados se presentan en el Anexo A.

Para las siguientes conclusiones se aplicó la prueba de correlación de Pearson sobre los puntajes numéricos del componente genérico de las pruebas y se consideró rechazar la hipótesis nula teniendo en cuenta un p-valor menor o igual a 0.05, de este modo al relacionar los puntajes obtenidos y algunas características de los estudiantes se encontró correlación entre la variable edad y los puntajes en razonamiento cuantitativo, lectura crítica e inglés, además indicando que mientras mayor sea la edad menor será el puntaje en dichas categorías (correlación negativa).

En cuanto a los desempeños del componente genérico de la prueba se obtuvo los siguientes resultados a través del test ji-cuadrado de Pearson indicando que para el desempeño en comunicación escrita se encuentra asociado el sexo del estudiante y el número de horas que trabaja a la semana, de igual manera para el desempeño en inglés se encontró dependencia con el nivel educativo del padre y el salario mensual del grupo familiar en el hogar.

Se determinó realizar el test de normalidad de Shapiro Wilk (ver Anexo A) de cada variable respuesta numérica, obteniendo como resultado que ninguno de los puntajes del componente genérico de la prueba tienden a distribuirse normalmente. Por tal motivo no fue necesario hacer el test de varianzas de Leven's para observar varianzas iguales y determinar diferencia de medias entre las variables.

En este orden de ideas al asumir la no normalidad de los puntajes numéricos del componente genérico de las pruebas Saber Pro, se realizó la prueba no paramétrica de Kruskal Wallis con el fin de determinar asociación entre variables, obteniendo que:

El puntaje en razonamiento cuantitativo tiene relación con las variables género, horas de trabajo a la semana, tener lavadora, nivel educativo de los padres, ocupación de la madre y valor anual de la matrícula.

Tabla I.

Variables	Media
Mujeres	10.7
Hombres	11.1
Trabaja entre 1 y 5 horas	11.5
Trabaja entre 6 y 10 horas	10.6
Nivel educativo universitario de los padres	11.6
Nivel educativo sin estudio de los padres	10.9
Ocupación administración en la madre	10.95
Ocupación Pensionada de la madre	11.7

Así mismo para el puntaje en lectura crítica existe nexo con las variables tener trabajo, horas de trabajo a la semana, nivel educativo de la madre, ocupación del padre y valor anual de la matrícula.

Tabla II

Variables	Media
No Tiene trabajo	10.7
Trabaja como ayudante sin remuneracion	10.3
Ocupación padre: Administrador	10.5
Ocupación padre: Obrero	10.7
Matricula anual: No paga	10.8
Matricula anual: Entre 3 y 5 millones	9.9

Además se halló correlación entre los puntajes en inglés y las variables estrato, horas de trabajo a la semana, nivel de educación de los padres, ocupación de la madre, salario mensual del grupo familiar en el hogar y tiene trabajo.

Tabla III

Variable	Media
Estrato 2	10.14
Estrato 1	10.4
Trabaja 0 horas	10.3
Trabaja entre 16 y 20 horas	10.0
Madre con estudios universitarios	10.55
Madre sin estudios	9.86
Salario mensual familiar más de 10 SMLV	13.94
Salario mensual familiar menos de 1 SMLV	10.18

Finalmente, para el puntaje en comunicación escrita no se halló ningún tipo de asociación con otra variable.

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 (teniendo en cuenta el modelo más parsimonioso), y posteriormente la evaluación, teniendo en cuenta la bondad de ajuste, para finalmente interpretar los resultados.

5.1 ESPECIFICACIÓN DE LOS MODELOS

La Tabla 10 muestra la distribución asignada para cada variable respuesta y sus correspondientes funciones de enlace.

Tabla 10. Distribuciones, funciones de enlace

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	δ	ν	τ
Skew t type 3 (ST3)	Razonamiento cuantitativo Lectura crítica	identity	log	log	log
Logistic (LO)	Comunicación escrita	identity	log	---	---
Skew t type 2 (ST2)	Inglés	identity	log	identity	log

Para **RAZONAMIENTO CUANTITATIVO**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución ST3:

ST3	GT	ST2	TF	ST5	ST1
804.1670	804.5448	805.2190	806.1180	806.6338	808.0552

Figura 5-1: Ajuste datos Razonamiento cuantitativo

A continuación se muestra la prueba gráfica que respalda el ajuste de la distribución ST3 sobre el puntaje en **RAZONAMIENTO CUANTITATIVO**.

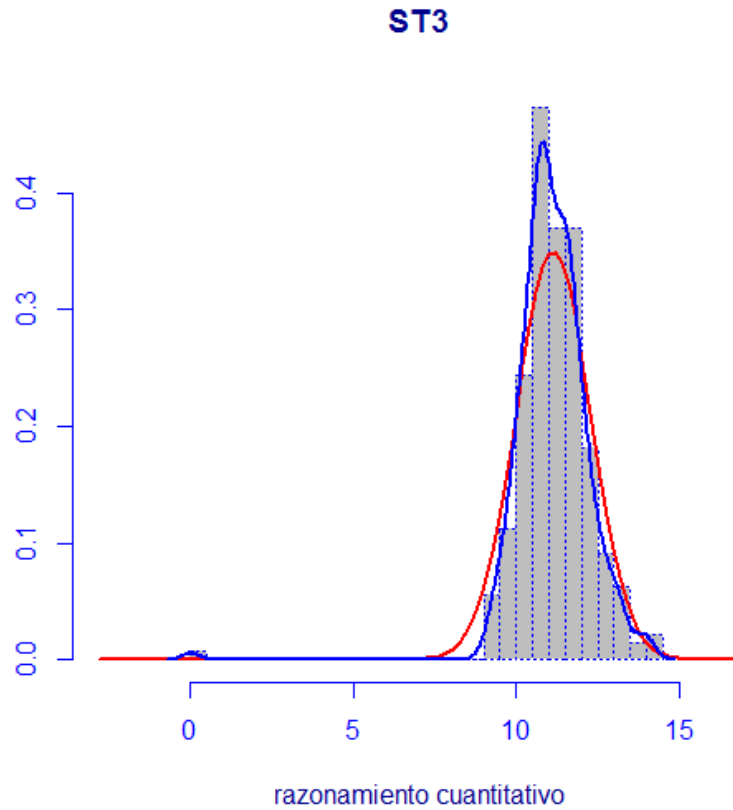


Figura 5-2: Prueba gráfica razonamiento cuantitativo

Así mismo para la variable respuesta puntaje en **LECTURA CRÍTICA**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución ST3:

ST3	TF	ST2	GT	ST5
737.1300	737.3544	737.4413	737.8011	738.4151

Figura 5-3: Ajuste datos Lectura Crítica

A continuación se muestra la prueba gráfica que respalda el ajuste de la distribución ST3 sobre el puntaje en **LECTURA CRÍTICA**.

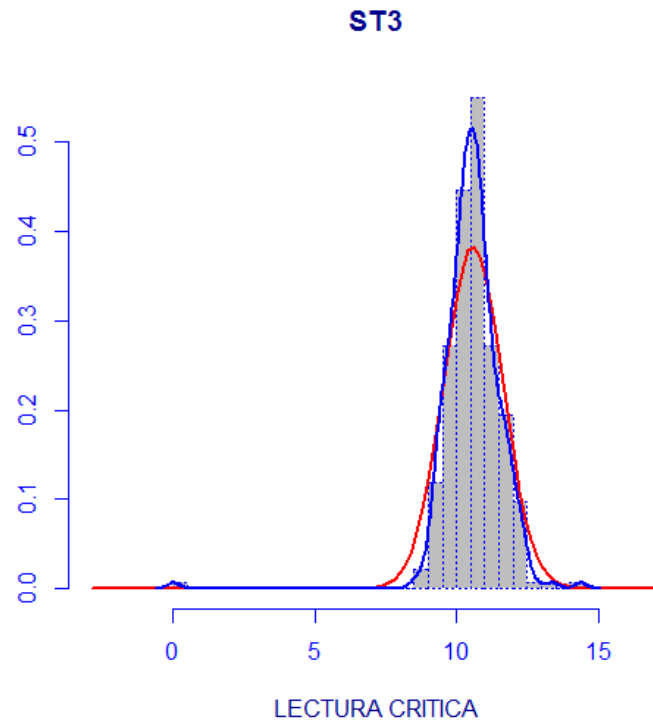


Figura 5-4: Prueba gráfica lectura crítica

En cuanto a la variable respuesta puntaje en **COMUNICACIÓN ESCRITA**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución Logistic:

LO	ST4	TF	ST3	SEP2
772.2682	773.6006	773.9917	775.2930	775.4489

Figura 5-5: Ajuste datos comunicación escrita

A continuación se muestra la prueba gráfica que respalda el ajuste de la distribución Logistic sobre el puntaje en **COMUNICACIÓN ESCRITA**.

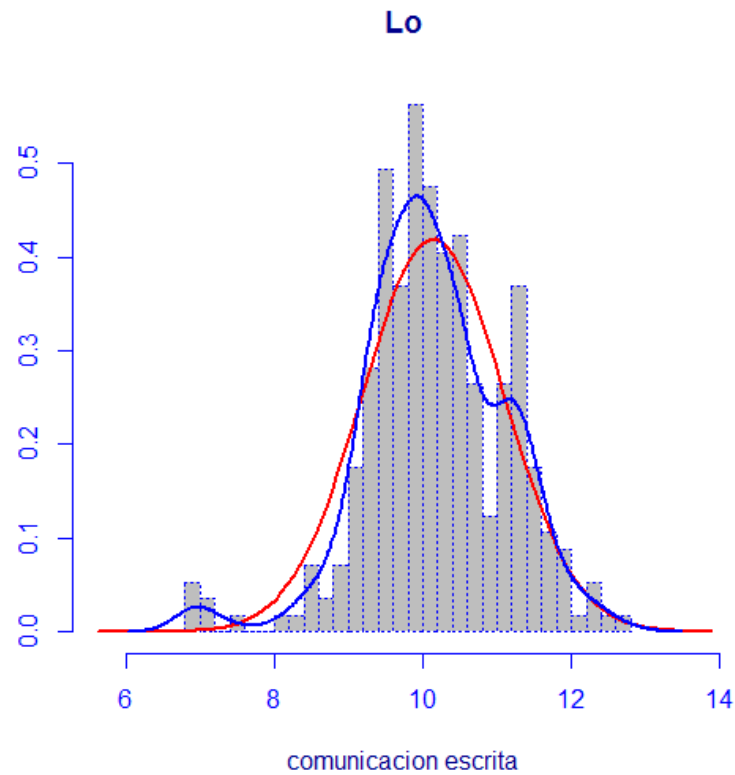


Figura 5-6: Prueba gráfica comunicación escrita

Para la variable respuesta **PUNTAJE EN INGLÉS**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución ST2:

ST2	ST3	JSU	ST1	ST5
798.2387	798.8719	801.8557	804.6990	806.0088

Figura 5-7: Ajuste datos Inglés

A continuación se muestra la prueba gráfica que respalda el ajuste de la distribución Skew t type 2 sobre el **PUNTAJE EN INGLÉS**.

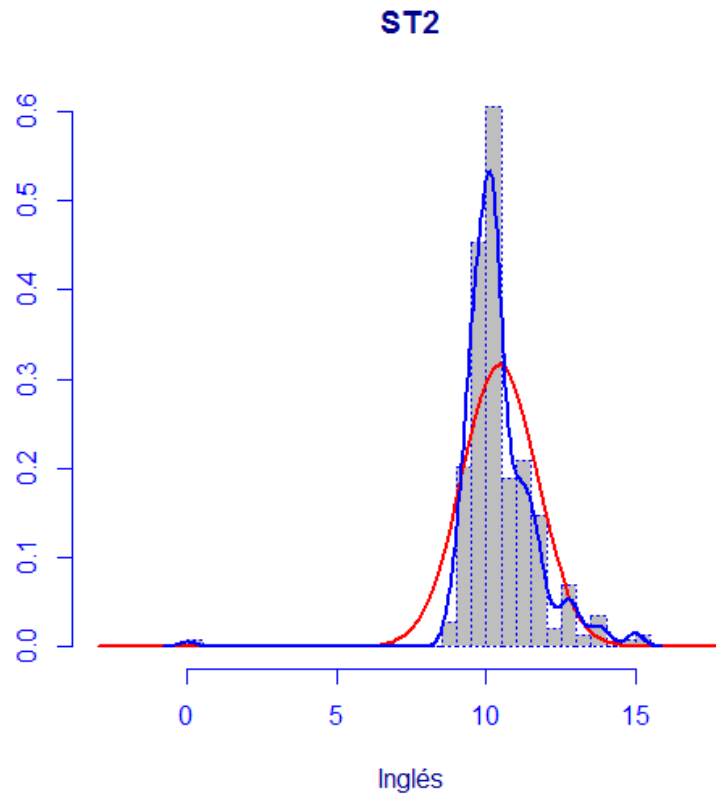


Figura 5-8: Prueba gráfica inglés

5.2 FACTORES DEMOGRÁFICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

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

Tabla 11. Componente Sistemático Factor Demográfico

Notación	Nombre
X_1	<i>Edad</i>
X_2	<i>Número de personas que conforman el hogar</i>
X_3	<i>Personas de las que usted se encuentra a cargo</i>
X_4	<i>Hogar actual</i>
X_5	<i>Sexo</i>
X_6	<i>Estado civil</i>

El modelo lineal queda determinado por el siguiente predictor lineal:

$$\eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_j X_4 + \alpha_j X_5 + \tau_k X_6$$

Dónde:

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_j = Efecto de la situación del hogar actual, con $j = 1, 2$

α_j = Efecto del género del estudiante, con $j = 1, 2$

τ_k = Efecto del estado civil, con $k = 1, 2, 3, 4, 5$

A continuación, se presenta el modelo óptimo para **RAZONAMIENTO CUANTITATIVO** teniendo en cuenta los factores demográficos; en la selección y mejor ajuste del modelo se utilizó el criterio StepGAICALLA.

```

-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.08499    0.43990  32.019 < 2e-16 ***
Edad          -0.15345    0.01755  -8.745 < 2e-16 ***
Situacion.hogar[T.Temporal]  0.28935    0.10314   2.805 0.00538 **
Genero[T.Masculino]  0.41674    0.18471   2.256 0.02483 *
---
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) -0.26620    0.04649  -5.726 0.0000000262 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function:  log
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.41914    0.29156  -1.438  0.1517
Edad         0.02931    0.01184   2.475  0.0139 *
---
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)  -6.11958    1.92857  -3.173  0.00168 **
Edad         1.19795    0.07534  15.901 < 2e-16 ***
personas.a.cargo  0.52327    0.05331   9.815 < 2e-16 ***
Personas.en.el.hogar -0.15595    0.06829  -2.284 0.02314 *
Estado.civil[T.Separado] -26.09127  1.25160 -20.846 < 2e-16 ***
Estado.civil[T.Soltero] -16.90096  0.53840 -31.391 < 2e-16 ***
Estado.civil[T.Union libre] -16.46319  0.58687 -28.053 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  14
      Residual Deg. of Freedom:  270
                          at cycle:  20

Global Deviance:    702.7242
      AIC:          730.7242
      SBC:          781.8098
*****

```

Figura 5-9: Modelo óptimo razonamiento cuantitativo factor demográfico.

Interpretación:

A partir de la información del modelo óptimo se puede afirmar que, con un nivel de significancia del 1%, la variable edad influye en dicho puntaje, es decir, si comparamos dos estudiantes, por cada año de más que tenga uno respecto al otro, su puntaje disminuirá 0.153 puntos.

Así mismo en este puntaje, un estudiante que sea de sexo masculino obtendrá 0.416 puntos más en dicha prueba frente a una estudiante de sexo femenino.

Por otra parte, se puede afirmar, con un nivel de significancia inferior al 1%, que al comparar dos estudiantes donde el segundo difiere del primero en la situación de su hogar, es decir si es permanente frente a uno temporal, obtendrá 0.289 puntos más.

Respecto a la variación en el puntaje de Razonamiento cuantitativo, se puede afirmar que las variables edad, número de personas a cargo, número de personas en el hogar y estado civil influyen en la variación del puntaje en este componente, con una significancia inferior al 5%.

Una vez ajustado el modelo GAMLSS, se utilizan los residuales cuantiles aleatorizados descritos en Dunn y Smyth (1996) y a los cuales hace referencia Pérez (Pérez, 2016) en su trabajo, y comprobar la adecuación del modelo y, más específicamente, la distribución seleccionada para la variable respuesta.

Las siguientes gráficas muestran el ajuste, a partir de los residuales, los cuales sea cual sea la distribución de la variable respuesta, siempre tienen una distribución normal estándar, siempre y cuando el modelo sea correcto.

```
*****  
Summary of the Quantile Residuals  
      mean      = 0.003762947  
      variance   = 1.001134  
      coef. of skewness = -0.05100746  
      coef. of kurtosis  = 2.719101  
Filliben correlation coefficient = 0.9984742  
*****
```

Figura 5-10: Prueba ajuste modelo demográfico razonamiento cuantitativo

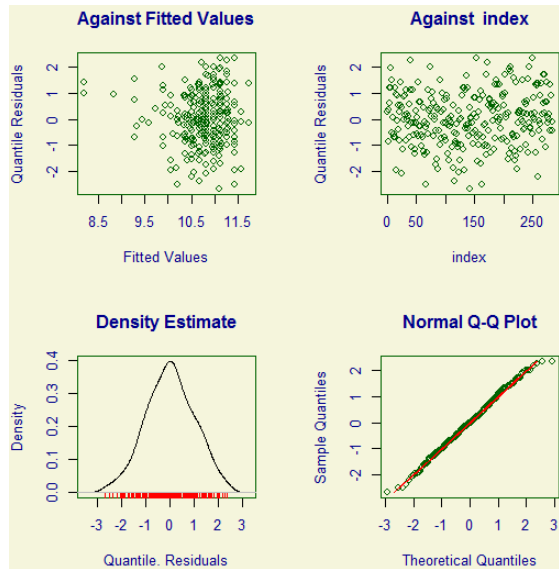


Figura 5-11: Ajuste de la distribución modelo puntaje razonamiento cuantitativo factor demográfico

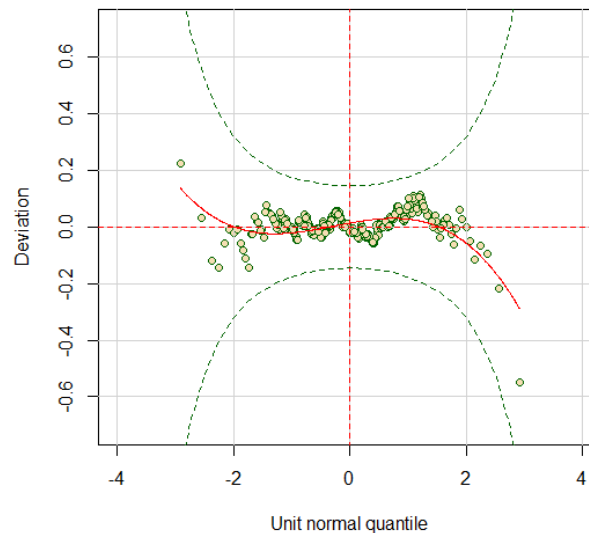


Figura 5-12: Ajuste de la distribución modelo razonamiento cuantitativo factor demográfico

En las gráficas anteriores se puede observar que los residuales aleatorizados para el modelo óptimo presentan un buen ajuste a una distribución normal estándar, esto quiere decir que la distribución Skew t type 3 es la adecuada para representar el comportamiento de la variable puntaje en razonamiento cuantitativo.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **LECTURA CRÍTICA** se tuvo en cuenta el criterio StepGAICALL.A.

```

Mu link function: identity
Mu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.13924    0.34361  35.329  < 2e-16 ***
Edad        -0.07250    0.01386  -5.232  0.000000328 ***
---
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)   -0.44687    0.05076  -8.803  < 2e-16 ***
Situacion.hogar[T.Temporal] 0.34373    0.10134   3.392  0.000793 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: log
Nu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.46347    0.11308   4.099  0.0000544 ***
Personas.en.el.hogar -0.05549    0.02394  -2.317   0.0212 *
---
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)   9.48609441  0.00010568  89764  <2e-16 ***
Estado.civil[T.Separado] -9.58000750  0.00008599 -111409  <2e-16 ***
Estado.civil[T.Soltero] -11.75799847  0.00010497 -112018  <2e-16 ***
Estado.civil[T.Union libre] -26.79137360  0.02357104  -1137  <2e-16 ***
personas.a.cargo    10.59293090  0.00002843  372549  <2e-16 ***
Personas.en.el.hogar  1.08146514  0.00002576  41978  <2e-16 ***
Situacion.hogar[T.Temporal]  3.43433188  0.00002832  121277  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 284
Degrees of Freedom for the fit: 13
      Residual Deg. of Freedom: 271
      at cycle: 20

Global Deviance: 672.2958
      AIC: 698.2958
      SBC: 745.7325
*****

```

Figura 5-13: Modelo óptimo lectura crítica factor demográfico

Interpretación:

Luego de observar el modelo óptimo de **LECTURA CRÍTICA** se deduce que, si comparamos dos estudiantes que difieren en la edad, se tiene que por cada año de más un estudiante obtendrá 0.44 puntos menos con relación a otro con una significancia inferior al 1%.

Además se puede indicar que la variabilidad del puntaje se ve afectado por las variables situación del hogar, número de personas en el hogar y estado civil, establecido con un criterio de significancia menor al 1% para cada estimación.

Con base a lo anterior se puede decir que el puntaje en lectura crítica a través de las variables demográficas se ve afectado únicamente por la edad de los estudiantes al momento de presentar la prueba Saber Pro en la carrera de Ingeniería Electromecánica de la UPTC seccional Duitama.

Las siguientes gráficas muestran el ajuste de los residuos, los cuales sin importar la distribución de la variable respuesta, deben acoplarse a una normal estándar.

```
*****  
Summary of the Quantile Residuals  
      mean      = -0.002504872  
      variance   =  0.9818973  
      coef. of skewness =  0.06364699  
      coef. of kurtosis  =  3.061046  
Filliben correlation coefficient =  0.9985599  
*****
```

Figura 5-14: Prueba ajuste modelo demográfico lectura crítica

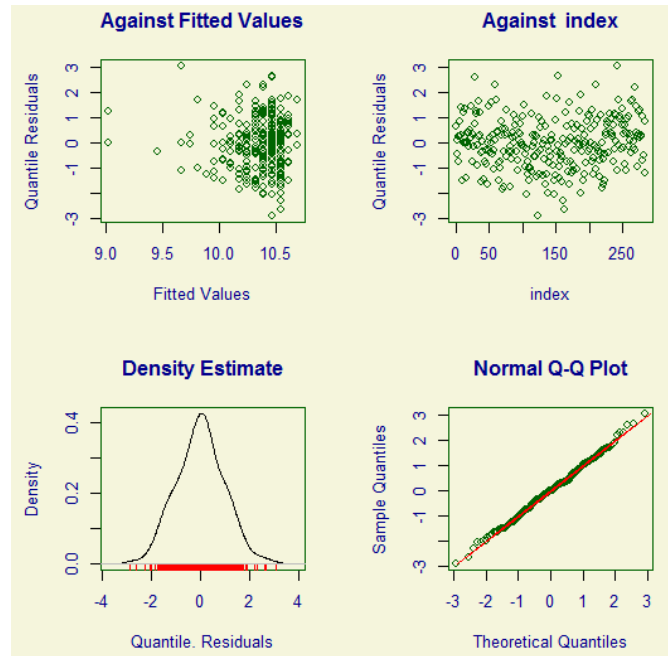


Figura 5-15: Ajuste de la distribución modelo puntaje lectura crítica factor demográfico

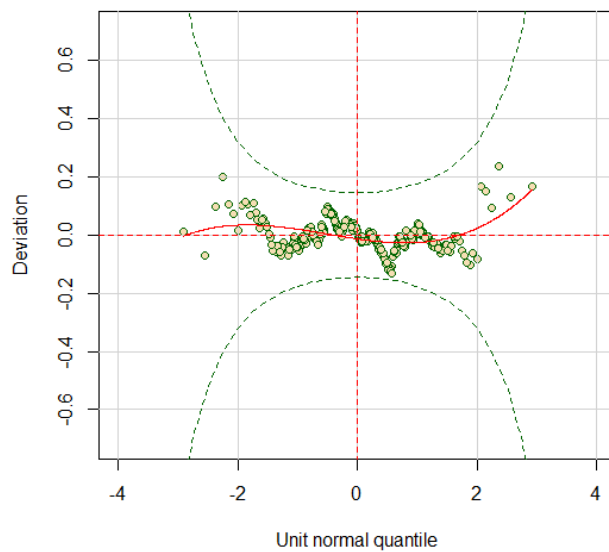


Figura 5-16: Ajuste de la distribución modelo lectura crítica factor demográfico

En las gráficas anteriores se observa que los residuales aleatorizados para el modelo óptimo en **LECTURA CRÍTICA** presentan un buen ajuste a una normal estándar, indicando que la distribución Skew t type 3 es la adecuada para explicar el comportamiento de dicha variable.

El modelo óptimo en cuanto a **COMUNICACIÓN ESCRITA** se trabajó con el criterio StepGAICALL.A para su selección y mejor ajuste.

```

Mu link function:  identity
Mu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.72186    1.10259  10.631  <2e-16 ***
Edad           -0.04308    0.01979  -2.177  0.0303 *
Estado.civil[T.Separado]  1.26772    1.23720   1.025  0.3064
Estado.civil[T.Soltero]  -0.32328    0.95806  -0.337  0.7360
Estado.civil[T.Union libre] -0.31904    0.99752  -0.320  0.7493
Genero[T.Masculino]      -0.23720    0.17448  -1.359  0.1751
---
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)  -0.66135    0.04942  -13.38  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  7
Residual Deg. of Freedom:  277
                             at cycle:  3

Global Deviance:    759.7095
AIC:                773.7095
SBC:                799.2523
*****

```

Figura 5-17: Modelo Óptimo comunicación escrita factor demográfico

Interpretación:

Si se compara la edad de dos estudiantes se tiene que, por cada año de más que tenga uno respecto al otro, su puntaje disminuirá 0.04 puntos en el **PUNTAJE EN COMUNICACIÓN ESCRITA**, con un nivel de significancia inferior al 5%.

Además se puede afirmar que dadas las variables explicativas, ninguna determina influencia en la variabilidad del puntaje en comunicación escrita.

De esta forma se concluye que de los factores demográficos analizados, tan sólo la edad explica el puntaje en comunicación escrita de los estudiantes de Ingeniería Electromecánica. Las siguientes gráficas mostrarán si el ajuste de los residuos, poseen una distribución normal estándar.

```

*****
Summary of the Quantile Residuals
      mean      = 0.002179308
      variance   = 1.002036
      coef. of skewness = -0.1473503
      coef. of kurtosis  = 3.063238
Filliben correlation coefficient = 0.9920197
*****

```

Figura 5-18: Prueba ajuste modelo demográfico comunicación escrita

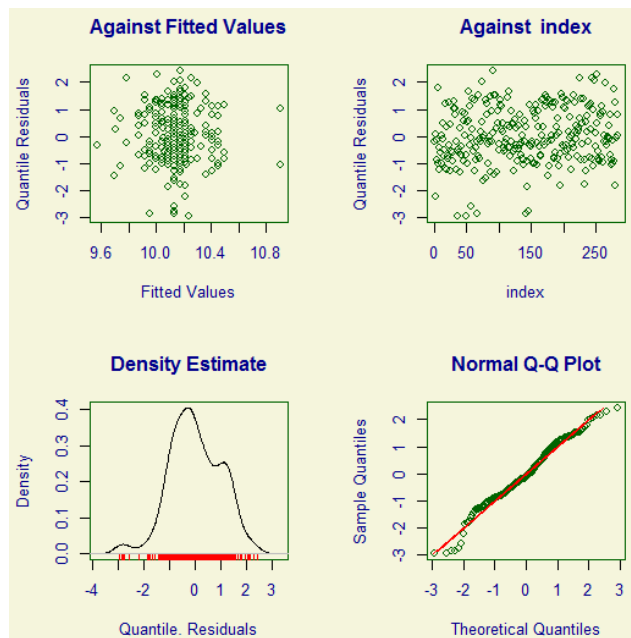


Figura 5-19: Ajuste de la distribución modelo puntaje comunicación escrita factor demográfico

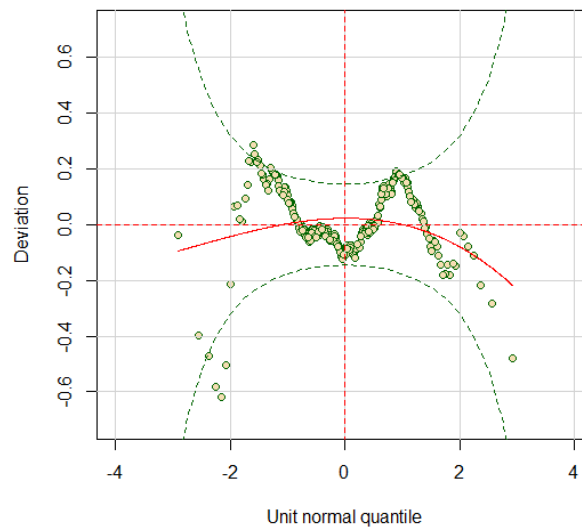


Figura 5-20: Ajuste de la distribución modelo comunicación escrita factor demográfico

En las gráficas anteriores se puede corroborar que los residuales aleatorizados para el modelo óptimo en comunicación escrita presentan un buen ajuste a una distribución normal estándar, esto quiere decir que la distribución Logistic es la adecuada para dicha variable.

Ahora bien, al **DESEMPEÑO EN COMUNICACIÓN ESCRITA** se aplica regresión logística ordinal. La Tabla muestra el modelo óptimo según factores demográficos.

Tabla 12. Resumen modelo óptimo: Comunicación escrita factor demográfico

	Coef. estimado	Error	Wald	p-valor	Odds Ratio	I.C. 95%	
Genero[T.Masculino]	-0.7349	0.35	-2.1	0.03	0.6758796	0.2415	0.9522
Nivel 1 Nivel 2	-4.7024	0.5570	-8.4428				
Nivel 2 Nivel 3	-4.2227	0.4853	-8.7019				
Nivel 3 Nivel 4	-2.7782	0.3770	-7.3699				
Nivel 4 Nivel 5	-0.4297	0.3338	-1.2875				
Nivel 5 Nivel 6	1.0100	0.3379	2.9888				
Nivel 6 Nivel 7	3.0535	0.4799	6.3631				

Interpretación del modelo:

En la Tabla 12 al analizar la variable sexo del estudiante se observa que su razón de probabilidad y su intervalo de confianza al 95% es menor que uno, lo que significa que la probabilidad acumulada, comenzando de Nivel de desempeño 1 hasta el nivel 8, decrece en todos los valores que puede tomar Comunicación escrita, así un estudiante de sexo masculino es menos probable que obtenga niveles altos en el desempeño en comunicación escrita, comparado con una estudiante de sexo femenino.

A continuación se realiza la selección del modelo óptimo para el **PUNTAJE EN INGLÉS** teniendo en cuenta el criterio StepGAICALL.A para la selección y mejor ajuste del modelo óptimo.

```

Mu link function:  identity
Mu Coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept)      11.81953    0.28251  41.838  <2e-16 ***
Edad             -0.03191    0.01023  -3.120  0.0020 **
Estado.civil[T.Separado] -1.06148    0.44450  -2.388  0.0176 *
Estado.civil[T.Soltero]  -1.60046    0.05420 -29.526  <2e-16 ***
Estado.civil[T.Union libre] -1.54818    0.13110 -11.809  <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)     -1.60539    0.79065  -2.030   0.0433 *
Edad            -0.04134    0.01937  -2.135   0.0337 *
Estado.civil[T.Separado]  3.98628    1.61006   2.476   0.0139 *
Estado.civil[T.Soltero]   2.76122    0.59399   4.649 0.00000517 ***
Estado.civil[T.Union libre] 2.84702    0.66379   4.289 0.00002479 ***
personas.a.cargo  -0.26068    0.10037  -2.597   0.0099 **
---
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)     2.9723     0.4169   7.129 8.44e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  13
    Residual Deg. of Freedom:  271
                                at cycle:  20

Global Deviance:  717.5968
AIC:              743.5968
SBC:              791.0334
*****

```

Figura 5-21: Modelo óptimo puntaje inglés factor demográfico

Interpretación:

Al observar el modelo óptimo para el puntaje en inglés, se tiene que, por cada año de más que tenga un estudiante frente a otro, el puntaje en inglés se reducirá 0.032 puntos. Así mismo al comparar dos estudiantes donde el segundo difiere del primero en el estado civil se tiene que, con un nivel de significancia inferior al 2%, que aquellos estudiantes solteros obtienen en promedio 1.60 puntos menos respecto a los estudiantes casados, similar situación a los estudiantes en unión libre que obtienen en promedio 1.54 puntos menos respecto a los estudiantes casados.

La variabilidad para el puntaje en inglés se ve afectada por las variables edad y número de personas a cargo, además de estado civil.

En conclusión, para el **PUNTAJE EN INGLÉS** expuesto a través de los factores demográficos, las variables edad y estado civil son influyentes, en el caso de los estudiantes de ingeniería electromecánica.

Las siguientes graficas muestran el ajuste de los residuos, que sin importar la distribución de la variable independiente, estos siempre deben asociarse a una distribución normal estándar.

```
*****
                Summary of the Quantile Residuals
                mean    =  0.001718739
                variance =  1.019979
                coef. of skewness =  0.04143454
                coef. of kurtosis =  2.694798
Filliben correlation coefficient =  0.9970208
*****
```

Figura 5-22: Prueba ajuste modelo demográfico puntaje inglés

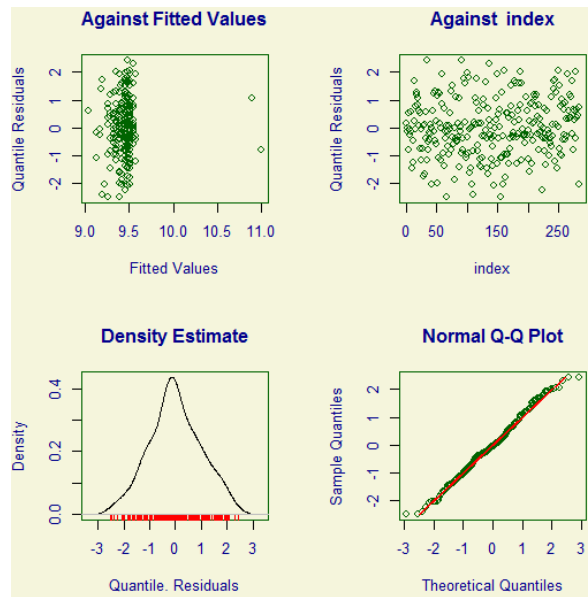


Figura 5-23: Ajuste de la distribución modelo puntaje inglés factor demográfico

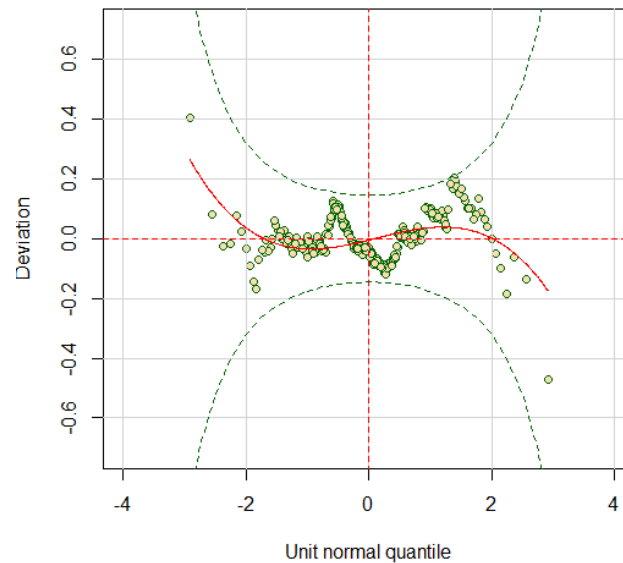


Figura 5-24: Ajuste de la distribución modelo inglés factor demográfico

En las gráficas anteriores se observa que los residuales aleatorizados para el modelo óptimo del **PUNTAJE EN INGLÉS** presentan un buen ajuste para esta variable, indicando que la distribución Skew t type 2 es la adecuada para explicar su comportamiento.

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

Tabla 13. Resumen modelo óptimo Inglés factor demográfico.

	Coef. estimado	Error	Wald	p-valor	Odds Ratio	I.C. 95%	
Edad	-0.1681	0.04	-4.133	0.000047	0.845	0.7805	0.9154
Estado.civil[T.Separado]	-12.630	1.34	-9.381	2.09e-18	3.2e-7	2.3e-7	4.5 e-5
Estado.civil[T.Soltero]	-14.231	0.40	-35.42	1.3e-106	6.6e-7	3.0e-7	1.4e-6
Estado.civil[T.Unionlibre]	-14.716	0.56	-26.1	1.20e-77	4.0e-7	1.3e-7	1.2e-6
A- A1	-20.5218	1.03	-19.733				
A1 A2	-18.7773	1.01	-18.51				
A2 B+	-17.4830	1.00	-17.455				
B+ B1	-17.0503	1.00	-17.041				

Interpretación del modelo

En la Tabla 13 al analizar la variable edad del estudiante se observa que su razón de probabilidad y su intervalo de confianza al 95% es menor que uno, lo que significa que la probabilidad acumulada, comenzando de Nivel A- hasta el nivel B1, decrece en todos los valores que puede tomar el desempeño en inglés, así un estudiante que tenga una edad inferior frente a otro, será menos probable que obtenga niveles altos en el desempeño en inglés. Así mismo un estudiante que tenga un estado civil de soltería, en unión libre o ser separado frente a uno casado, será menos probable que obtengan niveles altos en este mismo desempeño.

5.3 FACTORES ACADÉMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

La siguiente Tabla muestra las variables explicativas a tener en cuenta para los modelos académicos.

Tabla 14. Componente sistemático Factor académico

Notación	Nombre
X_1	<i>Tipo de bachillerato</i>
X_2	<i>Nivel de educación del padre</i>
X_3	<i>Nivel de educación de la madre</i>

El modelo lineal queda determinado por el siguiente predictor lineal:

$$\eta_i = \beta_0 + \gamma_j X_1 + \alpha_k X_2 + \delta_k X_3$$

β_0 = Intercepto

γ_j = Efecto del tipo de bachillerato, con j = 1; 2; 3

α_k = Efecto del nivel educativo del padre, con k = 1; 2; 3; 4; 5

δ_k = Efecto del nivel educativo de la madre, con k = 1; 2; 3; 4; 5

Para la selección y mejor ajuste del modelo óptimo en cuanto a **RAZONAMIENTO CUANTITATIVO** según factores académicos se tuvo en cuenta el criterio StepGAICALL.A, como se muestra a continuación:

```

Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.72063    0.11725  91.437 < 2e-16 ***
Nivel.edu.madre[T.Secundaria]    0.08032    0.10877   0.738  0.46087
Nivel.edu.madre[T.Sin estudio]  -0.06643    0.14927  -0.445  0.65667
Nivel.edu.madre[T.tecnicos]    -0.09477    0.23758  -0.399  0.69028
Nivel.edu.madre[T.Universitarios]  0.50475    0.18927   2.667  0.00812 **
---
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)   -0.19487    0.12667  -1.538    0.125
Nivel.edu.padre[T.Secundaria]   -0.02937    0.11315  -0.260    0.795
Nivel.edu.padre[T.Sin estudio]   0.02999    0.12201   0.246    0.806
Nivel.edu.padre[T.tecnicos]     0.03633    0.17481   0.208    0.836
Nivel.edu.padre[T.Universitarios] 0.26105    0.16132   1.618    0.107
Tipo.de.Bachillerato[T.Normalista] -2.42774    0.44510  -5.454 0.000000111 ***
Tipo.de.Bachillerato[T.Tecnico]   0.05003    0.08793   0.569    0.570
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function:  log
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.27749    0.06785   4.09 0.000057 ***
---
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)    5.923    25.639   0.231   0.817
-----
No. of observations in the fit: 284
Degrees of Freedom for the fit: 14
Residual Deg. of Freedom: 270
at cycle: 20

Global Deviance: 740.4325
AIC: 768.4325
SBC: 819.5182
*****

```

Figura 5-25: Modelo óptimo razonamiento cuantitativo factor académico

Interpretación:

Al observar el modelo óptimo de Razonamiento Cuantitativo teniendo en cuenta factores académicos se asume que, al comparar nivel educativo de la madre aquellos estudiantes con madres universitarias tendrán 0.50 puntos más en esta prueba respecto a los estudiantes con madres con nivel académico de primaria, dicho con un nivel de significancia del 1%.

La variabilidad en el puntaje de Razonamiento Cuantitativo, con una significancia del 1% se ve influenciado por la variable tipo de bachillerato.

Por consiguiente, se puede determinar que el puntaje en razonamiento cuantitativo, según factores de tipo académico, está determinado por la variable educación de la madre.

Una vez ajustado el modelo GAMLSS, se utilizaron los residuales cuantiles aleatorizados para comprobar si la distribución elegida es la más apropiada para la variable respuesta con dichas variables explicativas.

Las siguientes gráficas muestran el ajuste de los residuos, los cuales siempre tienden a una distribución normal estándar.

```
*****  
Summary of the Quantile Residuals  
      mean      = 0.00587538  
      variance   = 0.9978092  
      coef. of skewness = 0.008690941  
      coef. of kurtosis  = 3.008858  
Filliben correlation coefficient = 0.9985143  
*****
```

Figura 5-26: Prueba ajuste modelo académico razonamiento cuantitativo

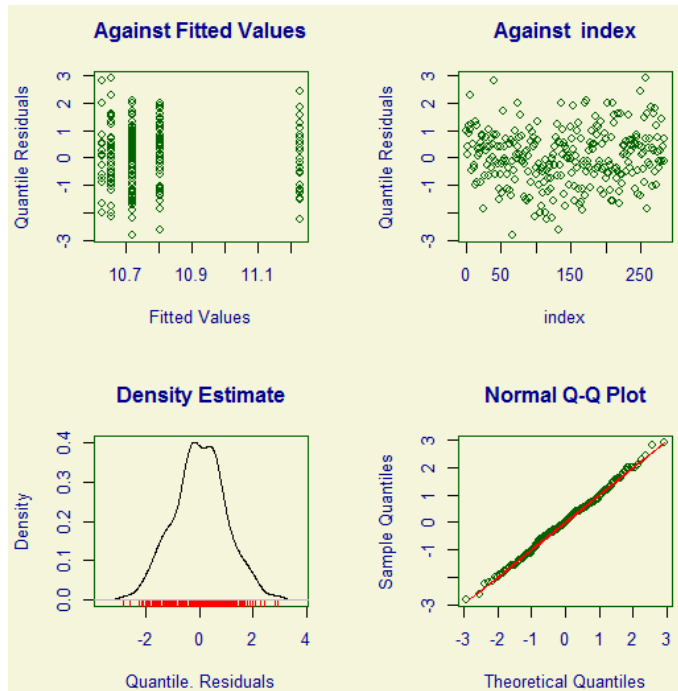


Figura 5-27: Ajuste de la distribución modelo puntaje razonamiento cuantitativo factor Académico

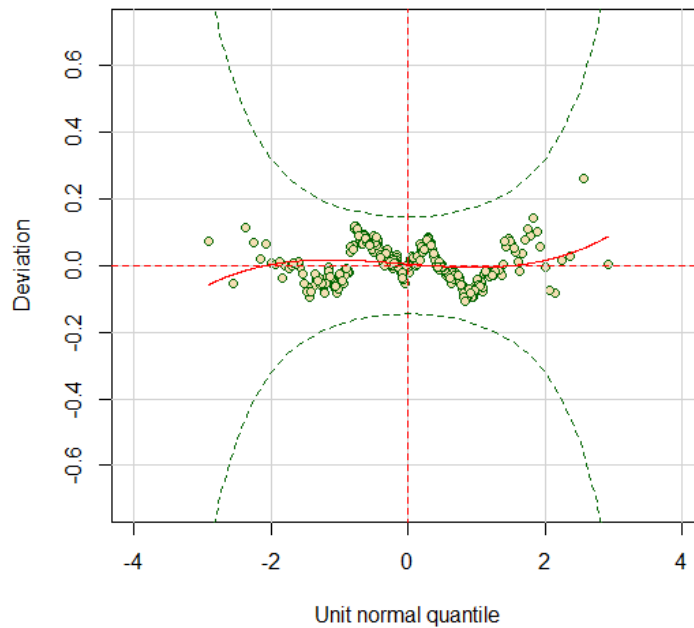


Figura 5-28: Ajuste de la distribución modelo razonamiento cuantitativo factor académico

En las figuras 5-27 y 5-28 se observa que los residuales aleatorizados para el modelo óptimo en **Razonamiento Cuantitativo** según factores académicos presentan un buen ajuste, esto quiere decir que la distribución Skew t type 3 es la más adecuada para explicar su comportamiento a través de las variables explicativas.

A continuación se presenta la selección y mejor ajuste del modelo óptimo en cuanto al **PUNTAJE EN LECTURA CRÍTICA** a través de factores académicos, se tuvo en cuenta el criterio StepGAICALL.A para dicha selección.

```

Mu link function:  identity
Mu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.32942    0.12238  84.404 < 2e-16 ***
Tipo.de.Bachillerato[T.Normalista]  2.23605    0.57829   3.867 0.000138 ***
Tipo.de.Bachillerato[T.Tecnico]     0.14187    0.09479   1.497 0.135626
---
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)  -0.29954    0.06703  -4.469 0.0000115 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function:  log
Nu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.08224    0.09838   0.836 0.40393
Nivel.edu.madre[T.Secundaria]    0.17315    0.09769   1.773 0.07742 .
Nivel.edu.madre[T.Sin estudio]   0.03428    0.10684   0.321 0.74854
Nivel.edu.madre[T.tecnicos]    -0.06218    0.15508  -0.401 0.68878
Nivel.edu.madre[T.Universitarios] 0.35599    0.13435   2.650 0.00853 **
---
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)    2.8373    0.8879   3.196 0.00156 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  10
Residual Deg. of Freedom:  274
at cycle:  20

Global Deviance:  683.9552
AIC:  703.9552
SBC:  740.445
*****

```

Figura 5-29: Modelo óptimo lectura crítica factor académico

Interpretación:

Según el modelo óptimo en Lectura Crítica se observa que, con un nivel de significancia del 1%, si se compara un estudiante que haya finalizado su bachillerato en un colegio normalista con otro que lo haya hecho en uno académico, su puntaje aumentará 2.23 puntos.

Por otra parte la variabilidad del puntaje en lectura crítica está asociada al nivel educativo de la madre.

Por lo anterior se deduce que el puntaje en **LECTURA CRÍTICA** en los estudiantes de Ingeniería Electromecánica, según factores académicos, está asociado al tipo de bachillerato que cursó al salir de grado once.

Las siguientes gráficas muestran el ajuste de los residuos.

```
*****  
Summary of the Quantile Residuals  
      mean    = -0.01290342  
      variance = 1.014789  
      coef. of skewness = 0.177667  
      coef. of kurtosis = 2.923983  
Filliben correlation coefficient = 0.9973456  
*****
```

Figura 5-30: Prueba ajuste modelo académico lectura crítica

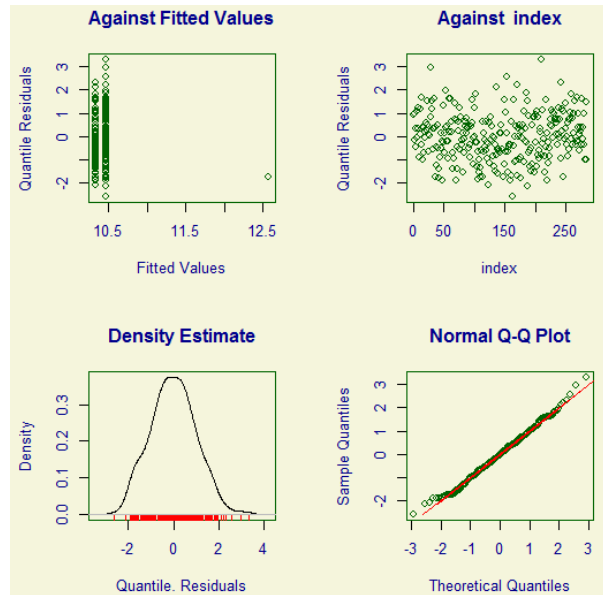


Figura 5-31: Ajuste de la distribución modelo puntaje lectura crítica factor académico

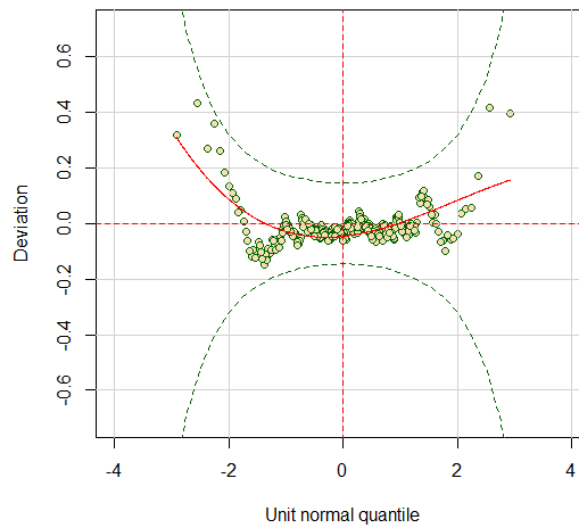


Figura 5-32: Ajuste de la distribución modelo lectura crítica factor académico

En las figuras 5-31 y 5-32 se puede determinar que los residuos aleatorizados para el modelo óptimo del puntaje en lectura crítica presenta un buen ajuste a una normal estándar, esto quiere decir que la distribución Skew t Type 3 es la adecuada teniendo en cuenta dicho factor en su comportamiento.

Ahora, para la selección y mejor ajuste del modelo óptimo en cuanto a **COMUNICACIÓN ESCRITA** según factores académicos se tuvo en cuenta el criterio StepGAICALL.A, como se describe a continuación:

```

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

Call:  gamlss(formula = puntaje.comuni.escri ~ Nivel.edu.padre,
            sigma.formula = ~1, family = LO, data = Dataset)

Fitting method: RS()

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

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.02981    0.08531  117.575  <2e-16 ***
Nivel.edu.padre[T.Secundaria]  0.15329    0.14057   1.091  0.2764
Nivel.edu.padre[T.Sin estudio]  0.02256    0.14115   0.160  0.8731
Nivel.edu.padre[T.tecnicos]    0.51445    0.21531   2.389  0.0175 *
Nivel.edu.padre[T.Universitarios] 0.38619    0.19727   1.958  0.0513 .
---
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) -0.66154    0.04942  -13.38  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  6
Residual Deg. of Freedom:  278
                        at cycle:  3

Global Deviance:  759.3439
AIC:  771.3439
SBC:  793.2377
*****

```

Figura 5-33: Modelo óptimo comunicación escrita factor académico

Interpretación:

Según el modelo óptimo de comunicación escrita, con un nivel de significancia del 1%, al comparar el nivel educativo del padre se tiene que, aquellos estudiantes con papás con nivel educativo Técnico tendrán 0.51 puntos más en esta prueba, respecto a los estudiantes con progenitores con nivel educativo de primaria.

Se puede afirmar que dadas las variables respuesta, éstas no explican la variabilidad del puntaje en comunicación escrita, según factores académicos.

De esta manera se puede interpretar que, de los factores académicos que inciden en el puntaje en comunicación escrita de los estudiantes de Ingeniería Electromecánica en las pruebas saber pro sólo incide el nivel educativo del padre.

Las siguientes gráficas muestran el ajuste de los residuos.

```

*****
Summary of the Quantile Residuals
      mean      = -0.0009187135
      variance   = 1.000915
      coef. of skewness = -0.1474808
      coef. of kurtosis  = 3.003434
Filliben correlation coefficient = 0.9943392
*****

```

Figura 5-34: Prueba ajuste modelo académico comunicación escrita

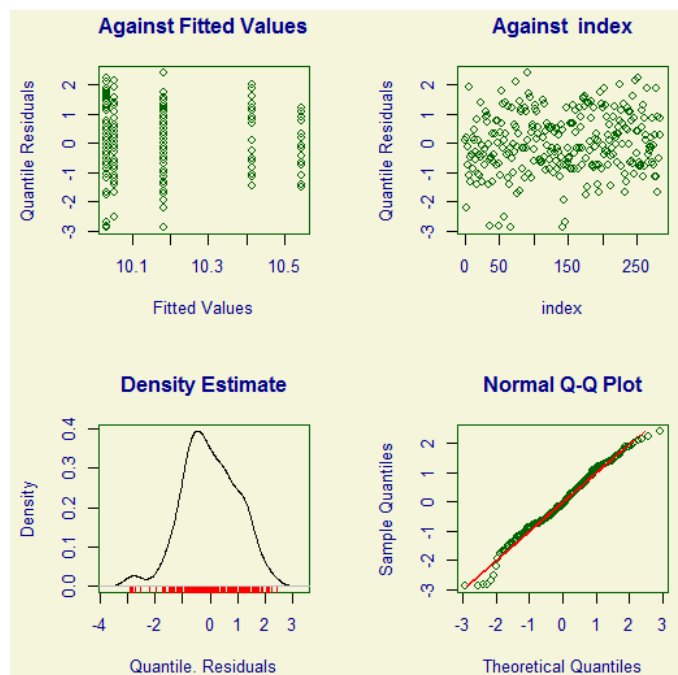


Figura 5-35: Ajuste de la distribución modelo puntaje comunicación escrita factor académico

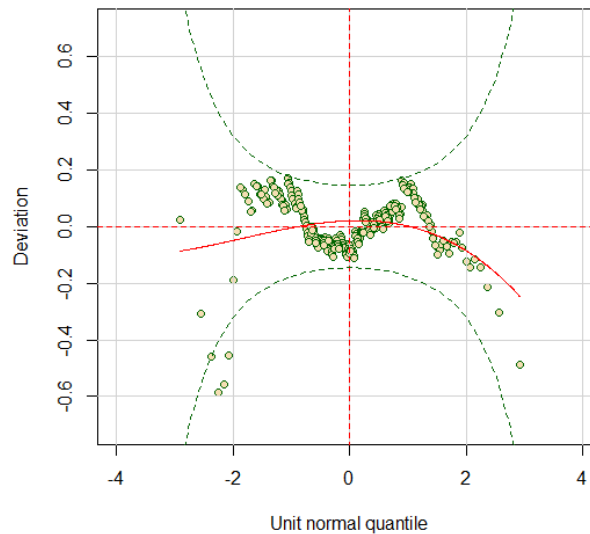


Figura 5-36: Ajuste de la distribución modelo comunicación escrita factor académico

En las gráficas anteriores se determina que los residuales aleatorizados para el modelo óptimo en **comunicación escrita** según factores académicos presenta un buen ajuste, es decir, que la distribución Logística es la adecuada para explicar su comportamiento.

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

Tabla 15 resumen modelo Óptimo Comunicación escrita factor académico

	Coef. estimado	Error	Wald	p-valor	Odds Ratio	I.C. 95%
Nivel 1 Nivel 2	-4.0218	0.4512	-8.9135			
Nivel 2 Nivel 3	-3.5410	0.3586	-9.8733			
Nivel 3 Nivel 4	-2.0994	0.1903	-11.032			
Nivel 4 Nivel 5	0.2263	0.1194	1.8948			
Nivel 5 Nivel 6	1.6436	0.1611	10.2050			
Nivel 6 Nivel 7	3.6781	0.3827	9.6107			

Interpretación del modelo

Al analizar las variables académicas asociadas al desempeño del estudiante en comunicación escrita se observa que, ninguna de estas explica dicho nivel de desempeño.

A continuación se presenta el modelo óptimo cuanto al puntaje en inglés, además para la selección y mejor ajuste se tuvo en cuenta el criterio StepGAICALL.A.

```

Mu link function:  identity
Mu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.435218   0.056926 165.747 < 2e-16 ***
Nivel.edu.padre[T.Secundaria]  0.464759   0.056926   8.164 1.12e-14 ***
Nivel.edu.padre[T.Sin estudio] -0.006061   0.101471  -0.060 0.95241
Nivel.edu.padre[T.tecnicos]    0.550938   0.143310   3.844 0.00015 ***
Nivel.edu.padre[T.Universitarios] 1.104811   0.056926  19.408 < 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)    -0.22089   0.08646  -2.555 0.0111 *
Tipo.de.Bachillerato[T.Normalista] -12.61048   1.10264 -11.437 <2e-16 ***
Tipo.de.Bachillerato[T.Tecnico]     0.02695   0.11615   0.232 0.8167
---
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)    2.325376   0.597582   3.891 0.000125 ***
Nivel.edu.padre[T.Secundaria]  -1.269408   0.691642  -1.835 0.067518 .
Nivel.edu.padre[T.Sin estudio]  -0.440882   0.800958  -0.550 0.582456
Nivel.edu.padre[T.tecnicos]    -0.001391   1.301951  -0.001 0.999148
Nivel.edu.padre[T.Universitarios] -2.089215   0.675876  -3.091 0.002196 **
---
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)    0.9538     0.1692     5.636 0.0000000421 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  14
Residual Deg. of Freedom:  270
at cycle:  20

Global Deviance:  705.3406
AIC:  733.3406
SBC:  784.4262
*****

```

Figura 5-37: Modelo óptimo inglés factor académico

Interpretación:

Al observar el modelo óptimo del **PUNTAJE EN INGLÉS** se tiene que, al comparar el nivel educativo del padre, aquellos estudiantes con padres universitarios tendrán 1.10 puntos más en esta prueba, respecto a los estudiantes con padres con nivel académico de primaria, al igual que estudiantes con padres técnicos que obtendrán 0.55 puntos más respecto a estudiantes con padres de nivel educativo de primaria, con un nivel de significancia del 1%.

La variabilidad en el puntaje de Inglés se ve influenciada por las variables tipo de bachillerato y nivel educativo del padre.

Por lo anterior se deduce que el **PUNTAJE EN INGLÉS** de los estudiantes de Ingeniería electromecánica, según factores académicos, está asociado al nivel educativo de sus progenitores.

Las siguientes gráficas muestran el ajuste de estos residuos, que deben ser similares a una distribución normal estándar.

```
*****
      Summary of the Quantile Residuals
      mean      = 0.01780301
      variance  = 0.9925207
      coef. of skewness = 0.1568344
      coef. of kurtosis  = 2.388413
      Filliben correlation coefficient = 0.9946103
*****
```

Figura 5-38: Prueba ajuste modelo académico ingles

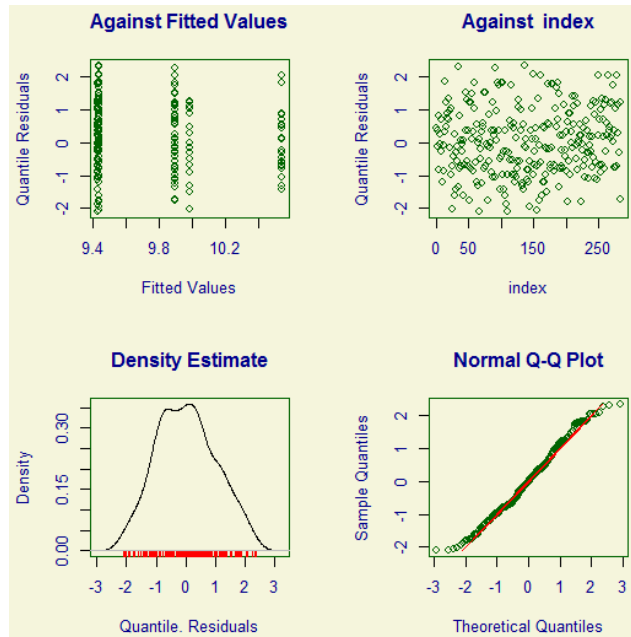


Figura 5-39: Ajuste de la distribución modelo puntaje inglés factor académico

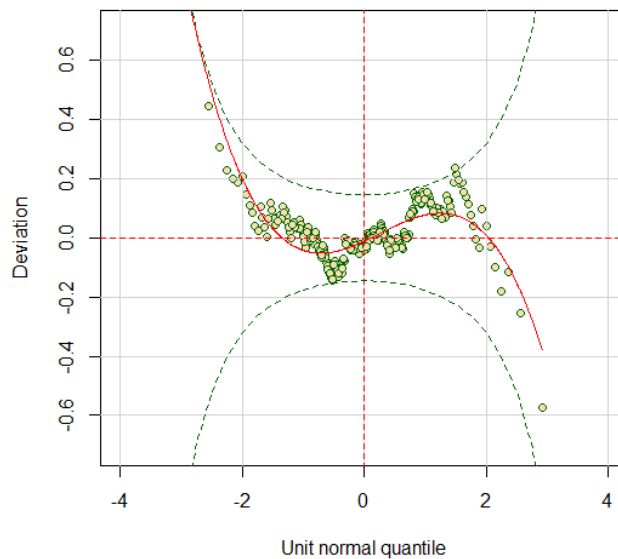


Figura 5-40: Ajuste de la distribución modelo inglés factor académico

En las gráficas anteriores se corrobora que los residuales aleatorizados para el modelo óptimo en el Puntaje en Inglés está asociado a una normal estándar (0,1). Por lo tanto la distribución Skew t type 2 es la adecuada para explicar su comportamiento.

En cuanto al **DESEMPEÑO EN INGLÉS** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo resultante a través del factor académico.

Tabla 16. Resumen modelo óptimo desempeño inglés factor académico

	Coef. estimado	Error	Wald	p-valor	Odds Ratio	I.C. 95%	
Nivel.edu.padre[T.Secundaria]	0.1215	0.2761	0.4401	0.660	1.12	0.65	1.93
Nivel.edu.padre[T.Sin estudio]	-0.1987	0.2908	-0.6833	0.494	0.81	0.46	1.44
Nivel.edu.padre[T.tecnicos]	1.1182	0.4312	2.5931	0.010	3.05	1.31	7.12
Nivel.edu.padre[T.Universitarios]	1.0544	0.3834	2.7503	0.006	2.87	1.35	6.08
A- A1	-1.9642	0.2313	-8.4905				
A1 A2	-0.2576	0.1824	-1.4126				
A2 B+	1.0087	0.1914	5.2693				
B+ B1	1.4293	0.2018	7.0834				

Interpretación del modelo

En la Tabla 16 se observa que las categorías de la variable nivel educativo del padre con estudios técnicos y universitarios en sus razones de probabilidad y sus intervalo de confianza al 95% son mayores que uno, lo que significa que la probabilidad acumulada, comenzando de Nivel de desempeño A- hasta el nivel B1, aumentan en todos los valores que puede tomar inglés, así un estudiante con un padre que tenga estudios universitarios o técnicos es más probable que obtenga niveles altos en el desempeño en inglés de la prueba Saber Pro.

Por otra parte los niveles educativos del padre en secundaria y sin estudios contienen a uno en su intervalo de confianza, es decir que no influyen en este desempeño para estudiantes de Ingeniería Electromecánica.

5.4 FACTORES SOCIOECÓMICOS QUE INFLUYEN EN EL COMPONENTE GENÉRICO

La siguiente Tabla muestra las variables explicativas a tener en cuenta para los modelos socioeconómicos.

Tabla 17. Componente sistemático socioeconómico

Notación	Nombre
X_1	<i>Número de dormitorios</i>
X_2	<i>Estrato de residencia según factura de energía</i>
X_3	<i>Valor anual de la matrícula</i>
X_4	<i>Ocupación del padre</i>
X_5	<i>Ocupación de la madre</i>
X_6	<i>Tiene celular</i>
X_7	<i>El hogar cuenta con internet</i>
X_8	<i>El hogar cuenta con televisión</i>
X_9	<i>El hogar cuenta con lavadora</i>
X_{10}	<i>El hogar cuenta con teléfono fijo</i>
X_{11}	<i>Trabaja actualmente</i>
X_{12}	<i>Horas que trabaja semanalmente</i>
X_{13}	<i>Ingresos mensuales familiares</i>
X_{14}	<i>El hogar cuenta con computador</i>

$$\eta_i = \beta_0 + \beta_1 X_1 + \varepsilon_a X_2 + \theta_b X_3 + \gamma_c X_4 + \alpha_d X_5 + \tau_f X_6 + \vartheta_h X_7 + \pi_i X_8 + \rho_j X_9 + \sigma_k X_{10} + \varphi_m X_{11} + \omega_n X_{12} + \epsilon_p X_{13} + \partial_r X_{14}$$

Dónde:

β_0 = Intercepto

β_1 = Efecto del número de dormitorios.

ε_a = Efecto del estrato del hogar, con a=1, 2, 3, 4

θ_b = Efecto del valor anual de la matrícula, con b=1, 2, 3, 4

γ_c = Efecto de la ocupación del padre, con c = 1, 2, 3, 4, 5

α_d = Efecto de la ocupación de la madre, con d = 1, 2, 3, 4, 5

τ_f = Efecto de tener celular, con f = 1, 2

ϑ_h = Efecto de contar con internet, h=1, 2

π_i = Efecto de contar con televisión, i=1, 2

ρ_j = Efecto de contar con lavadora, j=1, 2

σ_k = Efecto de contar con teléfono fijo, k=1, 2

φ_m = Efecto de tener trabajo, con m = 1, 2, 3, 4

ω_n = Efecto del número de horas que labora en la semana, con n= 1, 2, 3, 4

ϵ_p = Efecto de los ingresos mensuales familiares, con p = 1, 2, 3, 4, 5

∂_r = Efecto de contar con computador, r=1, 2

A continuación se presenta el modelo óptimo en cuanto a razonamiento cuantitativo según factores socioeconómicos. Para la selección y mejor ajuste del se tuvo en cuenta el criterio StepGAICALL.A.

```

Mu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.629285  0.193980  54.796 < 2e-16 ***
Horas.trabajo.semanal[T.Entre 1 y 5 horas]  0.287307  0.313549  0.916  0.36031
Horas.trabajo.semanal[T.Entre 11 y 15 horas] -0.168250  0.193987 -0.867  0.38652
Horas.trabajo.semanal[T.Entre 16 y 20 horas] -0.000378  0.123741 -0.003  0.99756
Horas.trabajo.semanal[T.Entre 6 y 10 horas] -0.512881  0.171902 -2.984  0.00310 **
ocupa.madre[T.Obrero] 0.682032  0.231890  2.941  0.00355 **
ocupa.madre[T.Ocupacion No remunerada] 0.240283  0.199044  1.207  0.22840
ocupa.madre[T.Otra actividad u ocupacion] 0.193196  0.222959  0.867  0.38697
ocupa.madre[T.Pensionado] 0.424407  0.352831  1.203  0.23006
---
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)    -0.12552  0.05732  -2.190  0.0294 *
Trabaja[T.si, como ayudante sin remuneracion] -0.09465  0.09242  -1.024  0.3067
Trabaja[T.si, por experiencia o remuneracion]  0.14427  0.13897  1.038  0.3001
Trabaja[T.si, por ser practica en plan de estudios] 21.82248  0.05855 372.706 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: log
Nu Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.23541  0.03483  6.759 7.94e-11 ***
---
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)    11.20  26.29  0.426  0.67

-----
No. of observations in the fit: 284
Degrees of Freedom for the fit: 15
Residual Deg. of Freedom: 269
at cycle: 20

Global Deviance: 779.2223
AIC: 809.2223
SBC: 863.9569
*****

```

Figura 5-41: Modelo óptimo razonamiento cuantitativo factor socioeconómico

Interpretación

Al analizar el modelo óptimo de Razonamiento Cuantitativo según factores socioeconómicos, se observa que, al comparar la ocupación de la madre aquellos estudiantes con madres obreras tendrán 0.68 puntos más en esta prueba, respecto a los estudiantes con madres administradoras; con un nivel de significancia del 1%. Así mismo, un estudiante que trabaje entre 6 y 10 horas obtendrá 0.51 puntos menos que un estudiante que no labore.

La variabilidad en el puntaje de Razonamiento Cuantitativo se ve afectado por la variable horas de trabajo del estudiante.

Por consiguiente, se puede determinar que el **PUNTAJE EN RAZONAMIENTO CUANTITATIVO**, según factores de tipo socioeconómico, está determinado por la variable educación del padre y número de horas que trabaje, para los estudiantes de Ingeniería Electromecánica.

Se muestra las gráficas con el ajuste de los residuos:

```
*****
                          Summary of the Quantile Residuals
                          mean      = -0.008476217
                          variance   =  1.001238
                          coef. of skewness =  0.07716451
                          coef. of kurtosis =  2.945748
Filliben correlation coefficient =  0.9981655
*****
```

Figura 5-42: Prueba ajuste modelo socioeconómico razonamiento cuantitativo

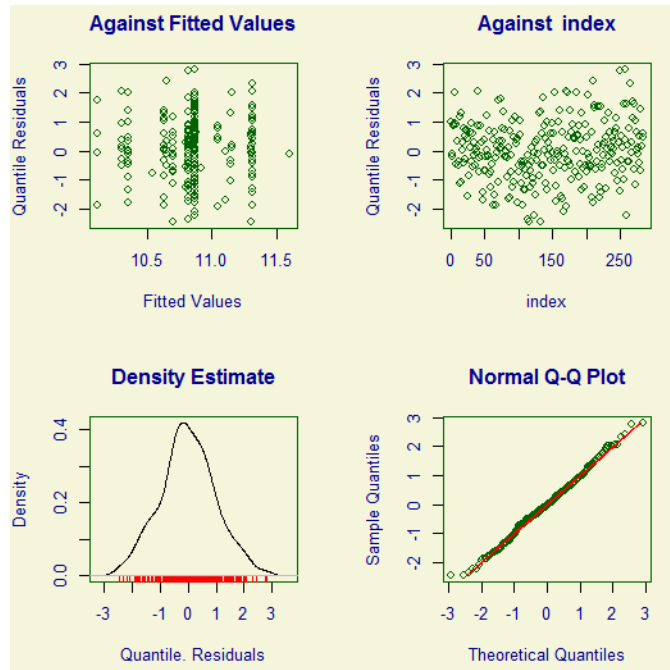


Figura 5-43: Ajuste de la distribución modelo puntaje razonamiento cuantitativo factor socioeconómico

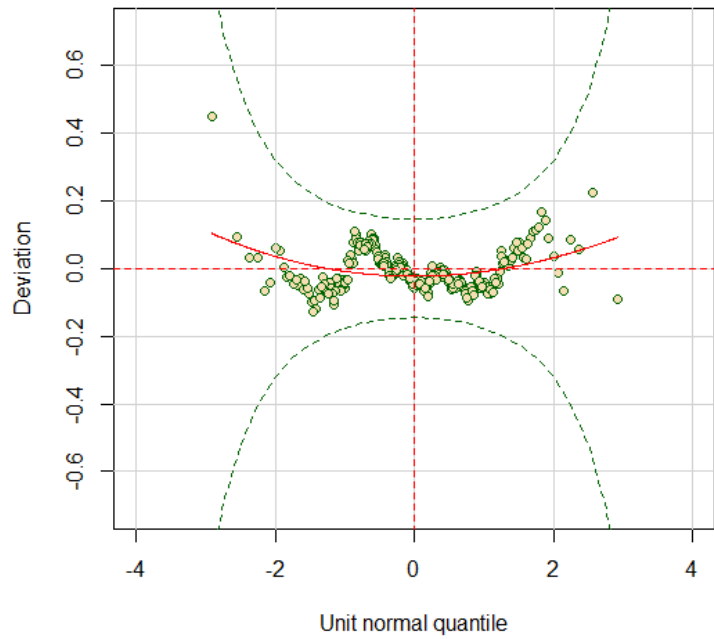


Figura 5-44: Ajuste de la distribución modelo razonamiento cuantitativo factor socioeconómico

En las gráficas anteriores se puede ver que los residuales aleatorizados para el modelo óptimo en el **PUNTAJE DE RAZONAMIENTO CUANTITATIVO** según factores socioeconómicos presenta un buen ajuste a una normal estándar, esto quiere decir que la distribución ST3 es la adecuada para observar su comportamiento.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **LECTURA CRÍTICA** según factores socioeconómicos se tuvo en cuenta el criterio StepGAICALL.A, así:

```

Mu link function: identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.1916	0.1983	51.391	< 2e-16 ***
ocupa.madre[T.Obrero]	0.6416	0.2339	2.743	0.00648 **
ocupa.madre[T.Ocupacion No remunerada]	0.2091	0.2108	0.992	0.32211
ocupa.madre[T.Otra actividad u ocupacion]	0.6083	0.1983	3.068	0.00237 **
ocupa.madre[T.Pensionado]	-0.2918	0.1983	-1.471	0.14233

```

---
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)	-11.5183	1.3901	-8.286	5.39e-15 ***
Horas.trabajo.semanal[T.Entre 1 y 5 horas]	-38.9232	0.5614	-69.326	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 11 y 15 horas]	-38.5826	0.1652	-233.548	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 16 y 20 horas]	-38.4826	0.1345	-286.036	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 6 y 10 horas]	-38.4614	0.2119	-181.535	< 2e-16 ***
Trabaja[T.si, como ayudante sin remuneracion]	38.5236	0.1218	316.202	< 2e-16 ***
Trabaja[T.si, por experiencia o remuneracion]	38.6983	0.2605	148.545	< 2e-16 ***
Trabaja[T.si, por ser practica en plan de estudios]	29.3586	1.6505	17.788	< 2e-16 ***
Valor.anual.matri[T.Entre 500 mil y un millon]	11.0185	1.3880	7.939	5.40e-14 ***
Valor.anual.matri[T.Entre un millon y 3 millones]	11.2436	1.4084	7.983	4.02e-14 ***
Valor.anual.matri[T.Menos de 500 mil]	11.2257	1.3892	8.081	2.12e-14 ***
Valor.anual.matri[T.No pago]	10.9409	1.5247	7.176	6.84e-12 ***

```

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

-----
Nu link function: log
Nu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.09308	0.04104	2.268	0.0241 *

```

---
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)	2.2953	0.2963	7.747	1.69e-13 ***

```

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

-----
No. of observations in the fit: 284
Degrees of Freedom for the fit: 19
Residual Deg. of Freedom: 265
at cycle: 20

Global Deviance: 725.0106
AIC: 763.0106
SBC: 832.3411
*****

```

Figura 5-45: Modelo óptimo lectura crítica factor socioeconómico

Interpretación:

A partir de la información del modelo óptimo se puede afirmar que, con un nivel de significancia del 1%, si se compara un estudiante cuya madre sea de ocupación obrera frente a otra que tenga ocupación de administradora, su puntaje aumentará 0.64 puntos.

Se puede afirmar que dada la variable respuesta, las variables que explican la variabilidad del puntaje en lectura crítica son: tener trabajo y valor anual de la matrícula, así como número de horas de trabajo a la semana.

Por lo anterior se deduce que el **PUNTAJE EN LECTURA CRÍTICA** en los estudiantes de Ingeniería Electromecánica, según factores socioeconómicos, está ligado al tipo de ocupación que tenga la madre.

Las siguientes gráficas muestran el ajuste de los residuos, los cuales deben seguir una distribución normal estándar:

```
*****
                Summary of the Quantile Residuals
                mean      = -0.01224167
                variance  =  1.071633
                coef. of skewness =  0.2342847
                coef. of kurtosis =  2.763181
Filliben correlation coefficient =  0.9945179
*****
```

Figura 5-46: Prueba ajuste modelo socioeconómico lectura crítica

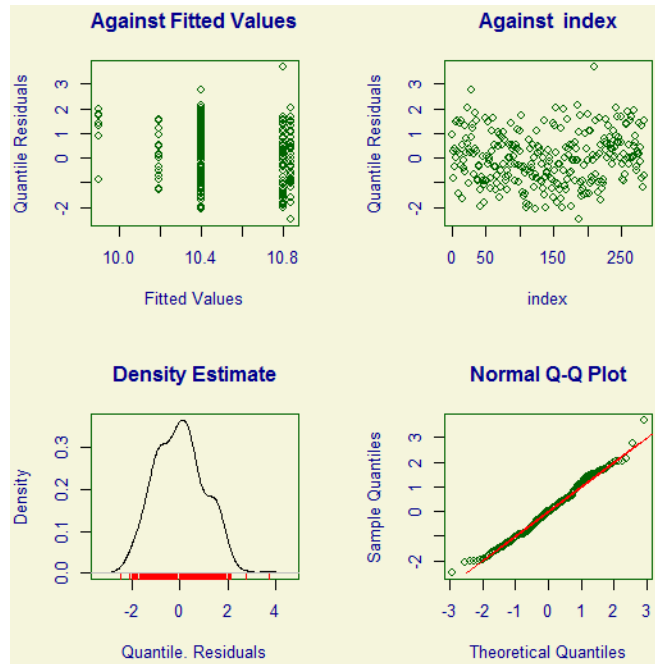


Figura 5-47: Ajuste de la distribución modelo puntaje lectura crítica factor socioeconómico

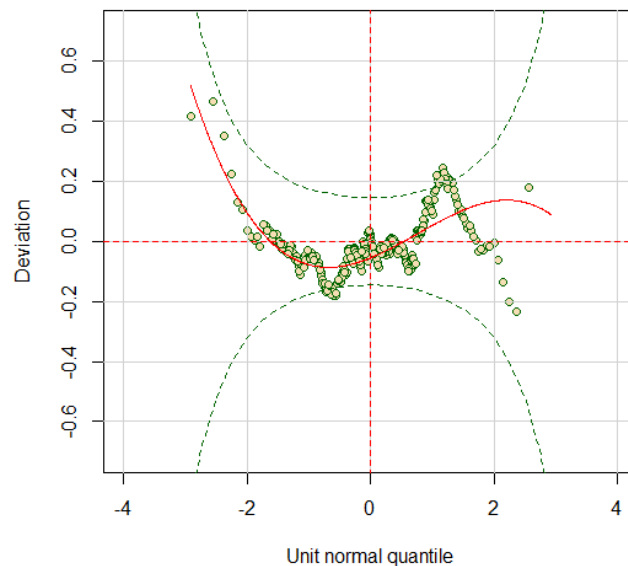


Figura 5-48: Ajuste de la distribución modelo lectura crítica factor socioeconómico

En las gráficas anteriores se determina que los residuales aleatorizados para el modelo óptimo en lectura crítica según factor socioeconómico presenta un buen ajuste, esto quiere decir que la distribución ST3 es la más adecuada para dicho comportamiento. Sin embargo es importante ver

que en la **Figura 5-48** se observa a los residuos en forma de S estirada a la izquierda y da un giro, lo que indica residuos platicúrticos, es decir que existe una altísima curtosis.

Para la selección y mejor ajuste del modelo óptimo en cuanto a comunicación escrita se tuvo en cuenta el criterio StepGAICALL.A.

```

Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.04923    0.06739 149.126  <2e-16 ***
Internet[T.Si]  0.21128    0.10987   1.923   0.0555 .
---
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)      -0.4122    0.1421  -2.901  0.00402 **
ocupa.padre[T.Obrero]  -0.3350    0.1737  -1.928  0.05483 .
ocupa.padre[T.Ocupacion No remunerada] -0.3095    0.5620  -0.551  0.58232
ocupa.padre[T.Otra actividad u ocupacion] -0.2103    0.1584  -1.328  0.18528
ocupa.padre[T.Pensionado]  -0.5102    0.2095  -2.436  0.01550 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  284
Degrees of Freedom for the fit:  7
  Residual Deg. of Freedom:  277
                at cycle:  3

Global Deviance:    758.0177
      AIC:          772.0177
      SBC:          797.5606
*****

```

Figura 5-49: Modelo óptimo comunicación escrita factor socioeconómico

Interpretación:

A partir de la información del modelo óptimo en comunicación escrita se puede afirmar que, con un nivel de significancia del 1%, si se compara un estudiante que tenga servicio de internet con otro que no, su puntaje aumentará 0.21 puntos.

Se puede afirmar que la variabilidad del puntaje en comunicación escrita está determinada en el modelo por la variable ocupación del padre.

Por lo anterior se deduce que el **PUNTAJE EN COMUNICACIÓN ESCRITA** en los estudiantes de Ingeniería electromecánica, según factores socioeconómicos, está asociado al hecho de tener servicio de internet en el domicilio.

Las siguientes gráficas muestran el ajuste de los residuos y se podrá observar si su distribución se asemeja a una normal estándar:

```

*****
Summary of the Quantile Residuals
      mean      = 0.009913552
      variance   = 0.9952019
      coef. of skewness = -0.1294461
      coef. of kurtosis  = 2.839661
Filliben correlation coefficient = 0.993331
*****

```

Figura 5-50: Prueba ajuste modelo socioeconómico comunicación escrita

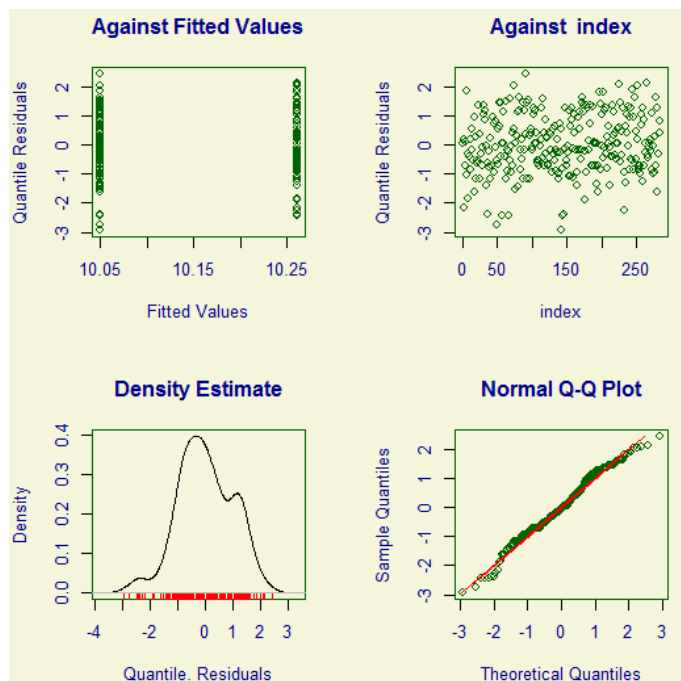


Figura 5-51: Ajuste de la distribución modelo puntaje comunicación escrita factor socioeconómico

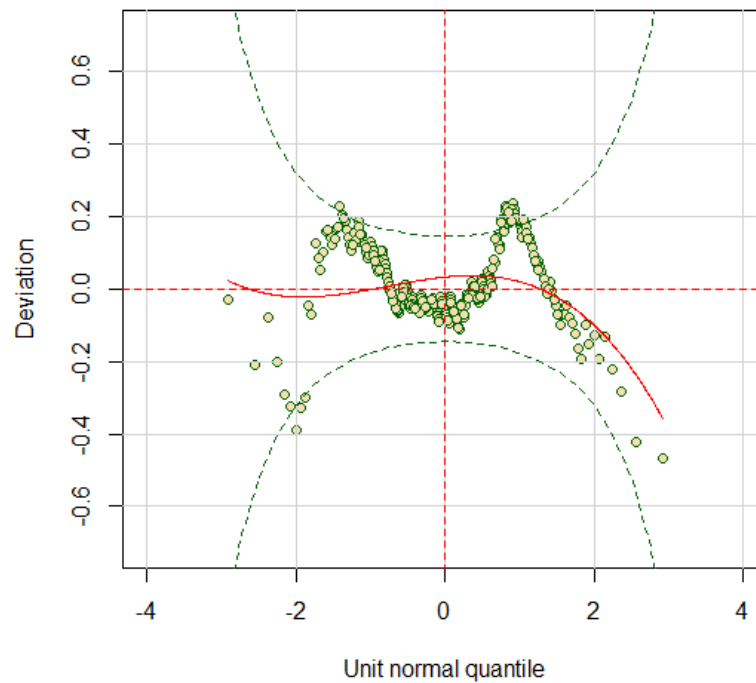


Figura 5-52: Ajuste de la distribución modelo puntaje comunicación escrita factor socioeconómico

En las gráficas anteriores se puede observar que los residuales aleatorizados para el modelo óptimo en el **PUNTAJE DE COMUNICACIÓN ESCRITA** según factores socioeconómicos presentan un buen ajuste a una normal estándar, esto quiere decir que la distribución Logistic es la más adecuada. Sin embargo es importante ver que se presenta residuos platocúrticos, es decir que existe una alta curtosis.

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

Tabla 18. Resumen modelo óptimo desempeño comunicación escrita factor socioeconómico

	Coef. estimado	Error	Wald	p-valor	OR	I.C. 95%	
Horas.trabajo.semanal[T.Entre 1 y 5 horas]	1.5908	0.8151	1.951	0.0519	4.90	0.993	24.24
Horas.trabajo.semanal[T.Entre 11 y 15 horas]	0.0855	0.4217	0.202	0.8400	1.08	0.476	2.489
Horas.trabajo.semanal[T.Entre 16 y 20 horas]	-0.4787	0.2629	-1.821	0.0696	0.619	0.370	1.03
Horas.trabajo.semanal[T.Entre 6 y 10 horas]	-0.0265	0.3668	-0.072	0.9426	0.973	0.474	1.99
Internet[T.si]	0.4588	0.2307	1.98	0.0486	1.582	1.004	2.48
Nivel 1 Nivel 2	-4.0048	0.4758	-8.41				
Nivel 2 Nivel 3	-3.5237	0.3890	-9.05				
Nivel 3 Nivel 4	-2.0759	0.2425	-8.56				
Nivel 4 Nivel 5	0.3025	0.1932	1.56				
Nivel 5 Nivel 6	1.7627	0.2246	7.84				
Nivel 6 Nivel 7	3.8433	0.4178	9.19				

Interpretación del modelo

Según se observa, la variable tener internet, en su razón de probabilidad y su intervalo de confianza al 95% es mayor que uno, lo que significa que la probabilidad acumulada, comenzando de Nivel de desempeño A- hasta el nivel B1, aumenta los valores que puede tomar inglés, así un estudiante que cuente con el servicio de internet es más probable que obtenga niveles altos en el desempeño de comunicación escrita en la prueba Saber Pro.

Por otra parte la variable horas de trabajo a la semana contiene a uno en sus distintos intervalos de confianza, es decir que no influyen en este desempeño para estudiantes de Ingeniería Electromecánica (significancia del 5%), asegurando que trabajar un determinado número de horas a la semana no incide en el desempeño en comunicación escrita de la prueba.

A continuación se presenta el modelo óptimo en cuanto a inglés, según factores socioeconómicos. Para la selección y mejor ajuste del modelo se tuvo en cuenta el criterio StepGAICALL.A.

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.72628	0.10216	95.208	< 2e-16 ***
Estrato[T.Estrato 2]	-0.35258	0.10928	-3.226	0.00140 **
Estrato[T.Estrato 3]	-0.17139	0.12481	-1.373	0.17080
Estrato[T.Estrato 4]	0.06673	0.19501	0.342	0.73247
salario.mensual.fami[T.Entre 2 y menos de 3 SM]	0.26853	0.09071	2.960	0.00334 **
salario.mensual.fami[T.Entre 3 y menos de 5 SM]	0.52019	0.18257	2.849	0.00471 **
salario.mensual.fami[T.Entre 5 y menos de 7 SM]	2.04492	0.19346	10.570	< 2e-16 ***
salario.mensual.fami[T.Mas de 10 SM]	2.94847	0.67748	4.352	0.000019 ***
salario.mensual.fami[T.Menos de 1 SM]	-0.03541	0.08719	-0.406	0.68495

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)	24.5017686	0.2647255	92.555	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 1 y 5 horas]	-23.1230779	0.4200918	-55.043	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 11 y 15 horas]	-22.9203926	0.2430917	-94.287	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 16 y 20 horas]	-23.1153315	0.1861786	-124.157	< 2e-16 ***
Horas.trabajo.semanal[T.Entre 6 y 10 horas]	-22.8260440	0.2455301	-92.966	< 2e-16 ***
salario.mensual.fami[T.Entre 2 y menos de 3 SM]	-0.0067532	0.1737198	-0.039	0.96902
salario.mensual.fami[T.Entre 3 y menos de 5 SM]	-0.0005569	0.3419379	-0.002	0.99870
salario.mensual.fami[T.Entre 5 y menos de 7 SM]	-0.2842856	0.3923689	-0.725	0.46937
salario.mensual.fami[T.Mas de 10 SM]	10.6032441	3.0881353	3.434	0.00069 ***
salario.mensual.fami[T.Menos de 1 SM]	-0.0240795	0.1731484	-0.139	0.88950
Trabaja[T.si, como ayudante sin remuneracion]	22.6645271	0.1525913	148.531	< 2e-16 ***
Trabaja[T.si, por experiencia o remuneracion]	22.7387013	0.2937531	77.408	< 2e-16 ***
Trabaja[T.si, por ser practica en plan de estudios]	26.3026148	3.3273701	7.905	7.09e-14 ***
Valor.anual.matri[T.Entre 500 mil y un millon]	-24.4513744	0.2574962	-94.958	< 2e-16 ***
Valor.anual.matri[T.Entre un millon y 3 millones]	-24.7658399	0.3903479	-63.446	< 2e-16 ***
Valor.anual.matri[T.Menos de 500 mil]	-24.5265670	0.2530829	-96.911	< 2e-16 ***
Valor.anual.matri[T.No pago]	-21.6206222	2.9555469	-7.315	3.01e-12 ***

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)	2.6098	0.5864	4.450	0.000124 ***
salario.mensual.fami[T.Entre 2 y menos de 3 SM]	0.1980	1.0573	0.187	0.852
salario.mensual.fami[T.Entre 3 y menos de 5 SM]	-0.9096	1.4473	-0.628	0.530
salario.mensual.fami[T.Entre 5 y menos de 7 SM]	-5.1330	3.2317	-1.588	0.113
salario.mensual.fami[T.Mas de 10 SM]	426574.3495	100000.0000	4.266	0.0000273 ***
salario.mensual.fami[T.Menos de 1 SM]	-0.4112	0.8557	-0.481	0.631

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)	1.0093	0.1837	5.495	0.0000000872 ***

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

No. of observations in the fit: 284
Degrees of Freedom for the fit: 33
Residual Deg. of Freedom: 251
at cycle: 20

Global Deviance: 827.0577
AIC: 893.0577
SBC: 1013.474

Figura 5-53: Modelo óptimo puntaje inglés factor socioeconómico

Interpretación:

A partir de la información del modelo óptimo en el puntaje de Inglés se puede afirmar que, con un nivel de significancia del 1%, si se compara un estudiante de estrato 2 frente a otro que pertenezca al estrato 1, su puntaje en la prueba disminuirá 0.35 puntos.

Por otra parte si se realiza una comparación teniendo en cuenta el salario mensual familiar se tiene que, un estudiante en cuyo hogar reciban más de dos salarios al mes, frente a otro estudiante que en su hogar ingresen entre 1 y menos de 2 salarios mínimos, con cualquier significancia, el puntaje en inglés aumentará.

Se puede afirmar que dada la variable respuesta, las variables que explican la variabilidad del puntaje en inglés son: horas de trabajo a la semana, valor anual de la matrícula, tener trabajo y salario mensual familiar.

Por lo anterior se deduce que el puntaje inglés en los estudiantes de Ingeniería Electromecánica, según factores socioeconómicos, está asociado al estrato según recibo de energía y al número de salarios mensuales que ingresen en la familia.

Las siguientes gráficas muestran el ajuste de los residuos:

```
*****  
Summary of the Quantile Residuals  
      mean      = 0.005025226  
      variance   = 1.242472  
      coef. of skewness = -1.063972  
      coef. of kurtosis  = 9.033719  
Filliben correlation coefficient = 0.9687856  
*****
```

Figura 5-54: Prueba ajuste modelo socioeconómico inglés

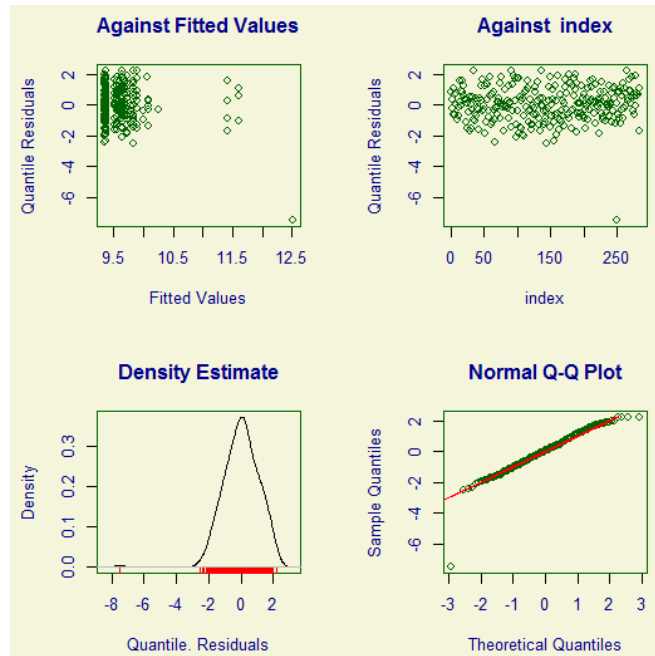


Figura 5-55: Ajuste de la distribución modelo puntaje inglés factor socioeconómico

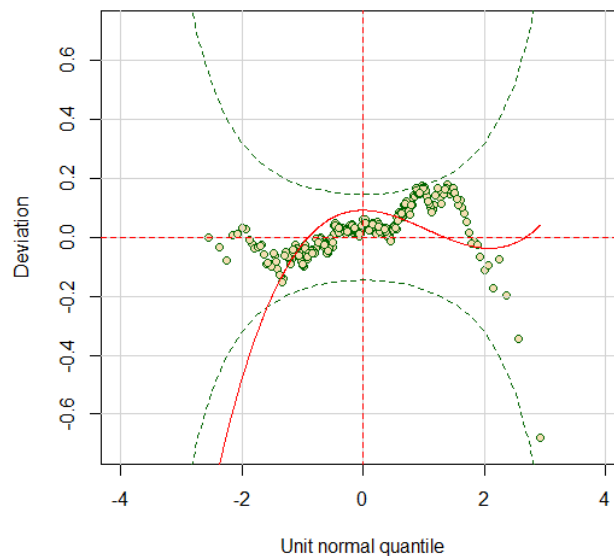


Figura 5-56: Ajuste de la distribución modelo inglés factor socioeconómico

En las figuras 5-55 y 5-56 se observa que los residuales aleatorizados para el modelo óptimo del **PUNTAJE EN INGLÉS** a través de factores socioeconómicos presentan un buen ajuste, esto quiere decir que la distribución Skew t type 2 es la más adecuada para explicar su comportamiento.

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

Tabla 19. Resumen modelo óptimo Inglés factor socioeconómico

	Coef. estimado	Error	Wald	p-valor	Odds Ratio	I.C. 95%	
Estrato[T.Estrato 2]	-0.97	0.372	-2.604	0.0097	0.37	0.18	0.78
Estrato[T.Estrato 3]	-0.46	0.415	-1.117	0.2649	0.63	0.27	1.4
Estrato[T.Estrato 4]	-1.10	0.835	-1.318	0.1885	0.33	0.064	1.7
Trabaja[T.si, como ayudante sin remuneracion]	-0.78	0.233	-3.337	0.0009	0.45	0.28	0.72
Trabaja[T.si, por experiencia o remuneracion]	-0.71	0.398	-1.802	0.0726	0.48	0.22	1.08
Trabaja[T.si, por ser práctica plan de estudios]	14.67	4e-8	3e+9	0	2364554	2.3e+6	2.3e+6
Valor.anual.matri[T.Entre 500 mil y un millon]	-12.63	0.306	-41.20	1.3e-121	3.2e-6	1.7e-6	5.9e-6
Valor.anual.matri[T.Entre un millon y 3 millones]	-12.38	0.409	-30.24	1.2e-90	4.19e-6	1.8e-6	9.3e-6
Valor.anual.matri[T.Menos de 500 mil]	-13.11	0.286	-45.84	5.3e-133	2.01e-6	1.1e-6	3.5e-6
Valor.anual.matri[T.No pago]	-11.35	0.956	-11.8	1.1e-26	1.1e-5	1.8e-6	7.6e-5
A- A1	-16.22	0.417	-38.87				
A1 A2	-14.47	0.394	-36.73				
A2 B+	-13.14	0.385	-34.12				
B+ B1	-12.69	0.385	-32.90				

En la Tabla 19 se observa que la variable estrato en su razón de probabilidad y su intervalo de confianza al 95% es menor a uno únicamente para el estrato 2 vs estrato 1 (para los demás no es significativo), lo que significa que la probabilidad acumulada, comenzando de Nivel de desempeño A- hasta el nivel B1, decrece en todos los valores que puede tomar inglés, así un estudiante perteneciente al estrato 2 es más probable que obtenga niveles más bajos en el desempeño en inglés de la prueba Saber Pro comparado con un estudiante de estrato 1.

Además se puede decir que la variable tener trabajo aumenta en todos los valores que puede tomar inglés, si se trabaja por ser práctica en el plan de estudios frente a estudiantes que no tengan trabajo; por el contrario, tener trabajo como ayudante sin remuneración reduce la

posibilidad de tener un nivel de desempeño más alto en la prueba Saber Pro, frente a un estudiante sin carga laboral.

Por otra parte el valor anual de la matrícula es menor a 1 en todos sus intervalos de confianza para cada una de sus categorías, es decir que, todos los valores que puede tomar inglés en este desempeño disminuyen, en contraste con estudiantes que paguen entre 3 y 5 millones por el costo anual de la matrícula en la carrera de Ingeniería Electromecánica.

6 DISCUSIÓN

Situados en este contexto, el planteamiento de Bogoya (2013), dice que, “el estrato socioeconómico refleja una diferencia significativa en los puntajes del componente genérico de las pruebas Saber Pro”, pero en nuestro caso no se evidencia esto (con excepción del puntaje en inglés y su desempeño donde sí incide). Así mismo Bogoya concluye que: “La UPTC posee cierto nivel por debajo de la media en aspectos puntuales como inglés, y lectura crítica, pero al mismo tiempo supera la media nacional en aspectos como razonamiento cuantitativo y escritura” (Bogoya, 2013), lo cual es apoyado por este trabajo, ya que, para la carrera de Ingeniería Electromecánica se tiene un nivel de desempeño en la categoría “excelente” en razonamiento cuantitativo, que se contrarresta con la categoría “regular” para inglés.

Con base en los resultados de este trabajo se puede decir que la ocupación de la madre determina un aspecto crucial en los resultados de las pruebas, ya que, al ser obreras o amas de casa influyen de manera positiva a los puntajes en Razonamiento cuantitativo y lectura crítica. Así mismo tener servicio de internet aumenta la posibilidad de obtener mejores habilidades en comunicación escrita y su respectivo nivel de desempeño. En cuanto a la variable estrato se puede afirmar que comparando dos estudiantes de estrato 1 y 2, el puntaje en inglés y su desempeño mostrará por lo general mejor calificación en el estudiante de estrato 1.

Ahora bien, las conclusiones de Arias y Ávila en su trabajo: influencia de los padres en el rendimiento académico de los hijos, concluyen: “ a un mayor nivel educativo de los padres mayor será la probabilidad del estudiante de obtener un puntaje alto” (Arias y Ávila, 2014), lo cual se ajusta a lo encontrado en esta investigación, siendo más precisos, un mayor nivel educativo de la madre, es decir, con estudios técnicos o universitarios aumentan la probabilidad de obtener mejores puntajes en razonamiento cuantitativo, así mismo un estudiante de Ingeniería Electromecánica con un padre cuyo nivel educativo sea universitario o técnico favorecerá su puntaje en las pruebas saber pro en comunicación escrita, inglés y su respectivo desempeño. Los autores sustentan este hecho en la relación padres-hijos en el sentido de cercanía (a mayor cercanía mejor rendimiento).

En esta investigación se logró además establecer que, estudiantes egresados de colegios con títulos de normalistas y técnicos obtienen mejores resultados sobre estudiantes con título académico en el puntaje en lectura crítica.

En el caso del factor demográfico al analizar la edad de los estudiantes se encontró que, al comparar dos de ellos que difieran en su edad, aquel que sea menor tendrá más probabilidad de obtener mejores puntajes en razonamiento cuantitativo, lectura crítica, comunicación escrita e inglés, sin embargo para el desempeño en inglés la relación de la edad es inversa a la anterior, es decir, que estudiantes de mayor edad obtendrán un nivel más alto que estudiantes más jóvenes.

Al analizar la variable sexo del estudiante se encontró que, los hombres tienen más posibilidad de obtener mejores puntajes en razonamiento cuantitativo, mientras una mujer se destaca más en comunicación escrita y sus respectivos niveles de desempeño. Lo anterior genera un contraste particular con lo enunciado por Arias y Ávila, quienes no hallaron diferencias significativas en las pruebas de comunicación escrita, pero sí lo hicieron en las pruebas de razonamiento cuantitativo.

Respecto al estado civil, esta investigación arrojó que, en el puntaje en inglés y su respectivo desempeño, el ser casado favorece en cierta medida los resultados en estos componentes, frente a estudiantes que sean solteros, separados o vivan en unión libre.

Finalmente, un estudiante cuya situación del hogar pertenezca a la categoría temporal, aumenta la probabilidad de obtener mejores puntajes en la prueba de razonamiento cuantitativo en las Saber Pro, comparado con un estudiante que tenga un hogar permanente.

Las anteriores conclusiones se pueden complementar con las siguientes tablas resumen, donde se muestra los factores que influyen aun nivel de significancia del 1%, 5% o 10% para cada componente, siendo RC: razonamiento cuantitativo, LC: lectura crítica, CE: comunicación escrita, IN: inglés, D.CE: desempeño comunicación escrita, D.IN: desempeño inglés, ahora observemos que:

“↑” indica que la variable influye a la hora de obtener puntajes altos.

“↓” indica la posibilidad de obtener valores bajos.

“-” indica que la variable no hace parte, o no resultado significativa en el modelo óptimo.

Tabla IV. Resumen variables demográficas influyentes

VARIABLES DEMOGRÁFICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Sexo (M)	↑	-	-	-	↓	-
Sexo (F)	↓	-	-	-	↑	-
Edad	↑	↑	↑	↑	↑	-
Estado civil (Casado)	-	-	-	↑	-	↑
Estado civil (Soltero)	-	-	-	↓	-	↓
Estado civil (Unión libre)	-	-	-	↓	-	↓
Estado civil (Separado)	-	-	-	↓	-	↓

Hogar actual (Permanente)	↓	-	-	-	-	-
Hogar actual (Temporal)	↑	-	-	-	-	-

Se observa que tener estado civil de casado genera mejores puntajes en los componentes de inglés y desempeño en inglés, en comparación a estudiantes con otro tipo de estado civil, además contar con un hogar temporal también aumenta la posibilidad de tener puntajes altos en comparación a quienes cuentan con un hogar permanente en razonamiento cuantitativo.

En cuanto a la edad para cada componente con excepción del desempeño en inglés se tiene que, el ser más joven aumenta la posibilidad de obtener puntajes más altos. Los hombres obtienen mejor puntaje en el componente de razonamiento cuantitativo mientras las mujeres lo hacen en el componente de desempeño en comunicación.

Tabla V. Resumen variables académicas influyentes

VARIABLES ACADÉMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Tipo de bachillerato (Técnico)	-	↑	-	-	-	-
Tipo de bachillerato (Académico)	-	↓	-	-	-	-
Tipo de bachillerato (Normalista superior)	-	↑	-	-	-	-
Nivel educativo padre (Universitario)	-	-	↑	↑	-	↑
Nivel educativo padre (primaria)	-	-	↓	↓	-	↓
Nivel educativo padre (Técnico)	-	-	↑	↑	-	↑
Nivel educativo madre (Universitario)	↑	-	-	-	-	-
Nivel educativo madre (Primaria)	↓	-	-	-	-	-
Nivel educativo madre (Secundaria)	↑	-	-	-	-	-
Nivel educativo madre (Sin estudio)	↓	-	-	-	-	-

Por otra parte, en cuanto al factor académico se tiene que, tener madre con nivel educativo universitario y secundaria mejora los puntajes en razonamiento cuantitativo, en comparación a tener niveles educativos de primaria y sin estudio, así mismo tener padre con nivel educativo universitario y técnico aumenta el puntaje en comunicación escrita e inglés, además obtener un tipo de bachillerato académico disminuye la posibilidad de lograr puntajes altos respecto a un bachiller de tipo técnico en lectura crítica.

Tabla VI. Resumen variables socioeconómicas influyentes

VARIABLES SOCIOECONÓMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Valor anual matrícula (Menos de 500 mil pesos)	-	-	-	-	-	↓
Valor anual matrícula (Entre 1 y 3 millones de pesos)	-	-	-	-	-	↓
Valor anual matrícula (Entre 3 y 5 millones de pesos)	-	-	-	-	-	↑
Ocupación madre (Administrador)	↓	↓	-	-	-	-
Ocupación madre (Obrera)	↑	↑	-	-	-	-
Conexión a internet (Sí)	-	-	↑	-	-	-
Conexión a internet (No)	-	-	↓	-	-	-
Trabaja actualmente (Sí, con remuneración en dinero)	-	-	-	-	-	↓
Trabaja actualmente (Sí, por experiencia)	-	-	-	-	-	↓
Trabaja actualmente (Sí, práctica obligatoria)	-	-	-	-	-	↓
Trabaja actualmente (No)	-	-	-	-	-	↑
Horas de trabajo a la semana (1-5)	↑	-	-	-	↑	-
Horas de trabajo a la semana (6-10)	↓	-	-	-	↓	-
Horas de trabajo a la semana (11-15)	↓	-	-	-	↑	-
Horas de trabajo a la semana (16-20)	↓	-	-	-	↓	-
Estrato 1	-	-	-	↑	-	↑
Estrato 2	-	-	-	↓	-	-
Estrato 3	-	-	-	↓	-	-
Estrato 4	-	-	-	↑	-	-
Salario mensual familiar(1-2)	-	-	-	↓	-	-
Salario mensual familiar(2-3)	-	-	-	↑	-	-
Salario mensual familiar(3-5)	-	-	-	↑	-	-
Salario mensual familiar(5-7)	-	-	-	↑	-	-
Salario mensual familiar(+10)	-	-	-	↑	-	-

Se aprecia que realizar pagos de matrícula entre 3 y 5 millones de pesos, aumenta la posibilidad de obtener puntajes altos en desempeño en inglés, en comparación a quienes realizan pagos de cualquier otro valor. Por otra parte, al tener madre con ocupación de administradora se obtiene puntajes más bajos en el componente de razonamiento cuantitativo y lectura crítica, comparado con cualquier otro tipo de ocupación.

En cuanto a la conexión a internet, obtiene más puntos en comunicación escrita quien posea este servicio, así también como un estudiante que no trabaje.

Ahora bien, un estudiante que trabaje semanalmente entre 1 y 5 horas obtendrá mejores puntajes respecto a cualquier categoría comparativa en razonamiento cuantitativo, igual caso en el desempeño en comunicación escrita, con excepción de la categoría de trabajar entre 11 y 15 horas a la semana donde también aumenta esta probabilidad.

En el puntaje en inglés, pertenecer a estrato 1 y 4 además de tener un salario mensual familiar superior a 2 salarios mínimos legales vigentes mejora este valor.

Dada la naturaleza de algunos factores de estudio, se tratará de establecer acciones encaminadas al mejoramiento institucional, para ello se podría indicar que:

- ✓ La carrera de Ingeniería Electromecánica y en general la Universidad debe continuar y ampliar el programa de residencias universitarias, ya que estudiantes con situación de hogar temporal destacan por lo general en las pruebas Saber Pro.
- ✓ Cada carrera otorgue algún beneficio o reconocimiento al mejor puntaje de las pruebas Saber Pro durante cada año, con el fin de incentivar (aspecto motivacional) la competencia, evitando que la prueba sea vista como un simple requisito para la graduación.
- ✓ Al asumir que el internet es una herramienta que incrementa la posibilidad de mejores rendimientos en la prueba Saber Pro, optimizar las áreas de cobertura de dicho servicio, garantizando la opción de investigación y búsqueda de información (En mi concepto motiva al ser humano a leer y escribir con comprensión).
- ✓ Se debe evaluar cómo mejorar los puntajes y desempeños en comunicación escrita, el cual es el punto débil de la carrera, y para ello se sugiere analizar si los contenidos que se enseñan en el área general está capacitando apropiadamente a estos futuros profesionales.
- ✓ Aunque no se trabajó la variable: tomar curso por la cantidad de datos faltantes, en la literatura consultada se pudo establecer que un refuerzo, curso o inducción de lo que es la prueba y su componente genérico, puede llegar a mejorar considerablemente el rendimiento en la misma, por ello la universidad podría ofrecer un refuerzo gratuito o por una módica suma a los estudiantes, lo cual brindaría un mayor nivel de preparación.

- ✓ En los resultados se evidenció que los estudiantes que mantienen una relación estrecha con sus padres adquieren un sentido mayor de responsabilidad, punto que podría usar a favor de la Universidad involucrando a los padres en charlas o pequeñas conferencias sobre el apoyo a los hijos (al menos por los primeros semestres), trayendo consigo el mejoramiento, no solo, de las pruebas Saber Pro, si no del nivel académico en general. “A mayor cercanía mejor rendimiento” (Arias y Ávila, 2.014).

- ✓ También se sugiere realizar capacitaciones para los docentes y directivos sobre las características de las pruebas Saber Pro, favoreciendo la adecuación y/o transición de contenidos por parte del docente al estudiante, reforzando intrínsecamente las habilidades para afrontar dicha prueba.

7 CONCLUSIONES

Al momento de iniciar a trabajar en la base de datos suministrada por el ICFES se procedió a seleccionar aquellas variables que daban un sentido realmente crítico a lo que se deseaba hallar, es así como se pasó de trabajar con 63 variables a 24, ya que algunas presentaban bastantes datos faltantes, otras no aportaban información relevante y además se tuvo en cuenta codificaciones más precisas con el fin de concluir de manera más significativa.

Ahora bien, al determinar aspectos de correlación y asociación de variables se logró observar que la edad, el sexo, el número de horas de trabajo a la semana, así como la educación de los padres y el salario mensual familiar inciden en los puntajes y desempeños de los estudiantes de Ingeniería Electromecánica. Luego de establecer estos aspectos se procedió a comprobar la normalidad en las variables respuesta, es decir en los puntajes del componente genérico, hallando que no seguían una distribución normal, por lo tanto se decidió implementar los modelos aditivos generalizados para localización, escala y forma (GAMLSS) a la investigación. En este sentido se obtuvo una adecuada distribución para cada uno de los puntajes asociados al componente genérico, para finalmente asociar las variables respuesta cada tipo de factor de manera independiente, obteniendo como resultado 18 modelos (6 por cada factor).

Para **Razonamiento Cuantitativo** los factores demográficos que inciden en este componente son la edad del estudiante al momento de presentar la prueba, la situación o tipo de hogar en el que se encuentre y finalmente su sexo. El puntaje según variables académicas sólo es determinado por el nivel educativo de la madre, mientras que, para el factor socioeconómico se asocian el número de horas de trabajo en la semana, y la ocupación de la madre.

En **Lectura Crítica** la edad en el aspecto demográfico, el tipo de bachillerato realizado por el estudiante al finalizar la educación media por la parte académica, y en el aspecto socioeconómico la ocupación de la madre.

Para **Comunicación Escrita** el puntaje se explica por la edad, el estado civil y el sexo del estudiante en el aspecto demográfico. Para el factor académico el nivel educativo del padre, y en las variables socioeconómicas el tener servicio de Internet en la residencia.

En **Inglés** como variables demográficas la edad y el estado civil, en las académicas el nivel educativo del padre, y para el factor socioeconómico el estrato, además del salario mensual familiar.

Para los casos del **desempeño en comunicación escrita** y **desempeño en inglés** se realizó una regresión logística ordinal y se tuvo en cuenta el criterio Akaike (AIC) en la selección del modelo.

Para **desempeño en comunicación escrita** se encontró que la variable demográfica sexo del estudiante incide en el desempeño, además no existe factor académico que explique el nivel en comunicación escrita, por otra parte el tener servicio de internet si se relaciona con este módulo.

En **el desempeño en inglés** el estado civil como variable demográfica, el nivel educativo del padre en cuanto a factor académico, el estrato, tener trabajo y el valor anual de la matrícula como parte socioeconómica explica el nivel de desempeño de un estudiante.

REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (2002), *Análisis Categórico de Datos*, Universidad de la Florida, Gainesville, Florida, segunda edición, Publicaciones corporación John Wiley e hijo, Hoboken, Nueva Jersey.
- Arias I., Ávila 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 DC, *Revista educación y desarrollo social*, Bogotá.
- Bogoya, D. (2013), *Benchmarking: elementos de calidad de la educación superior en Colombia*, SCIMAGO: base de datos universidad nacional de Colombia.
- Bogoya, D. (2009), *Evaluación de la Calidad de la Educación Superior y valor académico agregado*, *Revista: Anfibios académicos: pedagogías, docencia y evaluación en la educación superior*, Bogotá, Colombia, Universidad Nacional de Colombia.
- Brunner, J. Gomes, C. Fordham, E. (2016), *Revisión de políticas nacionales de educación: La educación en Colombia*, Organización para la Cooperación y el Desarrollo Económicos (OCDE), paginas 15-75, 268-329.
- Cañadas, L. (2013), *Regresión logística: tratamiento computacional con R*, Universidad de Granada, España.
- Díaz, L. (2012), *Análisis estadístico de datos categóricos*, Universidad Nacional de Colombia UNAL, Bogotá.
- Egea J., Kessler M., *Regresión lineal con R commander*, Universidad Politécnica de Cartagena, Departamento de Matemática Aplicada y Estadística.
- Erazo, O. (2012). *El rendimiento académico, un fenómeno de múltiples relaciones y complejidades*. *Revista vanguardia psicológica, clínica teórica y práctica*, 2(2) ,144-173.
- Escuela de ingeniería electromecánica, Universidad Pedagógica y Tecnológica de Colombia UPTC (2016), *Acta de comité curricular sobre las pruebas saber pro en la escuela de ingeniería electromecánica de la UPTC seccional Duitama*.
- Guzmán Tovar C., Serna, C. & Hoyos, D. (2012). *Las pruebas ECAES en Colombia: una evaluación a la evaluación*. *Panorama*, 6 (10), 33-54.
- Iglesias, T. (2013), *Métodos de bondad de ajuste en Regresión logística*, Universidad de Granada.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2014). Bases de datos pruebas Saber pro. [On-line]. Disponible en: <http://www.icfesinteractivo.gov.co>.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2011), Orientaciones para el examen de Estado de calidad de la educación superior saber pro (ecaes), prueba de competencias comunes del área de educación.

Isáziga, C. Coello, J (2014), Estudios sobre calidad de la educación en Colombia, Instituto colombiano para la evaluación de la educación (ICFES).

Mesa, H. (2009), De los ECAES, la autonomía universitaria y el derecho, Revista Electrónica.

Ministerio de educación, cultura y deporte, subdirección general de cooperación internacional europea (2003), Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación, Ministerio de educación, cultura y deporte, subdirección general de cooperación internacional europea, consejo de Europa para la publicación en inglés y francés.

Mundial, (2012), “La educación superior en Colombia 2012”, Organización para la Cooperación y el Desarrollo Económicos (OCDE).

OCDE, (2016). Revisión de políticas nacionales de educación: Educación en Colombia.

Salas, W. formación por competencias en educación superior. Una aproximación conceptual a propósito del caso colombiano, Revista Iberoamericana de Educación (ISSN: 1681-5653) Universidad de Antioquia, Colombia.

Pérez, M. (2016), Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS), Trabajo para optar al título de master, Universidad de Vigo, Pontevedra, España.

Sancho, R. (2012), Ministerio de Ciencia y Tecnología, Madrid, España.

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.

Torrado C., Teichler U. (2014), Estudios sobre calidad de la educación en Colombia: 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, Instituto Colombiano para la Evaluación de la Educación – ICFES, Bogotá.

ANEXOS

*ANEXO A:
RESULTADOS
ESTADÍSTICOS
ANÁLISIS
BIVARIADO*

A continuación aparecen los anexos que surgieron a lo largo del desarrollo del trabajo investigativo, en los cuales se soporta las técnicas estadísticas que se emplearon al analizar las variables tomadas de la base de datos suministrada por el ICFES.

Las siguientes imágenes muestran si existe o no correlación entre algunas variables, con un nivel de significancia del 5%.

Razonamiento cuantitativo vs Edad:

```
Pearson's product-moment correlation  
data: Edad and puntaje.razo.cuanti  
t = -5.2185, df = 285, p-value = 0.000000348  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.3975154 -0.1859025  
sample estimates:  
cor  
-0.2953268
```

Razonamiento cuantitativo vs número de personas a cargo:

```
Pearson's product-moment correlation  
data: personas.a.cargo and puntaje.razo.cuanti  
t = -0.39382, df = 285, p-value = 0.694  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.13872793 0.09270963  
sample estimates:  
cor  
-0.02332161
```

Razonamiento cuantitativo vs número de personas a en el hogar:

```
Pearson's product-moment correlation  
data: Personas.en.el.hogar and puntaje.razo.cuanti  
t = 0.12829, df = 285, p-value = 0.898  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1082774 0.1232712  
sample estimates:  
cor  
0.007598775
```

Razonamiento cuantitativo vs número de dormitorios:

```
Pearson's product-moment correlation  
data: Num.Dormitorios and puntaje.razo.cuanti  
t = -1.2533, df = 285, p-value = 0.2111  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.18820192 0.04210762  
sample estimates:  
cor  
-0.07403423
```

Lectura crítica vs edad:

```
Pearson's product-moment correlation  
data: Edad and puntaje.lectu.crit  
t = -3.741, df = 285, p-value = 0.0002216  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.3240151 -0.1031535  
sample estimates:  
cor  
-0.2163505
```

Lectura crítica vs número de personas a cargo:

```
Pearson's product-moment correlation
data: personas.a.cargo and puntaje.lectu.crit
t = -1.028, df = 285, p-value = 0.3048
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.17532851  0.05538925
sample estimates:
      cor
-0.06078145
```

Lectura crítica vs número de personas en el hogar:

```
Pearson's product-moment correlation
data: Personas.en.el.hogar and puntaje.lectu.crit
t = -0.1691, df = 285, p-value = 0.8658
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1256514  0.1058875
sample estimates:
      cor
-0.01001622
```

Lectura crítica vs número de dormitorios:

```
Pearson's product-moment correlation
data: Num.Dormitorios and puntaje.lectu.crit
t = -1.8299, df = 285, p-value = 0.06831
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.220789800  0.008118521
sample estimates:
      cor
-0.1077637
```

Comunicación escrita vs Edad:

```
Pearson's product-moment correlation
data: Edad and puntaje.comuni.escri
t = -1.6284, df = 282, p-value = 0.1046
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2105442  0.0200999
sample estimates:
      cor
-0.09651765
```

Comunicación escrita vs número de personas a cargo:

```
Pearson's product-moment correlation
data: personas.a.cargo and puntaje.comuni.escri
t = 0.19847, df = 282, p-value = 0.8428
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1047179  0.1280336
sample estimates:
      cor
0.01181793
```

Comunicación escrita vs número de personas en el hogar:

```
Pearson's product-moment correlation
data: Personas.en.el.hogar and puntaje.comuni.escri
t = 0.37517, df = 282, p-value = 0.7078
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.0943012  0.1383677
sample estimates:
      cor
0.0223357
```


Comunicación escrita vs número de dormitorios:

```
Pearson's product-moment correlation  
data: Num.Dormitorios and puntaje.comuni.escri  
t = 0.681, df = 282, p-value = 0.4964  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.07623147 0.15617499  
sample estimates:  
cor  
0.0405198
```

Puntaje inglés vs edad:

```
Pearson's product-moment correlation  
data: Edad and puntaje.ingles  
t = -4.1868, df = 285, p-value = 0.00003774  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.3468273 -0.1285131  
sample estimates:  
cor  
-0.2407124
```

Puntaje en inglés vs número de personas a cargo:

```
Pearson's product-moment correlation  
data: personas.a.cargo and puntaje.ingles  
t = -1.9662, df = 285, p-value = 0.05024  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.22840807678 0.00009565566  
sample estimates:  
cor  
-0.1156865
```

Puntaje en inglés vs número de personas en el hogar:

```
          Pearson's product-moment correlation
data:  Personas.en.el.hogar and puntaje.ingles
t = -0.125, df = 285, p-value = 0.9006
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1230794  0.1084699
sample estimates:
          cor
-0.007404042
```

Puntaje en inglés vs número de dormitorios:

```
          Pearson's product-moment correlation
data:  Num.Dormitorios and puntaje.ingles
t = -1.9628, df = 285, p-value = 0.05064
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2282172385  0.0002969888
sample estimates:
          cor
-0.1154879
```

Las siguientes tablas de contingencia muestran si existe o no independencia entre algunas variables con los desempeños en comunicación escrita e inglés respectivamente

Desempeño en comunicación escrita vs género:

```
Frequency table:
                Genero
Desempe.comuni.escrit Femenino Masculino
    Nivel 1             2           3
    Nivel 2             0           3
    Nivel 3             0          23
    Nivel 4             9          118
    Nivel 5            11           69
    Nivel 6             8           31
    Nivel 7             0           7

    Pearson's Chi-squared test

data: .Table
X-squared = 15.055, df = 6, p-value = 0.01983
```

Desempeño en comunicación escrita vs estado civil:

```
Frequency table:
                Estado.civil
Desempe.comuni.escrit Casado Separado Soltero Union libre
    Nivel 1             0           0           5           0
    Nivel 2             0           0           3           0
    Nivel 3             1           0          21           1
    Nivel 4             0           1         121           5
    Nivel 5             0           0          78           2
    Nivel 6             1           1          36           1
    Nivel 7             0           0           6           1

    Pearson's Chi-squared test

data: .Table
X-squared = 13.806, df = 18, p-value = 0.7417
```

Desempeño en comunicación escrita vs situación del hogar:

```
Frequency table:
                Situacion.hogar
Desempe.comuni.escrit  Permanente  Temporal
Nivel 1                 3           2
Nivel 2                 2           1
Nivel 3                15           8
Nivel 4                90          37
Nivel 5                59          21
Nivel 6                26          13
Nivel 7                 7           0

        Pearson's Chi-squared test

data: .Table
X-squared = 4.2132, df = 6, p-value = 0.6479
```

Desempeño en comunicación escrita vs estrato:

```
Frequency table:
                Estrato
Desempe.comuni.escrit  Estrato 1  Estrato 2  Estrato 3  Estrato 4
Nivel 1                 1           4           0           0
Nivel 2                 1           1           1           0
Nivel 3                 4          13           6           0
Nivel 4                 9          81          34           3
Nivel 5                10          55          14           1
Nivel 6                 3          26           9           1
Nivel 7                 0           6           1           0

        Pearson's Chi-squared test

data: .Table
X-squared = 12.163, df = 18, p-value = 0.8387
```

Desempeño en comunicación escrita vs valor anual de la matrícula:

```

Frequency table:
                Desempe.comuni.escrit
Valor.anual.matri  Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Entre 3 millones y 5 millones  0     0     0     0     1     0     0
Entre 500 mil y un millon     1     0    10    39    31    11     1
Entre un millon y 3 millones  0     0     0    10     3     4     2
Menos de 500 mil             4     3    12    78    44    23     4
No pago                       0     0     1     0     1     1     0

    Pearson's Chi-squared test

data: .Table
X-squared = 22.208, df = 24, p-value = 0.5669
    
```

Desempeño en comunicación escrita vs ocupación del padre:

```

Frequency table:
                Desempe.comuni.escrit
ocupa.padre      Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Administrador     3     0     1    16     9     6     1
Obrero            1     0     5    31    25    10     1
Ocupacion No remunerada  0     0     0     1     1     0     0
Otra actividad u ocupacion  1     2    14    63    35    22     5
Pensionado        0     1     3    16    10     1     0

    Pearson's Chi-squared test

data: .Table
X-squared = 23.099, df = 24, p-value = 0.5139
    
```

Desempeño en comunicación escrita vs ocupación de la madre:

```

Frequency table:
                Desempe.comuni.escrit
ocupa.madre      Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Administrador     1     0     0     8     4     5     0
Obrero            0     0     5    17    11     9     2
Ocupacion No remunerada  3     1    11    74    45    19     4
Otra actividad u ocupacion  1     2     7    25    16     6     1
Pensionado        0     0     0     3     4     0     0

    Pearson's Chi-squared test

data: .Table
X-squared = 20.891, df = 24, p-value = 0.6452
    
```

Desempeño en comunicación escrita vs nivel educativo del padre:

```

Frequency table:
      Desempe.comuni.escrit
Nivel.edu.padre  Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Primaria         4       0       7      53      32       9       2
Secundaria       1       2       9      25      18      12       1
Sin estudio      0       1       5      29      18       6       2
tecnicos         0       0       1       8       5       7       0
Universitarios  0       0       1      12       7       5       2

      Pearson's Chi-squared test

data: .Table
X-squared = 27.912, df = 24, p-value = 0.2638

```

Desempeño en comunicación escrita vs nivel educativo de la madre:

```

Frequency table:
      Desempe.comuni.escrit
Nivel.edu.madre  Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Primaria         4       1       8      57      29      17       1
Secundaria       1       0       9      29      25       8       2
Sin estudio      0       1       2      22      15       4       2
tecnicos         0       1       2       6       4       4       0
Universitarios  0       0       2      13       7       6       2

      Pearson's Chi-squared test

data: .Table
X-squared = 23.133, df = 24, p-value = 0.512

```

Desempeño en comunicación escrita vs tiene celular:

```

Frequency table:
      Desempe.comuni.escrit
Celular Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
No       0       0       2       7       4       2       0
Si       5       3      21      120      76      37       7

      Pearson's Chi-squared test

data: .Table
X-squared = 1.4002, df = 6, p-value = 0.9658

```

Desempeño en comunicación escrita vs Tiene servicio de internet:

```
Frequency table:
Desempe.comuni.escrit
Internet Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
No      4      1     17     82     48     20     3
Si      1      2      6     45     32     19     4

Pearson's Chi-squared test

data: .Table
X-squared = 6.5584, df = 6, p-value = 0.3636
```

Desempeño en comunicación escrita vs tiene servicio de televisión:

```
Frequency table:
Desempe.comuni.escrit
Tv Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
No      3      1     15     69     49     20     2
Si      2      2      8     58     31     19     5

Pearson's Chi-squared test

data: .Table
X-squared = 4.9788, df = 6, p-value = 0.5465
```

Desempeño en comunicación escrita vs tiene lavadora:

```
Frequency table:
Desempe.comuni.escrit
Lavadora Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
No      2      1      8     42     32     15     2
Si      3      2     15     85     48     24     5

Pearson's Chi-squared test

data: .Table
X-squared = 1.358, df = 6, p-value = 0.9684
```

Desempeño en comunicación escrita vs tiene teléfono fijo:

```

Frequency table:
      Desempe.comuni.escrit
Telefono.fijo Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      No          3      3      19      82      57      26      3
      Si           2      0       4      45      23      13      4

      Pearson's Chi-squared test

data: .Table
X-squared = 6.9512, df = 6, p-value = 0.3254

```

Desempeño en comunicación escrita vs Tipo de bachillerato:

```

Frequency table:
      Desempe.comuni.escrit
Tipo.de.Bachillerato Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      Academico         4       2      12      53      32      17      4
      Normalista         0       0       0       2       0       0       0
      Tecnico            1       1      11      72      48      22      3

      Pearson's Chi-squared test

data: .Table
X-squared = 7.546, df = 12, p-value = 0.8195

```

Desempeño en comunicación escrita vs trabaja:

```

Frequency table:
      Desempe.comuni.escrit
Trabaja      Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      No          1       2      11      63      41      24      4
      si, como ayudante sin remuneracion  4       1       7      51      34      12      2
      si, por experiencia o remuneracion  0       0       5      12       5       3       1
      si, por ser practica en plan de estudios  0       0       0       1       0       0       0

      Pearson's Chi-squared test

data: .Table
X-squared = 12.817, df = 18, p-value = 0.8023

```


Desempeño en comunicación escrita vs horas de trabajo a la semana:

```

Frequency table:
                Desempe.comuni.escrit
Horas.trabajo.semanal Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
0 horas                1      2     11     64     41     24     4
Entre 1 y 5 horas      0      0      1      1      2      1      2
Entre 11 y 15 horas    1      0      2      7      9      3      0
Entre 16 y 20 horas    1      1      9     39     21     5      1
Entre 6 y 10 horas     2      0      0     16      7      6      0

                Pearson's Chi-squared test

data: .Table
X-squared = 41.39, df = 24, p-value = 0.01509

```

Desempeño en comunicación escrita vs salario mensual familiar:

```

Frequency table:
                Desempe.comuni.escrit
salario.mensual.fami Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
Entre 1 y menos de 2 SM  1      3     10     65     49     19     2
Entre 2 y menos de 3 SM  3      0      3     24     10     11     2
Entre 3 y menos de 5 SM  1      0      1      7      3      2      1
Entre 5 y menos de 7 SM  0      0      1      2      3      1      0
Mas de 10 SM            0      0      0      1      0      0      0
Menos de 1 SM           0      0      8     28     15     6      2

                Pearson's Chi-squared test

data: .Table
X-squared = 25.579, df = 30, p-value = 0.6964

```

Desempeño en comunicación escrita vs tiene computador:

```

Frequency table:
                Desempe.comuni.escrit
Computador Nivel 1 Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
No          1      0      6     20     16     3      1
Si          4      3     17    107    64     36     6

                Pearson's Chi-squared test

data: .Table
X-squared = 5.143, df = 6, p-value = 0.5256

```

Desempeño en inglés vs género:

```
Frequency table:
              desempe.ingles
Genero       A-  A1 A2 B+ B1
Femenino    3  5 12  2  8
Masculino   29 77 72 21 58

          Pearson's Chi-squared test

data:  .Table
X-squared = 3.2919, df = 4, p-value = 0.5102
```

Desempeño en inglés vs estado civil:

```
Frequency table:
              desempe.ingles
Estado.civil A-  A1 A2 B+ B1
Casado       0  0  0  0  2
Separado     1  0  1  0  0
Soltero      30 77 80 23 63
Union libre  1  5  3  0  1

          Pearson's Chi-squared test

data:  .Table
X-squared = 14.124, df = 12, p-value = 0.2929
```

Desempeño en inglés vs situación del hogar:

```
Frequency table:
              desempe.ingles
Situacion.hogar A-  A1 A2 B+ B1
Permanente     21 58 62 18 44
Temporal       11 24 22  5 22

          Pearson's Chi-squared test

data:  .Table
X-squared = 1.9441, df = 4, p-value = 0.746
```

Desempeño en inglés vs estrato:

```
Frequency table:
      desempe.ingles
Estrato  A- A1 A2 B+ B1
Estrato 1  1  7  7  3 10
Estrato 2 26 59 54 13 36
Estrato 3  5 14 21  6 20
Estrato 4  0  2  2  1  0

      Pearson's Chi-squared test

data: .Table
X-squared = 14.02, df = 12, p-value = 0.2994
```

Desempeño en inglés vs valor anual de la matrícula:

```
Frequency table:
      desempe.ingles
Valor.anual.matri  A- A1 A2 B+ B1
Entre 3 millones y 5 millones  0  0  0  0  1
Entre 500 mil y un millon      8 24 32  6 24
Entre un millon y 3 millones    0  5  6  2  6
Menos de 500 mil               24 53 45 15 33
No pago                         0  0  1  0  2

      Pearson's Chi-squared test

data: .Table
X-squared = 15.719, df = 16, p-value = 0.4727
```

Desempeño en inglés vs ocupación del padre:

```
Frequency table:
      desempe.ingles
ocupa.padre  A- A1 A2 B+ B1
Administrador  6 10 12  3  5
Obrero        7 17 21  8 23
Ocupacion No remunerada  1  0  0  0  1
Otra actividad u ocupacion 15 45 43 10 29
Pensionado    3 10  8  2  8

      Pearson's Chi-squared test

data: .Table
X-squared = 12.538, df = 16, p-value = 0.7062
```

Desempeño en inglés vs ocupación de la madre:

```
Frequency table:
                                     desempe.ingles
ocupa.madre      A-  A1 A2 B+ B1
Administrador      1  3  6  1  7
Obrero            2 11 14  6 11
Ocupacion No remunerada 19 55 41 11 32
Otra actividad u ocupacion 10 12 21  4 13
Pensionado        0  1  2  1  3

      Pearson's Chi-squared test

data: .Table
X-squared = 17.893, df = 16, p-value = 0.3302
```

Desempeño en inglés vs nivel educativo del padre:

```
Frequency table:
                                     desempe.ingles
Nivel.edu.padre  A-  A1 A2 B+ B1
Primaria        15 31 33  8 21
Secundaria       5 25 16  9 13
Sin estudio     10 21 15  4 12
tecnicos        0  3  9  0  9
Universitarios  2  2 11  2 11

      Pearson's Chi-squared test

data: .Table
X-squared = 30.083, df = 16, p-value = 0.01758
```

Desempeño en inglés vs nivel educativo de la madre:

```
Frequency table:
                                     desempe.ingles
Nivel.edu.madre  A-  A1 A2 B+ B1
Primaria        13 28 40 10 27
Secundaria       8 26 18  6 17
Sin estudio      8 19 10  3  7
tecnicos         2  2  9  1  3
Universitarios  1  7  7  3 12

      Pearson's Chi-squared test

data: .Table
X-squared = 21.322, df = 16, p-value = 0.1665
```

Desempeño en inglés vs tiene celular:

```
Frequency table:
      desempe.ingles
Celular A- A1 A2 B+ B1
No  1  5  5  1  3
Si 31 77 79 22 63

      Pearson's Chi-squared test

data:  .Table
X-squared = 0.59792, df = 4, p-value = 0.9633
```

Desempeño en inglés vs tiene internet:

```
Frequency table:
      desempe.ingles
Internet A- A1 A2 B+ B1
No 21 56 45 14 40
Si 11 26 39  9 26

      Pearson's Chi-squared test

data:  .Table
X-squared = 4.0735, df = 4, p-value = 0.3961
```

Desempeño en inglés vs tiene computador:

```
Frequency table:
      desempe.ingles
Computador A- A1 A2 B+ B1
No  7 15 10  3 13
Si 25 67 74 20 53

      Pearson's Chi-squared test

data:  .Table
X-squared = 2.7978, df = 4, p-value = 0.5922
```

Desempeño en inglés vs tiene televisión:

```
Frequency table:
      desempe.ingles
Tv   A- A1 A2 B+ B1
No  19 47 41 10 43
Si   13 35 43 13 23

      Pearson's Chi-squared test

data:  .Table
X-squared = 5.661, df = 4, p-value = 0.2259
```

Desempeño en inglés vs tiene lavadora:

```
Frequency table:
      desempe.ingles
Lavadora A- A1 A2 B+ B1
No  14 29 31  7 21
Si   18 53 53 16 45

      Pearson's Chi-squared test

data:  .Table
X-squared = 1.6716, df = 4, p-value = 0.7959
```

Desempeño en inglés vs tiene teléfono fijo:

```
Frequency table:
      desempe.ingles
Telefono.fijo A- A1 A2 B+ B1
No  22 58 55 15 46
Si   10 24 29  8 20

      Pearson's Chi-squared test

data:  .Table
X-squared = 0.69665, df = 4, p-value = 0.9517
```

Desempeño en inglés vs tipo de bachillerato:

```
Frequency table:
                    desempe.ingles
Tipo.de.Bachillerato A- A1 A2 B+ B1
Academico           15 34 40 10 27
Normalista          0  1  1  0  0
Tecnico             17 47 43 13 39

Pearson's Chi-squared test

data: .Table
X-squared = 2.5543, df = 8, p-value = 0.9592
```

Desempeño en inglés vs trabaja:

```
Frequency table:
                    desempe.ingles
Trabaja              A- A1 A2 B+ B1
No                   13 34 40 17 43
si, como ayudante sin remuneracion 15 39 37  5 16
si, por experiencia o remuneracion   4  9  7  1  6
si, por ser practica en plan de estudios 0  0  0  0  1

Pearson's Chi-squared test

data: .Table
X-squared = 19.912, df = 12, p-value = 0.06878
```

Desempeño en inglés vs horas de trabajo a la semana:

```
Frequency table:
                    desempe.ingles
Horas.trabajo.semanal A- A1 A2 B+ B1
0 horas                13 35 40 17 43
Entre 1 y 5 horas      0  2  4  0  1
Entre 11 y 15 horas    2  8  7  3  3
Entre 16 y 20 horas   10 30 23  1 14
Entre 6 y 10 horas     7  7 10  2  5

Pearson's Chi-squared test

data: .Table
X-squared = 25.423, df = 16, p-value = 0.06271
```

Desempeño en inglés vs Salario mensual familiar:

```
Frequency table:
                desempe.ingles
salario.mensual.fami  A- A1 A2 B+ B1
Entre 1 y menos de 2 SM 17 48 45 10 31
Entre 2 y menos de 3 SM  2 13 21  8  9
Entre 3 y menos de 5 SM  0  3  4  2  6
Entre 5 y menos de 7 SM  2  0  2  0  4
Mas de 10 SM           0  0  0  1  0
Menos de 1 SM         11 18 12  2 16

        Pearson's Chi-squared test

data:  .Table
X-squared = 39.851, df = 20, p-value = 0.005217
```

Para variables de tipo cuantitativo, se realizó la prueba de normalidad de Shapiro Wilk de cada variable respuesta, dando como resultado en este caso, que los puntajes no asumen una distribución normal, para ello se tuvo en cuenta el criterio del p-valor (Rechazar la hipótesis nula (normalidad) con un valor igual o inferior a 0.05).

PRUEBA DE NORMALIDAD PUNTAJE RAZONAMIENTO CUANTITATIVO:

```
        Shapiro-Wilk normality test

data:  puntaje.razo.cuanti
W = 0.81535, p-value < 2.2e-16
```

PRUEBA DE NORMALIDAD PUNTAJE LECTURA CRÍTICA:

```
        Shapiro-Wilk normality test

data:  puntaje.lectu.crit
W = 0.78949, p-value < 2.2e-16
```


PRUEBA DE NORMALIDAD PUNTAJE COMUNICACIÓN ESCRITA:

```
Shapiro-Wilk normality test
data: puntaje.comuni.escri
W = 0.96987, p-value = 0.00001122
```

PRUEBA DE NORMALIDAD PUNTAJE INGLÉS:

```
Shapiro-Wilk normality test
data: puntaje.ingles
W = 0.80087, p-value < 2.2e-16
```

Como ninguno de los puntajes tiene una distribución normal se realizó el test no paramétrico de Kruskal Wallis, donde se buscó determinar asociación entre algunas variables y los puntajes del componente genérico.

Razonamiento cuantitativo vs Tiene celular

```
> with(Dataset, tapply(puntaje.razo.cuanti, Celular, median, na.rm=TRUE))
  No  Si
11.4 11.0

> kruskal.test(puntaje.razo.cuanti ~ Celular, data=Dataset)

Kruskal-Wallis rank sum test

data: puntaje.razo.cuanti by Celular
Kruskal-Wallis chi-squared = 1.5397, df = 1, p-value = 0.2147
```

Razonamiento cuantitativo vs Tiene computador

```
> with(Dataset, tapply(puntaje.razo.cuanti, Computador, median, na.rm=TRUE))
  No  Si
10.9 11.2

> kruskal.test(puntaje.razo.cuanti ~ Computador, data=Dataset)

Kruskal-Wallis rank sum test

data: puntaje.razo.cuanti by Computador
Kruskal-Wallis chi-squared = 1.8544, df = 1, p-value = 0.1733
```

Razonamiento cuantitativo vs Estado civil

```
> with(Dataset, tapply(puntaje.razo.cuanti, Estado.civil, median, na.rm=TRUE))
  Casado   Separado   Soltero Union libre
  10.65    10.85     11.00    11.55

> kruskal.test(puntaje.razo.cuanti ~ Estado.civil, data=Dataset)

    Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Estado.civil
Kruskal-Wallis chi-squared = 2.2686, df = 3, p-value = 0.5186
```

Razonamiento cuantitativo vs Estrato

```
> with(Dataset, tapply(puntaje.razo.cuanti, Estrato, median, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4
  11.00    11.00    11.25    11.10

> kruskal.test(puntaje.razo.cuanti ~ Estrato, data=Dataset)

    Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Estrato
Kruskal-Wallis chi-squared = 2.4304, df = 3, p-value = 0.488
```

Razonamiento cuantitativo vs Sexo

```
> with(Dataset, tapply(puntaje.razo.cuanti, Genero, median, na.rm=TRUE))
Femenino Masculino
  10.7     11.1

> kruskal.test(puntaje.razo.cuanti ~ Genero, data=Dataset)

    Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Genero
Kruskal-Wallis chi-squared = 5.7866, df = 1, p-value = 0.01615
```

Razonamiento cuantitativo vs Número de horas que trabaja semanalmente

```
> with(Dataset, tapply(puntaje.razo.cuanti, Horas.trabajo.semanal, median, na.rm=TRUE))
      0 horas  Entre 1 y 5 horas Entre 11 y 15 horas Entre 16 y 20 horas
      11.10    11.50             10.90             11.25
Entre 6 y 10 horas
      10.60

> kruskal.test(puntaje.razo.cuanti ~ Horas.trabajo.semanal, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Horas.trabajo.semanal
Kruskal-Wallis chi-squared = 11.476, df = 4, p-value = 0.0217
```

Razonamiento cuantitativo vs Tiene servicio de internet

```
> with(Dataset, tapply(puntaje.razo.cuanti, Internet, median, na.rm=TRUE))
      No  Si
11.0 11.2

> kruskal.test(puntaje.razo.cuanti ~ Internet, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Internet
Kruskal-Wallis chi-squared = 0.97466, df = 1, p-value = 0.3235
```

Razonamiento cuantitativo vs Tiene Lavadora

```
> with(Dataset, tapply(puntaje.razo.cuanti, Lavadora, median, na.rm=TRUE))
      No  Si
10.9 11.2

> kruskal.test(puntaje.razo.cuanti ~ Lavadora, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Lavadora
Kruskal-Wallis chi-squared = 4.5094, df = 1, p-value = 0.03371
```

Razonamiento cuantitativo vs Nivel educativo de la madre

```
> with(Dataset, tapply(puntaje.razo.cuanti, Nivel.edu.madre, median, na.rm=TRUE))
      Primaria      Secundaria      Sin estudio      tecnicos Universitarios
      10.95         11.20         10.70         11.00         11.50

> kruskal.test(puntaje.razo.cuanti ~ Nivel.edu.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Nivel.edu.madre
Kruskal-Wallis chi-squared = 10.572, df = 4, p-value = 0.03182
```

Razonamiento cuantitativo vs Nivel educativo del padre

```
> with(Dataset, tapply(puntaje.razo.cuanti, Nivel.edu.padre, median, na.rm=TRUE))
      Primaria      Secundaria      Sin estudio      tecnicos Universitarios
      10.9         11.2         10.9         11.1         11.6

> kruskal.test(puntaje.razo.cuanti ~ Nivel.edu.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Nivel.edu.padre
Kruskal-Wallis chi-squared = 10.174, df = 4, p-value = 0.0376
```

Razonamiento cuantitativo vs Ocupación de la madre

```
> with(Dataset, tapply(puntaje.razo.cuanti, ocupa.madre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.95         11.50         11.00
Otra actividad u ocupacion      Pensionado
      10.95         11.70

> kruskal.test(puntaje.razo.cuanti ~ ocupa.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by ocupa.madre
Kruskal-Wallis chi-squared = 10.859, df = 4, p-value = 0.0282
```

Razonamiento cuantitativo vs Ocupación del padre

```
> with(Dataset, tapply(puntaje.razo.cuanti, ocupa.padre, median, na.rm=TRUE))
      Administrador      Obrero  Ocupacion No remunerada
      11.0            11.2            10.2
Otra actividad u ocupacion  Pensionado
      11.0            11.5

> kruskal.test(puntaje.razo.cuanti ~ ocupa.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by ocupa.padre
Kruskal-Wallis chi-squared = 8.6072, df = 4, p-value = 0.0717
```

Razonamiento cuantitativo vs Salario mensual familiar

```
> with(Dataset, tapply(puntaje.razo.cuanti, salario.mensual.fami, median, na.rm=TRUE))
Entre 1 y menos de 2 SM Entre 2 y menos de 3 SM Entre 3 y menos de 5 SM Entre 5 y menos de 7 SM
      11.00            11.40            11.50            11.25
      Mas de 10 SM      Menos de 1 SM
      14.00            10.90

> kruskal.test(puntaje.razo.cuanti ~ salario.mensual.fami, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by salario.mensual.fami
Kruskal-Wallis chi-squared = 8.7039, df = 5, p-value = 0.1215
```

Razonamiento cuantitativo vs Situación del hogar

```
> with(Dataset, tapply(puntaje.razo.cuanti, Situacion.hogar, median, na.rm=TRUE))
Permanente  Temporal
      11.0    11.1

> kruskal.test(puntaje.razo.cuanti ~ Situacion.hogar, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Situacion.hogar
Kruskal-Wallis chi-squared = 2.6088, df = 1, p-value = 0.1063
```

Razonamiento cuantitativo vs Tiene teléfono fijo

```
> with(Dataset, tapply(puntaje.razo.cuanti, Telefono.fijo, median, na.rm=TRUE))
  No  Si
11.0 11.2

> kruskal.test(puntaje.razo.cuanti ~ Telefono.fijo, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Telefono.fijo
Kruskal-Wallis chi-squared = 1.8324, df = 1, p-value = 0.1758
```

Razonamiento cuantitativo vs Tipo de bachillerato

```
> with(Dataset, tapply(puntaje.razo.cuanti, Tipo.de.Bachillerato, median, na.rm=TRUE))
Academico Normalista   Tecnico
      11.0           10.8           11.2

> kruskal.test(puntaje.razo.cuanti ~ Tipo.de.Bachillerato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Tipo.de.Bachillerato
Kruskal-Wallis chi-squared = 0.39566, df = 2, p-value = 0.8205
```

Razonamiento cuantitativo vs Trabaja

```
> with(Dataset, tapply(puntaje.razo.cuanti, Trabaja, median, na.rm=TRUE))
      No          si, como ayudante sin remuneracion
      11.1          11.0
si, por experiencia o remuneracion si, por ser practica en plan de estudios
      10.9          10.2

> kruskal.test(puntaje.razo.cuanti ~ Trabaja, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Trabaja
Kruskal-Wallis chi-squared = 2.6621, df = 3, p-value = 0.4467
```

Razonamiento cuantitativo vs Tiene servicio de televisión

```
> with(Dataset, tapply(puntaje.razo.cuanti, Tv, median, na.rm=TRUE))
  No  Si
11.0 11.1

> kruskal.test(puntaje.razo.cuanti ~ Tv, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Tv
Kruskal-Wallis chi-squared = 0.068772, df = 1, p-value = 0.7931
```

Razonamiento cuantitativo vs valor anual de la matrícula

```
> with(Dataset, tapply(puntaje.razo.cuanti, Valor.anual.matri, median, na.rm=TRUE))
Entre 3 millones y 5 millones  Entre 500 mil y un millon  Entre un millon y 3 millones
                        10.20                        11.15                        11.70
      Menos de 500 mil                No pago
                        11.00                11.80

> kruskal.test(puntaje.razo.cuanti ~ Valor.anual.matri, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.razo.cuanti by Valor.anual.matri
Kruskal-Wallis chi-squared = 13.923, df = 4, p-value = 0.007544
```

XX

Puntaje lectura crítica vs Tiene celular

```
> with(Dataset, tapply(puntaje.lectu.crit, Celular, median, na.rm=TRUE))
  No  Si
10.8 10.5

> kruskal.test(puntaje.lectu.crit ~ Celular, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Celular
Kruskal-Wallis chi-squared = 0.9362, df = 1, p-value = 0.3333
```

Puntaje lectura crítica vs Tiene computador

```
> with(Dataset, tapply(puntaje.lectu.crit, Computador, median, na.rm=TRUE))
  No    Si
10.55 10.60

> kruskal.test(puntaje.lectu.crit ~ Computador, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Computador
Kruskal-Wallis chi-squared = 0.28833, df = 1, p-value = 0.5913
```

Puntaje lectura crítica vs Estado civil

```
> with(Dataset, tapply(puntaje.lectu.crit, Estado.civil, median, na.rm=TRUE))
  Casado   Separado   Soltero Union libre
  10.50    10.05     10.60    10.40

> kruskal.test(puntaje.lectu.crit ~ Estado.civil, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Estado.civil
Kruskal-Wallis chi-squared = 1.207, df = 3, p-value = 0.7513
```

Puntaje lectura crítica vs Estrato

```
> with(Dataset, tapply(puntaje.lectu.crit, Estrato, median, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4
  10.8    10.6    10.5    10.4

> kruskal.test(puntaje.lectu.crit ~ Estrato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Estrato
Kruskal-Wallis chi-squared = 2.2953, df = 3, p-value = 0.5134
```


Puntaje lectura crítica vs Tiene Sexo

```
> with(Dataset, tapply(puntaje.lectu.crit, Genero, median, na.rm=TRUE))
Femenino Masculino
 10.8      10.5

> kruskal.test(puntaje.lectu.crit ~ Genero, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Genero
Kruskal-Wallis chi-squared = 1.0505, df = 1, p-value = 0.3054
```

Puntaje lectura crítica vs Número de horas que trabaja a la semana

```
> with(Dataset, tapply(puntaje.lectu.crit, Horas.trabajo.semanal, median, na.rm=TRUE))
 0 horas  Entre 1 y 5 horas Entre 11 y 15 horas Entre 16 y 20 horas
 10.7      10.9              10.3              10.5
Entre 6 y 10 horas
 10.1

> kruskal.test(puntaje.lectu.crit ~ Horas.trabajo.semanal, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Horas.trabajo.semanal
Kruskal-Wallis chi-squared = 11.165, df = 4, p-value = 0.02477
```

Puntaje lectura crítica vs Tiene servicio de internet

```
> with(Dataset, tapply(puntaje.lectu.crit, Internet, median, na.rm=TRUE))
No  Si
10.6 10.5

> kruskal.test(puntaje.lectu.crit ~ Internet, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Internet
Kruskal-Wallis chi-squared = 0.80028, df = 1, p-value = 0.371
```

Puntaje lectura crítica vs Tiene lavadora

```
> with(Dataset, tapply(puntaje.lectu.crit, Lavadora, median, na.rm=TRUE))
  No   Si
10.65 10.50

> kruskal.test(puntaje.lectu.crit ~ Lavadora, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Lavadora
Kruskal-Wallis chi-squared = 0.02279, df = 1, p-value = 0.88
```

Puntaje lectura crítica vs Nivel educativo de la madre

```
> with(Dataset, tapply(puntaje.lectu.crit, Nivel.edu.madre, median, na.rm=TRUE))
  Primaria      Secundaria   Sin estudio   tecnicos Universitarios
      10.5           10.7           10.4           10.6           10.7

> kruskal.test(puntaje.lectu.crit ~ Nivel.edu.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Nivel.edu.madre
Kruskal-Wallis chi-squared = 10.036, df = 4, p-value = 0.03983
```

Puntaje lectura crítica vs Nivel educativo del padre

```
> with(Dataset, tapply(puntaje.lectu.crit, Nivel.edu.padre, median, na.rm=TRUE))
  Primaria      Secundaria   Sin estudio   tecnicos Universitarios
      10.50           10.70           10.50           10.50           10.75

> kruskal.test(puntaje.lectu.crit ~ Nivel.edu.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Nivel.edu.padre
Kruskal-Wallis chi-squared = 8.3546, df = 4, p-value = 0.07942
```

Puntaje lectura crítica vs Tiene Ocupación de la madre

```
> with(Dataset, tapply(puntaje.lectu.crit, ocupa.madre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.35           10.85           10.50
Otra actividad u ocupacion      Pensionado
      10.70           11.20

> kruskal.test(puntaje.lectu.crit ~ ocupa.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by ocupa.madre
Kruskal-Wallis chi-squared = 12.493, df = 4, p-value = 0.01404
```

Puntaje lectura crítica vs Ocupación del padre

```
> with(Dataset, tapply(puntaje.lectu.crit, ocupa.padre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.50           10.70           10.75
Otra actividad u ocupacion      Pensionado
      10.50           10.60

> kruskal.test(puntaje.lectu.crit ~ ocupa.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by ocupa.padre
Kruskal-Wallis chi-squared = 4.0595, df = 4, p-value = 0.398
```

Puntaje lectura crítica vs Salario mensual familiar

```
> with(Dataset, tapply(puntaje.lectu.crit, salario.mensual.fami, median, na.rm=TRUE))
Entre 1 y menos de 2 SM Entre 2 y menos de 3 SM Entre 3 y menos de 5 SM Entre 5 y menos de 7 SM
      10.50           10.60           10.70           10.85
      Mas de 10 SM      Menos de 1 SM
      12.40           10.50

> kruskal.test(puntaje.lectu.crit ~ salario.mensual.fami, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by salario.mensual.fami
Kruskal-Wallis chi-squared = 6.8465, df = 5, p-value = 0.2323
```

Puntaje lectura crítica vs Situación del hogar

```
> with(Dataset, tapply(puntaje.lectu.crit, Situacion.hogar, median, na.rm=TRUE))
Permanente   Temporal
      10.60      10.55

> kruskal.test(puntaje.lectu.crit ~ Situacion.hogar, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Situacion.hogar
Kruskal-Wallis chi-squared = 0.0020587, df = 1, p-value = 0.9638
```

Puntaje lectura crítica vs Tiene teléfono fijo

```
> with(Dataset, tapply(puntaje.lectu.crit, Telefono.fijo, median, na.rm=TRUE))
No  Si
10.5 10.7

> kruskal.test(puntaje.lectu.crit ~ Telefono.fijo, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Telefono.fijo
Kruskal-Wallis chi-squared = 0.81744, df = 1, p-value = 0.3659
```

Puntaje lectura crítica vs Tipo de bachillerato

```
> with(Dataset, tapply(puntaje.lectu.crit, Tipo.de.Bachillerato, median, na.rm=TRUE))
Academico Normalista   Tecnico
      10.5      11.6      10.6

> kruskal.test(puntaje.lectu.crit ~ Tipo.de.Bachillerato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.lectu.crit by Tipo.de.Bachillerato
Kruskal-Wallis chi-squared = 5.3698, df = 2, p-value = 0.06823
```


Puntaje en comunicación escrita vs Tiene celular

```
> with(Dataset, tapply(puntaje.comuni.escri, Celular, median, na.rm=TRUE))
  No  Si
10.0 10.1

> kruskal.test(puntaje.comuni.escri ~ Celular, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Celular
Kruskal-Wallis chi-squared = 0.2213, df = 1, p-value = 0.638
```

Puntaje en comunicación escrita vs Tiene computador

```
> with(Dataset, tapply(puntaje.comuni.escri, Computador, median, na.rm=TRUE))
  No  Si
10.1 10.1

> kruskal.test(puntaje.comuni.escri ~ Computador, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Computador
Kruskal-Wallis chi-squared = 0.60435, df = 1, p-value = 0.4369
```

Puntaje en comunicación escrita vs Estado civil

```
> with(Dataset, tapply(puntaje.comuni.escri, Estado.civil, median, na.rm=TRUE))
  Casado  Separado  Soltero Union libre
  10.30    10.90    10.10    9.95

> kruskal.test(puntaje.comuni.escri ~ Estado.civil, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Estado.civil
Kruskal-Wallis chi-squared = 1.2712, df = 3, p-value = 0.736
```

Puntaje en comunicación escrita vs Estrato

```
> with(Dataset, tapply(puntaje.comuni.escri, Estrato, median, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4
      10.1      10.1      10.0      10.1

> kruskal.test(puntaje.comuni.escri ~ Estrato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Estrato
Kruskal-Wallis chi-squared = 1.0168, df = 3, p-value = 0.7972
```

Puntaje en comunicación escrita vs Sexo

```
> with(Dataset, tapply(puntaje.comuni.escri, Genero, median, na.rm=TRUE))
Femenino Masculino
      10.4      10.0

> kruskal.test(puntaje.comuni.escri ~ Genero, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Genero
Kruskal-Wallis chi-squared = 2.7395, df = 1, p-value = 0.09789
```

Puntaje en comunicación escrita vs Horas de trabajo a la semana

```
> with(Dataset, tapply(puntaje.comuni.escri, Horas.trabajo.semanal, median, na.rm=TRUE))
      0 horas  Entre 1 y 5 horas Entre 11 y 15 horas Entre 16 y 20 horas
      10.2      10.7      10.3      9.9
Entre 6 y 10 horas
      10.0

> kruskal.test(puntaje.comuni.escri ~ Horas.trabajo.semanal, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Horas.trabajo.semanal
Kruskal-Wallis chi-squared = 4.3611, df = 4, p-value = 0.3593
```

Puntaje en comunicación escrita vs Tiene servicio de internet

```
> with(Dataset, tapply(puntaje.comuni.escri, Internet, median, na.rm=TRUE))
  No   Si
10.0 10.2

> kruskal.test(puntaje.comuni.escri ~ Internet, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Internet
Kruskal-Wallis chi-squared = 3.6997, df = 1, p-value = 0.05442
```

Puntaje en comunicación escrita vs Tiene lavadora

```
> with(Dataset, tapply(puntaje.comuni.escri, Lavadora, median, na.rm=TRUE))
  No   Si
10.15 10.05

> kruskal.test(puntaje.comuni.escri ~ Lavadora, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Lavadora
Kruskal-Wallis chi-squared = 0.13142, df = 1, p-value = 0.717
```

Puntaje en comunicación escrita vs Nivel educativo de la madre

```
> with(Dataset, tapply(puntaje.comuni.escri, Nivel.edu.madre, median, na.rm=TRUE))
  Primaria      Secundaria   Sin estudio   tecnicos Universitarios
  10.00         10.10         10.05         10.00         10.20

> kruskal.test(puntaje.comuni.escri ~ Nivel.edu.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Nivel.edu.madre
Kruskal-Wallis chi-squared = 3.262, df = 4, p-value = 0.515
```


Puntaje en comunicación escrita vs Nivel educativo del padre

```
> with(Dataset, tapply(puntaje.comuni.escri, Nivel.edu.padre, median, na.rm=TRUE))
      Primaria      Secundaria      Sin estudio      tecnicos      Universitarios
      10.0         10.1         10.0         10.5         10.2

> kruskal.test(puntaje.comuni.escri ~ Nivel.edu.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Nivel.edu.padre
Kruskal-Wallis chi-squared = 8.5284, df = 4, p-value = 0.07403
```

Puntaje en comunicación escrita vs Ocupación de la madre

```
> with(Dataset, tapply(puntaje.comuni.escri, ocupa.madre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.1         10.2         10.0
Otra actividad u ocupacion      Pensionado
      9.9         10.5

> kruskal.test(puntaje.comuni.escri ~ ocupa.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by ocupa.madre
Kruskal-Wallis chi-squared = 5.3301, df = 4, p-value = 0.2551
```

Puntaje en comunicación escrita vs Ocupación del padre

```
> with(Dataset, tapply(puntaje.comuni.escri, ocupa.padre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.00         10.20         10.05
Otra actividad u ocupacion      Pensionado
      10.05         10.00

> kruskal.test(puntaje.comuni.escri ~ ocupa.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by ocupa.padre
Kruskal-Wallis chi-squared = 4.6004, df = 4, p-value = 0.3308
```

Puntaje en comunicación escrita vs Salario mensual familiar

```
> with(Dataset, tapply(puntaje.comuni.escri, salario.mensual.fami, median, na.rm=TRUE))
Entre 1 y menos de 2 SM Entre 2 y menos de 3 SM Entre 3 y menos de 5 SM Entre 5 y menos de 7 SM
      10.1              10.0              10.1              10.5
Mas de 10 SM      Menos de 1 SM
      9.9              9.9

> kruskal.test(puntaje.comuni.escri ~ salario.mensual.fami, data=Dataset)

Kruskal-Wallis rank sum test

data: puntaje.comuni.escri by salario.mensual.fami
Kruskal-Wallis chi-squared = 1.9992, df = 5, p-value = 0.8493
```

Puntaje en comunicación escrita vs situación del hogar

```
> with(Dataset, tapply(puntaje.comuni.escri, Situacion.hogar, median, na.rm=TRUE))
Permanente  Temporal
      10.1      10.0

> kruskal.test(puntaje.comuni.escri ~ Situacion.hogar, data=Dataset)

Kruskal-Wallis rank sum test

data: puntaje.comuni.escri by Situacion.hogar
Kruskal-Wallis chi-squared = 0.25509, df = 1, p-value = 0.6135
```

Puntaje en comunicación escrita vs Tiene teléfono fijo

```
> with(Dataset, tapply(puntaje.comuni.escri, Telefono.fijo, median, na.rm=TRUE))
No  Si
10.1 10.1

> kruskal.test(puntaje.comuni.escri ~ Telefono.fijo, data=Dataset)

Kruskal-Wallis rank sum test

data: puntaje.comuni.escri by Telefono.fijo
Kruskal-Wallis chi-squared = 0.22122, df = 1, p-value = 0.6381
```

Puntaje en comunicación escrita vs Tipo de bachillerato

```
> with(Dataset, tapply(puntaje.comuni.escri, Tipo.de.Bachillerato, median, na.rm=TRUE))
Academico Normalista   Tecnico
          10.0           9.6           10.1

> kruskal.test(puntaje.comuni.escri ~ Tipo.de.Bachillerato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Tipo.de.Bachillerato
Kruskal-Wallis chi-squared = 1.8546, df = 2, p-value = 0.3956
```

Puntaje en comunicación escrita vs Trabaja

```
> with(Dataset, tapply(puntaje.comuni.escri, Trabaja, median, na.rm=TRUE))
          No          si, como ayudante sin remuneracion
          10.2          10.1
si, por experiencia o remuneracion si, por ser practica en plan de estudios
          9.9          9.7

> kruskal.test(puntaje.comuni.escri ~ Trabaja, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Trabaja
Kruskal-Wallis chi-squared = 1.7846, df = 3, p-value = 0.6183
```

Puntaje en comunicación escrita vs Tiene servicio de televisión

```
> with(Dataset, tapply(puntaje.comuni.escri, Tv, median, na.rm=TRUE))
No  Si
10.1 10.0

> kruskal.test(puntaje.comuni.escri ~ Tv, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.comuni.escri by Tv
Kruskal-Wallis chi-squared = 0.20659, df = 1, p-value = 0.6495
```


Puntaje en inglés vs Estado civil

```
> with(Dataset, tapply(puntaje.ingles, Estado.civil, median, na.rm=TRUE))
  Casado   Separado   Soltero Union libre
    11.0     9.7     10.2     10.0

> kruskal.test(puntaje.ingles ~ Estado.civil, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Estado.civil
Kruskal-Wallis chi-squared = 4.2691, df = 3, p-value = 0.2338
```

Puntaje en inglés vs Estrato

```
> with(Dataset, tapply(puntaje.ingles, Estrato, median, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4
   10.40   10.14   10.35   10.30

> kruskal.test(puntaje.ingles ~ Estrato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Estrato
Kruskal-Wallis chi-squared = 8.6459, df = 3, p-value = 0.03439
```

Puntaje en inglés vs Sexo

```
> with(Dataset, tapply(puntaje.ingles, Genero, median, na.rm=TRUE))
Femenino Masculino
   10.3     10.2

> kruskal.test(puntaje.ingles ~ Genero, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Genero
Kruskal-Wallis chi-squared = 0.51848, df = 1, p-value = 0.4715
```

Puntaje en inglés vs Número de horas que trabaja a la semana

```
> with(Dataset, tapply(puntaje.ingles, Horas.trabajo.semanal, median, na.rm=TRUE))
      0 horas  Entre 1 y 5 horas Entre 11 y 15 horas Entre 16 y 20 horas
      10.3      10.2      10.2      10.0
Entre 6 y 10 horas
      10.1

> kruskal.test(puntaje.ingles ~ Horas.trabajo.semanal, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Horas.trabajo.semanal
Kruskal-Wallis chi-squared = 14.881, df = 4, p-value = 0.004955
```

Puntaje en inglés vs Tiene servicio de internet

```
> with(Dataset, tapply(puntaje.ingles, Internet, median, na.rm=TRUE))
      No  Si
10.19 10.30

> kruskal.test(puntaje.ingles ~ Internet, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Internet
Kruskal-Wallis chi-squared = 1.4983, df = 1, p-value = 0.2209
```

Puntaje en inglés vs Tiene lavadora

```
> with(Dataset, tapply(puntaje.ingles, Lavadora, median, na.rm=TRUE))
      No  Si
10.18 10.30

> kruskal.test(puntaje.ingles ~ Lavadora, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Lavadora
Kruskal-Wallis chi-squared = 1.2307, df = 1, p-value = 0.2673
```

Puntaje en inglés vs Nivel educativo de la madre

```
> with(Dataset, tapply(puntaje.ingles, Nivel.edu.madre, median, na.rm=TRUE))
      Primaria      Secundaria      Sin estudio      tecnicos Universitarios
      10.30         10.18         9.86         10.30         10.55

> kruskal.test(puntaje.ingles ~ Nivel.edu.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Nivel.edu.madre
Kruskal-Wallis chi-squared = 11.939, df = 4, p-value = 0.01781
```

Puntaje en inglés vs Nivel educativo del padre

```
> with(Dataset, tapply(puntaje.ingles, Nivel.edu.padre, median, na.rm=TRUE))
      Primaria      Secundaria      Sin estudio      tecnicos Universitarios
      10.18         10.20         10.09         10.50         10.52

> kruskal.test(puntaje.ingles ~ Nivel.edu.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Nivel.edu.padre
Kruskal-Wallis chi-squared = 13.851, df = 4, p-value = 0.007785
```

Puntaje en inglés vs Ocupación de la madre

```
> with(Dataset, tapply(puntaje.ingles, ocupa.madre, median, na.rm=TRUE))
      Administrador      Obrero      Ocupacion No remunerada
      10.50         10.35         10.10
Otra actividad u ocupacion      Pensionado
      10.20         10.64

> kruskal.test(puntaje.ingles ~ ocupa.madre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by ocupa.madre
Kruskal-Wallis chi-squared = 11.375, df = 4, p-value = 0.02266
```

Puntaje en inglés vs Ocupación del padre

```
> with(Dataset, tapply(puntaje.ingles, ocupa.padre, median, na.rm=TRUE))
      Administrador      Obrero  Ocupacion No remunerada
      10.100           10.300           10.385
Otra actividad u ocupacion  Pensionado
      10.200           10.200

> kruskal.test(puntaje.ingles ~ ocupa.padre, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by ocupa.padre
Kruskal-Wallis chi-squared = 4.1659, df = 4, p-value = 0.384
```

Puntaje en inglés vs Salario mensual familiar

```
> with(Dataset, tapply(puntaje.ingles, salario.mensual.fami, median, na.rm=TRUE))
Entre 1 y menos de 2 SM Entre 2 y menos de 3 SM Entre 3 y menos de 5 SM Entre 5 y menos de 7 SM
      10.18           10.20           10.60           10.75
      Mas de 10 SM      Menos de 1 SM
      13.94           10.18

> kruskal.test(puntaje.ingles ~ salario.mensual.fami, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by salario.mensual.fami
Kruskal-Wallis chi-squared = 12.826, df = 5, p-value = 0.02506
```

Puntaje en inglés vs Situación hogar

```
> with(Dataset, tapply(puntaje.ingles, Situacion.hogar, median, na.rm=TRUE))
Permanente  Temporal
      10.2      10.2

> kruskal.test(puntaje.ingles ~ Situacion.hogar, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Situacion.hogar
Kruskal-Wallis chi-squared = 0.15603, df = 1, p-value = 0.6928
```


Puntaje en inglés vs Tiene teléfono fijo

```
> with(Dataset, tapply(puntaje.ingles, Telefono.fijo, median, na.rm=TRUE))
  No  Si
10.2 10.3

> kruskal.test(puntaje.ingles ~ Telefono.fijo, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Telefono.fijo
Kruskal-Wallis chi-squared = 0.25325, df = 1, p-value = 0.6148
```

Puntaje en inglés vs Tipo de bachillerato

```
> with(Dataset, tapply(puntaje.ingles, Tipo.de.Bachillerato, median, na.rm=TRUE))
Academico Normalista   Tecnico
    10.20      10.22      10.20

> kruskal.test(puntaje.ingles ~ Tipo.de.Bachillerato, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Tipo.de.Bachillerato
Kruskal-Wallis chi-squared = 0.00022805, df = 2, p-value = 0.9999
```

Puntaje en inglés vs Tiene trabajo

```
> with(Dataset, tapply(puntaje.ingles, Trabaja, median, na.rm=TRUE))
          No          si, como ayudante sin remuneracion
          10.30          10.08
si, por experiencia o remuneracion si, por ser practica en plan de estudios
          10.20          11.00

> kruskal.test(puntaje.ingles ~ Trabaja, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Trabaja
Kruskal-Wallis chi-squared = 15.626, df = 3, p-value = 0.001353
```

Puntaje en inglés vs Tiene servicio de televisión

```
> with(Dataset, tapply(puntaje.ingles, Tv, median, na.rm=TRUE))
  No  Si
10.2 10.2

> kruskal.test(puntaje.ingles ~ Tv, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Tv
Kruskal-Wallis chi-squared = 0.031324, df = 1, p-value = 0.8595
```

Puntaje en inglés vs valor anual matrícula

```
> with(Dataset, tapply(puntaje.ingles, Valor.anual.matri, median, na.rm=TRUE))
Entre 3 millones y 5 millones      Entre 500 mil y un millon  Entre un millon y 3 millones
                11.92                10.30                10.50
      Menos de 500 mil                No pago
                10.10                11.00

> kruskal.test(puntaje.ingles ~ Valor.anual.matri, data=Dataset)

      Kruskal-Wallis rank sum test

data:  puntaje.ingles by Valor.anual.matri
Kruskal-Wallis chi-squared = 8.6599, df = 4, p-value = 0.07018
```

ANEXO B:
ARTÍCULO

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

CASO: INGENIERÍA ELECTROMECAÁNICA DE LA UNIVERSIDAD PEDAGÓGICA Y
TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA

Trabajo de grado

Yeison Darío Acero Báez *, Carmen Helena Cepeda Araque **

¹Escuela de Matemáticas y Estadística, Seccional Duitama, Universidad Pedagógica y
Tecnológica de
Colombia, Duitama, Colombia

Resumen

Este artículo presenta los factores demográficos, académicos y socioeconómicos que intervienen en los puntajes de la prueba Saber Pro presentado por los estudiantes de Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama. Teniendo como fuente las bases de datos suministradas por el ICFES desde el segundo semestre de 2011 a 2015, la cual cuenta con información del estudiante al momento de inscribirse a la prueba. En el trabajo, se hace un análisis con un perfil descriptivo e inferencial de tipo cuantitativo, además se establecieron las siguientes: Como primer momento se conformó el marco teórico, luego en la segunda etapa se definió el diseño metodológico seguido, el siguiente paso fue la aplicación de instrumentos, tabulación y depuración de la base de datos, para luego en el cuarto paso analizar e interpretar la información y construir los modelos que describen aquellos factores que influyen en el componente genérico de la prueba Saber Pro, donde se pudo establecer que: tener estado civil de casado genera mejores puntajes en los componentes de inglés y desempeño en inglés, en comparación a estudiantes con otro tipo de estado civil, además contar con un hogar temporal también aumenta la posibilidad de tener puntajes altos en comparación a quienes cuentan con un hogar permanente en razonamiento cuantitativo, además obtener un tipo de bachillerato académico disminuye la posibilidad de lograr puntajes altos respecto a un bachiller de tipo técnico en lectura crítica. En cuanto a la conexión a internet, obtiene más puntos en comunicación escrita quien posea este servicio, así también como un estudiante que no trabaje.

Palabras clave: Prueba Saber Pro, Educación Superior, Modelo de regresión logística ordinal, GAMLSS.

Abstract

This article presents the demographic, academic and socioeconomic factors that intervene in the test scores of the Saber Pro presented by the students of electromechanical engineering of the Pedagogical and Technological University of Colombia, faculty Sectional Duitama. Having as a source the databases provided by the ICFES from the second semester of 2011 to 2015, which has information from the student at the time of enrollment to the test. At work, an analysis is done with a

descriptive and inferential profile of quantitative type, and the following were established: As the first moment the theoretical framework was formed, then in the second stage the methodological design was defined followed, the following Step was the application of instruments, tabulation and debugging of the database, then in the fourth step analyze and interpret the information and build the models that describe those factors that influence the generic component of the test Saber Pro, where it could be established that: having marital status married generates better scores on the components of English and performance in English, compared to students with other civil status also have a temporary home also increases the chance of having high scores compared to those who have a permanent home in quantitative reasoning also get a kind of academic baccalaureate decreases the possibility of achieving high scores regarding a degree of technical critical reading. As for the internet, you get more points in written communication who owns this service, as well as a student who does not work.

Key words: Test Saber Pro, higher education, logistic regression model ordinal, GAMLSS.

1. Introducción

El Ministerio de Educación Nacional (MEN) en su proyecto de mejoramiento de la calidad ha implementado en la evaluación de la pruebas Saber Pro un componente genérico que evalúa las competencias de los futuros profesionales de cada centro de educación superior del país, con el fin de establecer un referente de medición externa de la calidad de la educación superior. Es por ello que se hace necesario analizar los factores demográficos, académicos y socioeconómicos que intervienen en los puntajes de los estudiantes de Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama. Por lo tanto, en este trabajo monográfico se buscó analizar dichos resultados, teniendo como fuente las bases de datos suministradas por el ICFES desde el segundo semestre de 2011 a 2015, y establecer algunos aspectos a mantener o mejorar, para potenciar las competencias genéricas en sus estudiantes.

En este orden de ideas, y para contextualizar al lector, se hará una breve descripción en cuanto a la prueba Saber Pro, la cual se divide en dos grandes bloques: el primero es el módulo de competencias genéricas y el segundo corresponde a competencias específicas comunes. “Estas pruebas constituyen un instrumento estandarizado para la evaluación externa de la calidad de la educación superior, y tienen como principales estándares caracterizar el grado de calidad de la educación y la evaluación” (ICFES, 2012), además de medir el desarrollo de las competencias que tienen los futuros profesionales, y que son necesarias para afrontar los desafíos en el mercado laboral, haciendo obligatorio que todos los estudiantes que están próximos a graduarse de los diferentes programas de educación superior presenten la prueba. Según la Organización para la Cooperación y el Desarrollo Económico (OCDE):

Incluir competencias genéricas de acuerdo con el proyecto, Evaluación de los Resultados del Aprendizaje en Educación Superior (AHELLO), implica competencias en pensamiento crítico, razonamiento analítico (la capacidad de generar nuevas ideas y la aplicación práctica de la teoría), resolución de problemas, facilidad de comunicación escrita, capacidad de liderazgo y capacidad de trabajo en equipo. A esta lista se puede añadir la competencia en una segunda lengua, sin embargo, aunque existe un acuerdo general sobre la importancia de estas

habilidades, las competencias genéricas son muy difíciles de medir, ya que existen pocos instrumentos para evaluarlas (Mundial, 2012).

Asimismo la OCDE y el Banco Mundial, en la revisión realizada en 2012 sobre la educación superior, encontró que los estudiantes Colombianos están menos preparados en dichas pruebas en comparación con otros países, debido a que en nuestro país son más jóvenes y sus estándares educativos son bajos, con excepción de estudiantes de altos ingresos económicos.

La prueba Saber Pro tiene un alto impacto en las universidades, generando cambios en los currículos y su forma de evaluar. También en la forma como la universidad prepara a los estudiantes para presentar la prueba, fija parámetros y criterios para organizar el sistema de evaluación. Además constituye un referente de medición para la universidad en un ámbito local y nacional, posicionándola a través de dicha prueba, de modo que conocer aquellos factores que inciden en los resultados contribuye a mejorar la calidad académica y genera procesos de autoevaluación a fin de mejorar en competencias como razonamiento cuantitativo, lectura crítica, inglés y comunicación escrita como lo afirma (Mesa 2009).

Al inicio del proceso de análisis de los datos, suministrados por el ICFES, se utilizó un modelo lineal generalizado (GLM), se obtuvo que la distribución de las variables dependientes no pertenecían a la familia exponencial de densidades. Para dar solución a este problema se emplearon los Modelos Aditivos Generalizados para localización, escala y forma (GAMLSS), los cuales incluyen distribuciones que no se asocian a la familia exponencial de densidades. En cuanto a los desempeños se aplicó regresión logística ordinal.

2. Referente Conceptual

“El modelamiento consiste en la aplicación de una serie de procesos con el objeto de conseguir una explicación apropiada del comportamiento de una variable respuesta (datos) a partir de una o más variables explicativas (modelo), la explicación del modelo en general no puede ser perfecta, existe un error o residual” (Díaz, 2012).

El modelo más sencillo que podemos crear es un Modelo Lineal (LM), este modelo aunque sencillo logra cubrir un amplio aspecto de casos a pesar de sus estrictas restricciones como pueden ser el hecho que la variable respuesta debe adaptarse a una distribución normal o bien que la relación entre la variable, o las variables explicativas, y la variable respuesta deba asumirse lineal (Pérez, 2016).

Podemos expresar un modelo LM de la siguiente forma:

$$Y_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{n_x r i} + e_i$$

donde $e_i \sim N(0, \sigma^2)$ para $i = 1, 2, 3, \dots, n$. para este modelo se asume que los errores e_i , son independientes y siguen una distribución normal de media cero y desviación constante. Podemos escribir el modelo de la siguiente forma:

$$Y_i \sim N(0, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{n_x r i} + e_i$$

Donde X es una matriz np ($p = r+1$) que contiene todas las variables explicativas (además de una columna de 1 si se necesita una constante) y β es un vector desconocido de longitud p que será estimado a partir de las variables explicativas. A pesar de que la distribución es adecuada para numerosos e importantes ejemplos estudiados desde que se empezó a trabajar como concepto de regresión, también se ve que la distribución normal no puede abarcar todos los casos existentes. Los modelos lineales son un tipo de modelos de regresión univariante muestra que, entre otras, se dan dos asociaciones. La variable respuesta se considera que sigue una distribución normal y, además que la relación existente entre las variables explicativas y una variable respuesta tenga una forma lineal. Atendiendo la necesidad de que la variable respuesta de un modelo de regresión no siga una distribución normal aparecen los GLM, que son introducidos a continuación (Pérez, 2016).

A principios de la década de los 70, Nelder y Wedderburn (1972) citados por Pérez (2016) propusieron los modelos lineales generalizados (GLM), que engloban varios modelos de regresión ya existentes, proporcionando un marco unificador para aquellos modelos en los que la distribución de la variable respuesta pertenece a la familia exponencial como pueden ser por ejemplo, los modelos de regresión lineal o uno de regresión logística.

Para flexibilizar ligeramente un modelo LM aparece un Modelo Lineal Generalizado (GLM). Este modelo sigue una relación lineal entre la variable respuesta y las covariables del modelo. No obstante, permite que la distribución asumida por la variable respuesta no sea necesariamente una distribución normal, pudiendo ser esta una Poisson o una Binomial (Pérez, 2016), entre otras.

Díaz (2012) afirma:

“Los Modelos lineales Generalizados (GLM) abarcan modelos de regresión ordinal con distribución no normal. Cuando se estudian varias mediciones simultáneamente, resulta útil conseguir una descripción parsimoniosa de los datos a través de un modelo matemático que explique, de alguna forma, las observaciones; a esto se le denomina Modelo óptimo, explicando la mayor variabilidad con el mínimo número de parámetros (principio de parsimonia)”.

Las etapas del modelamiento estadístico se pueden resumir en:

Especificación: Se determinan los supuestos del componente aleatorio, se establece la función del componente sistemático y se determina como los dos componentes son combinación en el modelo mediante la función enlace.

Selección: Se estiman los parámetros del componente sistemático y se valora la precisión de las estimaciones, calculando la discrepancia entre pares de modelos, lo anterior con el objetivo de seleccionar el modelo óptimo.

Evaluación: Donde se analiza si el modelo representa la realidad más aproximada posible. En ello se analiza a través de una prueba, si dicho modelo cumple con las especificaciones.

Una variable respuesta $\{Y_i\}$ con alguna distribución de probabilidad (componente aleatorio) $i = 1, 2, \dots, n$, la cual pertenezca a la familia exponencial de densidades.

Un conjunto de variables explicativas $\{x_i\}$ (componente sistemático) y un vector de parámetros β .

Una función de enlace $g(\cdot)$ entre el componente aleatorio y el componente sistemático la cual describe como se relaciona $x'\beta$ con el valor esperado de $\{Y_i\}$. Implica que $f(\beta_0, \dots, \beta_p, x_1, \dots, x_p)$ es una combinación lineal de parámetros desconocidos es decir:

$$g(\mu_i) = x' \beta = \beta_0 + \beta_1 x_i + \dots + \beta_p x_p = \sum_{j=1}^p \beta_j g_i(x_j)$$

Se puede indicar que las mejoras más importantes introducidas por los modelos GLM fueron los siguientes:

El supuesto de que la variable respuesta sigue una distribución normal fue sustituido por el de que la variable respuesta sigue una distribución de la familia exponencial, pudiendo incluir las distribuciones como la Gamma o una Poisson.

Se comenzó a utilizar una función de enlace monótona, denotada por $g(\cdot)$, para modelar la relación existente entre las variables explicativas y la media de la distribución escogida para la variable respuesta.

Atendiendo a estas consideraciones podemos escribir, utilizando una notación matricial, el modelo GLM de la siguiente forma:

$$Y \sim \text{ExpF}(\mu, \Phi)$$

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

donde la denominación *ExpF* se refiere a la familia exponencial.

Para cada observación, la función de densidad de probabilidad de la familia exponencial puede escribirse como sigue:

$$f_y(y; \mu, \sigma) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}$$

donde $E(Y) = \mu = b'(\theta)$ y $\text{Var}(Y) = \phi v$ siendo $v(\mu) = b''[\theta(\mu)]$.

Con la introducción de los modelos GLM se consigue suavizar considerablemente la restricción de que la distribución de la variable respuesta tuviese que ser necesariamente normal. No obstante, o de hecho de que la relación entre la variable respuesta y las variables explicativas no fuese lineal seguiría pendiente de modelación. Trabajando con modelos GLM a los que se le introdujeron, por ejemplo, polinomios. Por lo tanto, la relación de linealidad fue la principal motivación para la aparición de los modelos aditivos generalizados (GAM) que se exponen en el siguiente apartado (Pérez, 2016).

Los modelos GAM fueron populares en la década de los 80. Los primeros en introducir estos modelos fueron Hastie e Tibshirani (1990) y, posteriormente, fueron extendidos por Wood en el año 2006.

Dichos modelos se pueden considerar una extensión no paramétrica de los GLM; la idea de los modelos GAM es permitir que sean los propios datos los que determinen la relación entre un

predictor lineal η y a las variables explicativas. Un modelo GAM se puede escribir de la siguiente forma:

$$Y \sim \text{ExpF}(\mu, \Phi)$$

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j$$

Donde h son funciones de suavizado no paramétricas que se aplican sobre aquellas variables explicativas continuas que no presentan una relación lineal con respecto a la variable respuesta.

Como lo explica Pérez, 2016:

Un siguiente paso, antes de la creación de los Modelos Aditivos Generalizados (GAM), fue intentar construir relaciones no lineales entre la variable respuesta y las covariables del modelo utilizado, por ejemplo los polinomios. Aun así, las suavizaciones de hecho puntualmente sobre los GLM no se aproximaban suficiente y dieron lugar a la de los modelos GAM. Estos últimos permiten la utilización de funciones suavizadoras para describir la relación entre la variable respuesta y las variables explicativas en un sentido no paramétrico. No obstante, y a pesar de las modificaciones de hecho sobre los GAM, todos estos modelos quedan cortos para aquellas distribuciones que tienen más de un parámetro (que son la gran mayoría) porque todos los modelos mencionados hasta ahora solo son capaces de modelar un parámetro de localización directamente a partir de las covariables. En algunas ocasiones también se logra modelar un parámetro de escala, pero siempre a través de una relación con el parámetro de localización; por esta razón aparecieron los Modelos Aditivos de Localización, Escala y Forma (GAMLSS). Dichos modelos permiten la modelación del parámetro de escala, e incluso de los parámetros de forma (asimetría y curtosis) de la distribución utilizada directamente a partir de las variables explicativas.

Con la introducción de los GAM quedan flexibilizadas las dos condiciones más restrictivas de los modelos LM que eran la imposibilidad de considerar una familia distinta a la normal y de considerar una relación no lineal entre la variable respuesta y las distintas variables explicativas. Aun así, todo es mejorable y con la idea de introducir nuevas distribuciones adicionales a las de la familia exponencial y de poder modelar un parámetro de escala y los parámetros de forma (dependiendo de la distribución) directamente a partir de las variables explicativas, sin necesidad de utilizar la relación existente entre un parámetro de localización y los restantes parámetros, apareciendo los modelos GAMLSS (Pérez, 2016).

Así mismo, Pérez (2016) indica que:

Estos modelos hacen referencia a un grupo de modelos estadísticos realizados por Rigby y Stasinopoulos (2002), permitiendo modelar la media o mediana de las variables dependientes y además, la variabilidad y la asimetría, en relación con variables independientes. Los modelos aditivos generalizados para la localización, escala y forma son unos modelos de regresión univariante que se definen como objetivo de superar varias limitaciones de los modelos ya existentes en este campo, como pueden ser los Modelos Lineales Generalizados (GLM) o los Modelos Aditivos Generalizados (GAM). Los GAMLSS permiten llevar a cabo diversas mejoras, entre las más importantes están poder abandonar la familia exponencial para escoger la

distribución de los datos de entre una amplia gama de opciones, incluida algunas que presentan una fuerte asimetría y/o curtosis. Otra de las ventajas desde modelos es la posibilidad de modelar todos los parámetros de distribución directamente a partir de variables explicativas.

“Un modelo lineal generalizado (GLM) y un modelo aditivo generalizado (GAM) ocupan un lugar importante en el campo de las técnicas estadísticas para la regresión univariante, Nelder e Wedderburn (1972) y Hastie y Tibshirani (1990)”, Pérez, (2016).

Estos dos modelos consideran, para la variable respuesta, una distribución de la familia exponencial. La media es modelada a partir de las variables explicativas mientras que la varianza está dada por la siguiente expresión $\text{Var}(Y) = \varphi v(\mu)$, donde φ es un parámetro de dispersión constante y, por lo tanto, depende de la media. Además si se considera una distribución de familia exponencial, tanto la asimetría como la curtosis están en función de la media y del parámetro φ . Por lo tanto, si se trabaja con modelos GLM o GAM ni la varianza, ni la asimetría, ni la curtosis es modelada explícitamente a partir de las variables explicativas sino a través de la relación de las variables independientes con la media y, la de esta última, con el resto de los parámetros (Pérez, 2016).

Los GAMLSS pueden considerarse modelos de regresión semi paramétricos. Son paramétricos en sentido no que necesitan una distribución paramétrica para la variable respuesta y semi en sentido no que lo modelado de los parámetros de la distribución en función de las variables explicativas puede implicar el uso de funciones suavizadoras no paramétricas.

También Pérez (2016), hace referencia a que:

En los modelos GAMLSS la suposición de que la variables respuesta “Y” pertenece a la familia exponencial y flexibilizada, permitiendo así trabajar con distribuciones con mayor asimetría o curtosis. Por lo tanto, la distribución de la variable respuesta “Y” puede ser seleccionada de entre una amplia gama de distribuciones, incluida aquellas con una alta asimetría o curtosis, tanto continuas como discretas. Además, la parte sistemática del modelo se amplía permitiendo modelar no solo la media (parámetro de localización) sino también la varianza (parámetro de escala), la asimetría o la curtosis (parámetro de forma). Incluye relaciones lineales como no lineales entre los parámetros de las distintas variables explicativas.

Recordando la forma de los GAM, un modelo GAMLSS puede escribirse de la siguiente forma:

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j$$

Y tenemos en cuenta que el modelo GAMLSS modela no solo la media sino todos los parámetros de la distribución a partir de las variables explicativas, siguiendo el modelo:

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

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

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

$$g_4(\mu) = X_4\beta_4 + \sum_{j=1}^{J_4} h_{j4}X_{j4}$$

Ya que en la mayoría de los casos prácticos se tienen como mucho 4 parámetros, siendo estos la media, varianza, asimetría y curtosis, estos 4 parámetros se denotarán por μ , σ , ν , τ respectivamente. Si se quiere generalizar un poco la notación utilizada podemos escribir:

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

Donde θ_k es el vector de parámetros de la distribución.

El modelo GAMLSS queda descrito de la siguiente forma:

$$y \sim D(\theta_k)$$

$$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\gamma_{jk}$$

donde D es la distribución de la variable respuesta y Z_{jk} y γ_{jk} para $k = 1, 2, 3, 4$ y $j = 1, \dots, J_k$ son matrices de diseño para los términos lineales y suavizados respectivamente. β_{jk} es un vector de parámetros de longitud J_k y γ_{jk} es una variable aleatoria q_{jk} dimensional. En GAMLSS se lleva a cabo maximizando la siguiente función de verosimilitud:

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

Donde $\sum_{i=1}^n \log f(y_i | \theta_i)$ es un algoritmo de verosimilitud, λ_{jk} son los parámetros y G_{jk} es una matriz simétrica que depende de los parámetros λ_{jk} . Por lo tanto, debemos estimar β , λ y γ . Siendo $M = D, G, T, \Lambda$ un modelo GAMLSS donde:

D = especifica la distribución de la variable respuesta

G = especifica el conjunto de funciones enlace (g_1, g_2, \dots, g_p) para los parámetros (1, ..., p)

T = engloba a las variables utilizadas en los predictores correspondientes a μ , σ , ν y τ

Λ = engloba los distintos parámetros de suavizado para las funciones de suavizado para un conjunto de datos en concreto, el proceso de selección consiste en comparar muchos modelos construidos combinando los distintos componentes de M. hablamos ahora de cada una de estas componentes por separado.

Componente D: La selección de una distribución apropiada cuenta con 2 pasos. El primer paso será ajustar diferentes modelos utilizando diferentes distribuciones y quedándonos con aquella que proporcione un modelo con menor AIC. El siguiente paso será validar de la escogida a través de un proceso de diagnóstico como puede ser la gráfica de gusano.

Componente G: La selección de la función de enlace viene determinada generalmente por el rango 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 aparece usando el criterio deviance (nos quedaremos con la 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 necesitaremos de su correspondiente parámetro de suavizado. Este puede ser previamente fijado o estimado a partir de los datos. Para este trabajo no se llevarán a cabo suavizadores.

Una vez tenido en cuenta todos los componentes del modelo M y construidos varios modelos se debe tratar de escoger uno entre todos los propuestos. A la hora de escoger el modelo más acertado podemos utilizar el criterio AIC.

“Una vez ajustado el modelo GAMLSS, se utilizarán los residuales cuantiles aleatorizados descritos en Dunn y Smyth (1996)”, para comprobar la adecuación del modelo y, más específicamente, la distribución elegida para la variable “Y” respuesta”, (Pérez, 2016).

Los residuos cuantiles aleatorizados vienen dados por la expresión $r_i = -1(u_i)$ donde -1 es la función inversa de la distribución acumulada de la variable normal estándar y u_i se define como:

* $F(y_i | \hat{\theta}_i)$ Si y_i es continua.

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

La principal ventaja de estos nuevos residuos es que sea cual sea la distribución de la variable respuesta, estos siempre tienden a una distribución normal estándar, siempre y cuando el modelo sea correcto (Pérez, 2016).

La función `plot.gamlss()` representa gráficamente si aquellos residuales aleatorizados presentan un buen ajuste, la función `plot()` tiene como salida 4 gráficos que representan lo siguiente:

- 1: residuos frente a los valores ajustados para a media
- 2: residuos frente a las variables explicativas del modelo.
- 3: estimación Kernel de densidad de los residuos.
- 4: QQ-plot de los residuos.

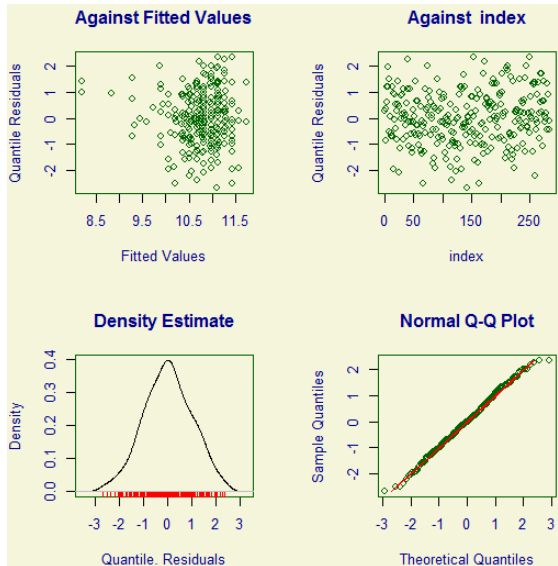


Figura 1: Función `plot.gamlss()`

Otra función que tiene la intención de identificar las relaciones de una variable explicativa donde el modelo no se ajusta adecuadamente a los datos, es la función `wp()` worm plot introducidos por van Buuren e Fredrils (2001) mencionados por Pérez (2016) y tiene como salida un único gráfico el cual representa lo siguiente:

- 1: Los puntos representan los residuos y la línea discontinua horizontal sus valores esperados, por lo tanto en esto se puede observar que tan lejos están los unos de los otros.
- 2: las líneas discontinuas negras marcan el intervalo de confianza del 95%, por lo tanto para que se pueda considerar que el modelo es correcto, sólo un 5% de los puntos podrían quedar fuera de este intervalo. Si esto no es así se debe considerar el modelo inadecuado para explicar la variable respuesta.
- 3: La curva continua horizontal es un ajuste cúbico de los puntos y puede indicar diferentes problemas en el modelo (Stasinopoulos, 2015).

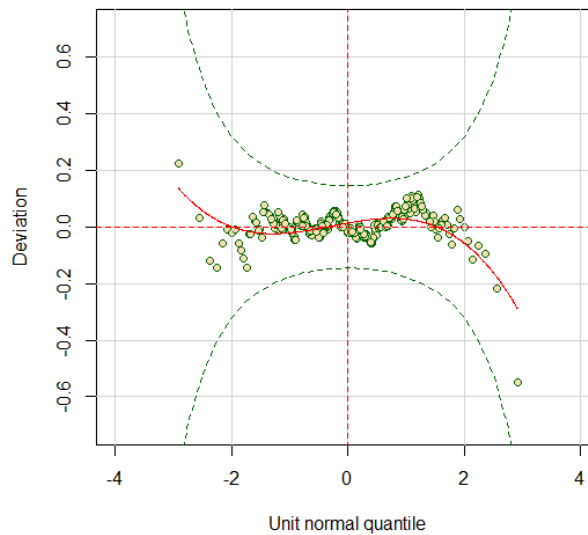


Figura 2: Función $w_p()$

SELECCIÓN DEL MODELO

Una vez ajustado el modelo se debe hacer una buena selección, una de las funciones de ayuda para una buena selección es `stepGAICALL.A()`, la estrategia utilizada por esta función para una distribución dada es:

1. Utilizar el criterio GAIC hacia adelante para seleccionar un modelo apropiado para μ , fijando δ , v y τ .
2. Dado el modelo para μ obtenido en 1 y para v y τ fijas, usa un proceso de selección hacia adelante para escoger un modelo apropiado para δ .
3. Dados los modelos para μ y δ obtenidos en 1 y 2 respectivamente y con τ fija, se utiliza un proceso de selección hacia adelante para escoger un modelo apropiado para v .
4. Dados los modelos para μ , δ y v obtenidos en 1, 2 y 3 respectivamente, se usa un proceso de selección hacia adelante para escoger un modelo apropiado para τ .
5. Dados los modelos para μ , δ y τ obtenidos en 1, 2 y 4 respectivamente, se utiliza un proceso de selección hacia atrás para escoger un modelo apropiado para v .
6. Dados los modelos para μ , v y τ obtenidos en 1, 5 e 4 respectivamente, se usa un proceso de selección hacia atrás para escoger un modelo apropiado para δ .
7. Dados los modelos para δ , v y τ obtenidos en 6, 5 y 4 respectivamente, se utiliza un proceso de selección hacia atrás para escoger un modelo apropiado para μ y finalizar así el proceso.

ALGUNAS DISTRIBUCIONES CONTINUAS PARA LOS GAMLSS

“Una vez realizado el ajuste de los datos a una determinada distribución para todas las variables respuesta, dio como resultado la siguiente tabla, la cual muestra la distribución para cada variable respuesta y sus correspondientes funciones de enlace” (Pérez, 2016):

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	α	ν	τ
Skew t type 3 (ST3)	Razonamiento cuantitativo Lectura crítica	identity	log	log	log
Logistic (LO)	Comunicación escrita	identity	log	---	---
Skew t type 2 (ST2)	Inglés	identity	log	identity	log

Distribución Skew t type 3 (ST3)

La distribución Skew t type 3, denotada por $ST3(\mu, \delta, \nu, \tau)$, definida por:

$$f_Y(y|\mu, \delta, \nu, \tau) = \frac{c}{\delta} \left\{ 1 + \frac{z^2}{\tau} \left[\nu^2 I(y < \mu) + \frac{1}{\nu^2} I(y \geq \mu) \right] \right\} \quad (2-21)$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$ y $\delta > 0$, $\nu > 0$ y $\tau > 0$, y cuando $z = \frac{(y-\mu)}{\delta}$. La media y la varianza están dadas por: $E(Y) = \mu + \delta E(Z)$ y $Var(Y) = \delta^2 V(Z)$. Fernández y Steel (1998), p360, eqn. (5). Con función de enlace $\mu = X\beta - \delta E(Z)$, $\delta^2 = \frac{e^{X\beta}}{V(Z)}$, $\nu = e^{X\beta}$ y $\tau = e^{X\beta}$

Distribución Logistic (LO)

La parametrización de la distribución Logistic, denotada aquí como $LO(\mu, \delta)$, es dada por:

$$f_Y(y|\mu, \delta) = \frac{1}{\delta} \left\{ \exp\left[-\left(\frac{y-\mu}{\delta}\right)\right] \right\} \left\{ 1 + \exp\left[-\left(\frac{y-\mu}{\delta}\right)\right] \right\}^{-2} \quad (2-18)$$

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

Con función de enlace $\mu = X\beta$ y $\delta^2 = \frac{3e^{X\beta}}{\pi^2}$

Distribución Skew t type 2 (ST2)

La distribución Skew t type 2, denotada por $ST2(\mu, \delta, \nu, \tau)$, definida por:

$$f_Y(y|\mu, \delta, \nu, \tau) = \frac{c}{\delta} \left\{ 1 + \frac{z^2}{\tau} [v^2 I(y < \mu) + \frac{1}{v^2} I(y \geq \mu)] \right\}$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$ y $\delta > 0$, $\nu > 0$ y $\tau > 0$, y cuando $z = \frac{(y - \mu)}{\delta}$. La media y la varianza estan dadas por: $E(Y) = \mu + \delta E(Z)$ y $Var(Y) = \delta^2 V(Z)$. Fernandez y Steel (1998), p360, eqn. (5). Con función de enlace $\mu = X\beta - \delta E(Z)$, $\delta^2 = \frac{e^{X\beta}}{V(Z)}$, $\nu = e^{X\beta}$ y $\tau = e^{X\beta}$

Para los desempeños en comunicación escrita e inglés se tuvo en cuenta la regresión logística ordinal que es una técnica estadística multivariante que nos permite estimar la relación existente entre una variable dependiente no Paramétrica, en particular dicotómica y un conjunto de variables independientes Paramétricas o no Paramétricas. Para este caso la variable dependiente indica la ocurrencia o no del suceso. El objetivo de la regresión logística es predecir la probabilidad de un evento de interés en una investigación, así como identificar las variables predictoras útiles para la predicción (Díaz, 2012).

En el análisis de datos es frecuente encontrarse con variables dicotómicas (sí/no, presencia /Ausencia), o variables medidas en escala ordinal. Una práctica usual, es tratar este tipo de variables como si fueran continuas, asignándoles una puntuación arbitraria basada en la codificación de las distintas categorías de respuesta, esta práctica, si bien pudiera considerarse correcta en el caso de variables ordinales, no lo es si las variables son simplemente nominales (Cañadas, 2013).

Los modelos de regresión logística son una herramienta que permite explicar el comportamiento de una variable respuesta discreta (binaria o con más de dos categorías) a través de una o varias variables independientes explicativas de naturaleza cuantitativa y/o cualitativa. Según el tipo de variable respuesta estaremos hablando de regresión logística binaria (variable dependiente con 2 categorías), o de regresión logística multinomial (variable dependiente con más de 2 categorías), pudiendo ser esta ultima de respuesta nominal u ordinal. Los modelos de respuesta discreta son un caso particular de los modelos lineales generalizados formulados por Nelder y Wedderburn en 1972, (Iglesias, 2013).

$$P(Y = 1) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 X_1)}}$$

En el caso particular de la RLO se utiliza una función de enlace para relacionar de forma lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la categoría i de la variable ordinal, y la probabilidad que la variable tome un valor mayor que la categoría i .

En la RLO las funciones de enlace más empleadas son la Logit y la Cloglog. Plantea que el enlace con Logit es 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 (Agresti, 2002).

Para el caso de este estudio, los valores de la variable ordinal representan los puntajes en cuanto al desempeño en inglés y comunicación escrita, cuya complejidad es valorada por los educandos como de Muy bajo a Excelente, por lo tanto no ocurre que los mayores valores (mejores puntajes) sean los más probables, siendo plausible considerar la función de enlace Logit como la más satisfactoria para este caso.

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:

$$O_i = P(\text{valor sea } \leq \text{categoría } i | \text{valores de } X) / (1 - P(\text{valor sea } \leq \text{categoría } i | \text{valores de } X)) \quad *2$$

Con el término valor en la ecuación (*2) se hace referencia a cualquier valor de la variable dependiente. Como se observa, las probabilidades de la ecuación (*2) es condicional, 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 variable independientes, existen p coeficientes, y βX se reemplaza por la combinación lineal entre $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$. Estos coeficientes cuantifican el efecto de las variables independientes sobre el logaritmo de la razón de probabilidad.

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 esta 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 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 tanto existen ecuaciones $k-1$ con los mismos coeficientes acompañando a las variables explicativas y que solo se diferencian en el valor del intercepto (Agresti, 2002).

“Para estimar los coeficientes de la ecuación de regresión se utilizan diversos procedimientos, siendo la estimación de máxima verosimilitud el más empleado” (Agresti, 2002).

Continúa, Agresti (2002) asegurando que:

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:

$$G = -2 \ln (\text{probabilidad de la muestra sin variables}) / (\text{probabilidad de la muestra con variables})$$

Este estadístico sigue distribución ji-cuadrado, con tantos grados de libertad como variables independientes existan, dicho por McCullagh, quien es citado por Agresti, quien recalca que: “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 estas, 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 variable. El estadístico utilizado es el siguiente:

$$Z_{\text{Wald}} = b_j / \text{ES}(b_j) \quad ,$$

Donde b_j es el coeficiente de regresión estimado para la variable independiente j . Bajo la hipótesis de que el coeficiente poblacional b_j para la variable j , la razón entre la estimación de este coeficiente (b_j) y el error estándar de esta estimación [$\text{ES}(b_j)$], sigue una distribución normal estándar. Mayores valores de este estadístico indican que el coeficiente b_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”. También, Agresti plantea que para comparar estas frecuencias generalmente se calcula el estadístico ji cuadrado de Pearson según la fórmula:

$$X^2 = \sum_i \sum_l \frac{(y_{il} - m_l p_{il})^2}{m_l p_{il} (1 - p_{il})} \quad ,$$

En la expresión se considera que la variable dependiente tiene k categorías y que se forman m combinaciones de valores con las variables explicativas, de manera que:

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.

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 X^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 que:

$$P(\text{valor sea} \leq \text{categoría } i | \text{valores de } X) = \ln(O_i) = \alpha_i + \beta X$$

De esta ecuación se deriva que:

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

La expresión es de suma utilidad, pues 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 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 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 $x = x_1$. Numéricamente sería:

$$\text{odds-ratio} = \frac{P(Y \leq ij | X = X_2)}{P(Y \geq ij | X = X_2)} \frac{P(Y \leq ij | X = X_1)}{P(Y \geq ij | X = X_1)}$$

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 cual es utilizado para interpretar el efecto de las variables explicativas sobre la variable objeto de estudio. Si este 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 dependiente” (Agresti, 2002).

3. Metodología

El tipo de investigación aplicada es de carácter descriptivo, inferencial para los resultados de la prueba saber Pro, con enfoque cuantitativo. Se contó con los datos de 287 estudiantes del programa Ingeniería Electromecánica de la UPTC que presentaron las pruebas entre el segundo semestre de 2011 a 2015. Este trabajo monográfico se ejecutó a partir de un diseño bibliográfico; se tomó una base de datos suministradas por el ICFES, a través del taller preparatorio para convocatorias 2016 orientado por el asesor de Gestión de Proyectos de Investigación Luis Eduardo Jaramillo, tomando como referencia la guía de interpretación de resultados suministradas por el ICFES para entender la codificación de las variables, esta base de datos tiene 287 filas por 58 columnas

Demográficas: A esta categoría pertenecen aquellas variables que se refieran a características de la población tales como: sexo, edad, situación del hogar, estado civil, número de personas en el hogar, número de personas a cargo, es cabeza de familia.

Académicas: Pertenecen aquellas variables asociadas al rendimiento académico previo y al esfuerzo mostrado por el estudiante, así como al currículo que sigue el mismo. Así, abarcan aquellos factores educativos que rodean directamente la intervención pedagógica o docente, como: tipo de bachillerato, nivel educativo de la madre y nivel educativo del padre.

Socioeconómicas: A esta categoría pertenecen aquellas variables asociadas al ingreso económico como: Estrato socioeconómico según recibo de energía, valor anual de la matrícula, tener celular, tener internet, tener servicio de tv, tener computador, tener lavadora, tener teléfono fijo en la residencia, salarios mensuales del grupo familiar, número de dormitorios, tener trabajo, número de horas que trabaja en la semana, ocupación del padre y ocupación de la madre.

Para este trabajo se utilizaron técnicas de tipo descriptivo, pruebas de diferencia de medias, pruebas de correlación, tablas de contingencia, pruebas de bondad de ajuste y técnicas de modelamiento como: GAMLSS y modelo multinomial ordinal. Para el procesamiento de los datos se utilizó el software libre R como procesador estadístico versión 3.4.1, como uno de los lenguajes más utilizados en investigación por la comunidad estadística, con sus paquetes: gamlss, MASS, gamlss.dist, gamlss.data, splines, nlme, parallel y Rcmdr.

4. Resultados

A continuación se presenta la descripción de las variables explicativas que se tuvieron en cuenta para este trabajo monográfico y su respectiva clasificación.

FACTORES DE ESTUDIO DEMOGRÁFICOS					
DESCRIPCIÓN		RANGO		RESUMEN	
Sexo del inscrito		Valores Posibles: 2		F=10.45%	M=89.55%
	Femenino	F			
	Masculino	M			
Edad del estudiante		Rango:[20-43]		n=287	
				$\bar{x}=24.54$	
				De=3.12	
				Cv=0.12	
				Cu=8.75	
				Q ₁ =23	
				Q ₂ =24	
				Q ₃ =26	
Estado civil del estudiante		Valores posibles: 5		1=95.12%	
Soltero		1		2=0.70%	
Casado		2		3=0.0%	
Viudo		3		4=0.70%	
Separado		4		5=3.48%	
Unión libre		5			
Situación de su hogar actual		Valores posibles:2		1=70.73%	
Es habitual o permanente		1		2=29.27%	
Es temporal por razones de estudio		2			

De los 287 estudiantes que presentaron la prueba la gran mayoría son hombres, entre ellos predomina ser solteros, con una edad promedio de 25 años y tan solo el 29% residen en un hogar temporal por razones de estudio u otras razones.

Continuación de los factores de estudio demográficos.

FACTORES DE ESTUDIO DEMOGRÁFICOS										
DESCRIPCIÓN	RANGO	RESUMEN								
Número de personas que conforman el hogar	Rango:[1,10]	<table border="1"> <tr> <td>$n=287$</td> </tr> <tr> <td>$\bar{x} = 4.46$</td> </tr> <tr> <td>$De=1.53$</td> </tr> </table>	$n=287$	$\bar{x} = 4.46$	$De=1.53$					
$n=287$										
$\bar{x} = 4.46$										
$De=1.53$										
<table border="1"> <tr> <td>Es cabeza de familia</td> </tr> <tr> <td><i>No</i></td> </tr> <tr> <td><i>Si</i></td> </tr> </table>	Es cabeza de familia	<i>No</i>	<i>Si</i>	<table border="1"> <tr> <td>Valores Posibles:2</td> </tr> <tr> <td><i>0</i></td> </tr> <tr> <td><i>1</i></td> </tr> </table>	Valores Posibles:2	<i>0</i>	<i>1</i>	<table border="1"> <tr> <td>$0=90.24\%$</td> </tr> <tr> <td>$1=9.76\%$</td> </tr> </table>	$0=90.24\%$	$1=9.76\%$
Es cabeza de familia										
<i>No</i>										
<i>Si</i>										
Valores Posibles:2										
<i>0</i>										
<i>1</i>										
$0=90.24\%$										
$1=9.76\%$										
Número de personas que usted tiene a cargo	Valores posibles: [0,4]	<table border="1"> <tr> <td>$n=287$</td> </tr> <tr> <td>$\bar{x} = 0.25$</td> </tr> <tr> <td>$De= 0.70$</td> </tr> </table>	$n=287$	$\bar{x} = 0.25$	$De= 0.70$					
$n=287$										
$\bar{x} = 0.25$										
$De= 0.70$										

En promedio, el número de personas que conforman el núcleo familiar de los estudiantes de Ingeniería Electromecánica es de 4 personas, además solo un pequeño porcentaje de los estudiantes son cabeza de familia (9.76%), por lo tanto se deduce que la mayor parte de ellos no tiene ninguna persona a cargo.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																									
DESCRIPCIÓN	RANGO	RESUMEN																							
<table border="1"> <tr> <td>Estrato residencia según factura de energía</td> </tr> <tr> <td><i>Estrato 1</i></td> </tr> <tr> <td><i>Estrato 2</i></td> </tr> <tr> <td><i>Estrato 3</i></td> </tr> <tr> <td><i>Estrato 4</i></td> </tr> <tr> <td><i>Estrato 5</i></td> </tr> <tr> <td><i>Estrato 6</i></td> </tr> <tr> <td><i>Zona rural sin estratificación</i></td> </tr> </table>	Estrato residencia según factura de energía	<i>Estrato 1</i>	<i>Estrato 2</i>	<i>Estrato 3</i>	<i>Estrato 4</i>	<i>Estrato 5</i>	<i>Estrato 6</i>	<i>Zona rural sin estratificación</i>	<table border="1"> <tr> <td>Rango:[1,7]</td> </tr> <tr> <td>1</td> </tr> <tr> <td>2</td> </tr> <tr> <td>3</td> </tr> <tr> <td>4</td> </tr> <tr> <td>5</td> </tr> <tr> <td>6</td> </tr> <tr> <td>7</td> </tr> </table>	Rango:[1,7]	1	2	3	4	5	6	7	<table border="1"> <tr> <td>1=9.76%</td> </tr> <tr> <td>2=65.51%</td> </tr> <tr> <td>3=23%</td> </tr> <tr> <td>4=1.74%</td> </tr> <tr> <td>5=0.0%</td> </tr> <tr> <td>6=0.0%</td> </tr> <tr> <td>7=0.0%</td> </tr> </table>	1=9.76%	2=65.51%	3=23%	4=1.74%	5=0.0%	6=0.0%	7=0.0%
Estrato residencia según factura de energía																									
<i>Estrato 1</i>																									
<i>Estrato 2</i>																									
<i>Estrato 3</i>																									
<i>Estrato 4</i>																									
<i>Estrato 5</i>																									
<i>Estrato 6</i>																									
<i>Zona rural sin estratificación</i>																									
Rango:[1,7]																									
1																									
2																									
3																									
4																									
5																									
6																									
7																									
1=9.76%																									
2=65.51%																									
3=23%																									
4=1.74%																									
5=0.0%																									
6=0.0%																									
7=0.0%																									
<table border="1"> <tr> <td>Valor anual de la matrícula del año anterior</td> </tr> <tr> <td><i>No pagó matrícula</i></td> </tr> <tr> <td><i>Menos de 500 mil</i></td> </tr> <tr> <td><i>Entre 500 mil y menos de 1 millón</i></td> </tr> <tr> <td><i>Entre 1 millón y 3 millones</i></td> </tr> <tr> <td><i>Entre 3 millones y 5 millones</i></td> </tr> <tr> <td><i>Más de 5 millones</i></td> </tr> </table>	Valor anual de la matrícula del año anterior	<i>No pagó matrícula</i>	<i>Menos de 500 mil</i>	<i>Entre 500 mil y menos de 1 millón</i>	<i>Entre 1 millón y 3 millones</i>	<i>Entre 3 millones y 5 millones</i>	<i>Más de 5 millones</i>	<table border="1"> <tr> <td>Valores posibles: 6</td> </tr> <tr> <td>0</td> </tr> <tr> <td>1</td> </tr> <tr> <td>2</td> </tr> <tr> <td>3</td> </tr> <tr> <td>4</td> </tr> <tr> <td>5</td> </tr> </table>	Valores posibles: 6	0	1	2	3	4	5	<table border="1"> <tr> <td>0=1.05%</td> </tr> <tr> <td>1=59.23%</td> </tr> <tr> <td>2=32.75%</td> </tr> <tr> <td>3=6.62%</td> </tr> <tr> <td>4=0.35%</td> </tr> <tr> <td>5=0%</td> </tr> </table>	0=1.05%	1=59.23%	2=32.75%	3=6.62%	4=0.35%	5=0%			
Valor anual de la matrícula del año anterior																									
<i>No pagó matrícula</i>																									
<i>Menos de 500 mil</i>																									
<i>Entre 500 mil y menos de 1 millón</i>																									
<i>Entre 1 millón y 3 millones</i>																									
<i>Entre 3 millones y 5 millones</i>																									
<i>Más de 5 millones</i>																									
Valores posibles: 6																									
0																									
1																									
2																									
3																									
4																									
5																									
0=1.05%																									
1=59.23%																									
2=32.75%																									
3=6.62%																									
4=0.35%																									
5=0%																									
<table border="1"> <tr> <td>Forma de pago de matrícula</td> </tr> <tr> <td><i>Pago por padres</i></td> </tr> <tr> <td><i>Varias formas de pago</i></td> </tr> <tr> <td><i>No canceló matrícula</i></td> </tr> <tr> <td><i>Pago por beca</i></td> </tr> <tr> <td><i>Pago por crédito</i></td> </tr> <tr> <td><i>Pago propio</i></td> </tr> </table>	Forma de pago de matrícula	<i>Pago por padres</i>	<i>Varias formas de pago</i>	<i>No canceló matrícula</i>	<i>Pago por beca</i>	<i>Pago por crédito</i>	<i>Pago propio</i>	<table border="1"> <tr> <td>Valores posibles: 6</td> </tr> <tr> <td>0</td> </tr> <tr> <td>1</td> </tr> <tr> <td>2</td> </tr> <tr> <td>3</td> </tr> <tr> <td>4</td> </tr> <tr> <td>5</td> </tr> </table>	Valores posibles: 6	0	1	2	3	4	5	<table border="1"> <tr> <td>0=58%</td> </tr> <tr> <td>1=6%</td> </tr> <tr> <td>2=2%</td> </tr> <tr> <td>3=4%</td> </tr> <tr> <td>4=3%</td> </tr> <tr> <td>5=31%</td> </tr> </table>	0=58%	1=6%	2=2%	3=4%	4=3%	5=31%			
Forma de pago de matrícula																									
<i>Pago por padres</i>																									
<i>Varias formas de pago</i>																									
<i>No canceló matrícula</i>																									
<i>Pago por beca</i>																									
<i>Pago por crédito</i>																									
<i>Pago propio</i>																									
Valores posibles: 6																									
0																									
1																									
2																									
3																									
4																									
5																									
0=58%																									
1=6%																									
2=2%																									
3=4%																									
4=3%																									
5=31%																									

Según la factura de energía que llega a la residencia familiar de los estudiantes se clasifica el estrato socioeconómico de los mismos, aquí se observa que la mayoría pertenecen al estrato 2. Respecto al valor anual de la matrícula, los estudiantes (170 personas) pagan menos de 500 mil pesos y la forma de pago es a través de los padres (Es importante mencionar que en la variable forma de pago existen varios datos faltantes: 72%).

Continuación de los factores de estudio socioeconómicos.

FACTORES DE ESTUDIO SOCIOECONÓMICOS			
DESCRIPCIÓN	RANGO	RESUMEN	
Ocupación del padre; madre	Rango:[1,26]		
<i>Empresarios</i>	1	1=0%	1=0%
<i>Administradores o gerentes</i>	2	2=0%	2=0%
<i>Profesionales independientes</i>	3	3=0%	3=0%
<i>Profesionales empleados</i>	4	4=0%	4=0%
<i>Trabajadores independientes</i>	5	5=0%	5=0%
<i>Trabajadores empleados</i>	6	6=0%	6=0%
<i>Rentistas</i>	7	7=0%	7=0%
<i>Obreros</i>	8	8=0%	8=0%
<i>Jubilados</i>	9	9=0%	9=0%
<i>Hogar</i>	10	10=0%	10=0%
<i>Estudiantes</i>	11	11=0%	11=0%
<i>No devengan ingreso o buscan trabajo</i>	12	12=0%	12=0%
<i>Empresario</i>	13	13=0.69%	13=0%
<i>Pequeño empresario</i>	14	14=9.06%	14=4.88%
<i>Empleado con cargo como empleador o gerente</i>	15	15=0.35%	15=0%
<i>Empleado de nivel directivo</i>	16	16=2.44%	16=1.40%
<i>Empleado de nivel técnico o profesional</i>	17	17=9.40%	17=3.83%
<i>Empleado de nivel auxiliar o administrativo</i>	18	18=1.74%	18=3.83%
<i>Empleado obrero u operario</i>	19	19=15.33%	19=7.66%
<i>Profesional independiente</i>	20	20=4.18%	20=3.13%
<i>Trabajador por cuenta propia</i>	21	21=34.14%	21=9.04%
<i>Hogar</i>	22	22=0.69%	22=55.05%
<i>Pensionado</i>	23	23=10.80%	23=2.44%
<i>Rentista</i>	24	24=11.14%	24=0%
<i>Estudiante</i>	25	25=0%	25=0%
<i>Otra actividad u ocupación</i>	26	26=0%	26=5.57%
<i>No sabe</i>	99	99=0%	99=0%

Al analizar la ocupación de los padres de los estudiantes de ingeniería electromecánica que presentaron la prueba saber pro entre el año 2011-2 y el 2015 se tiene que, en su mayoría son trabajadores por cuenta propia, es decir que tienen un ingreso asociado a alguna actividad comercial de compra y/o venta de servicios. En el caso de la ocupación de la madre, la mayoría están vinculadas a la labor doméstica en sus hogares.

Continuación de los factores de estudio socioeconómicos.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																
DESCRIPCIÓN	RANGO	RESUMEN														
<table border="1"> <tr><td>El estudiante tiene celular</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El estudiante tiene celular	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=5.23%</td></tr> <tr><td><i>1</i>=94.77%</td></tr> </table>	<i>0</i> =5.23%	<i>1</i> =94.77%						
El estudiante tiene celular																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =5.23%																
<i>1</i> =94.77%																
<table border="1"> <tr><td>El hogar cuenta con servicio de internet</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio de internet	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=61.32%</td></tr> <tr><td><i>1</i>=38.68%</td></tr> </table>	<i>0</i> =61.32%	<i>1</i> =38.68%						
El hogar cuenta con servicio de internet																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =61.32%																
<i>1</i> =38.68%																
<table border="1"> <tr><td>El hogar cuenta con servicio cerrado de televisión</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio cerrado de televisión	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=55.75%</td></tr> <tr><td><i>1</i>=44.25%</td></tr> </table>	<i>0</i> =55.75%	<i>1</i> =44.25%						
El hogar cuenta con servicio cerrado de televisión																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =55.75%																
<i>1</i> =44.25%																
<table border="1"> <tr><td>El hogar cuenta con servicio de teléfono fijo</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con servicio de teléfono fijo	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=68.29%</td></tr> <tr><td><i>1</i>=31.71%</td></tr> </table>	<i>0</i> =68.29%	<i>1</i> =31.71%						
El hogar cuenta con servicio de teléfono fijo																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =68.29%																
<i>1</i> =31.71%																
<table border="1"> <tr><td>El hogar cuenta con lavadora</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si</i></td></tr> </table>	El hogar cuenta con lavadora	<i>No</i>	<i>Si</i>	<table border="1"> <tr><td>Valores Posibles: 2</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> </table>	Valores Posibles: 2	<i>0</i>	<i>1</i>	<table border="1"> <tr><td><i>0</i>=35.54%</td></tr> <tr><td><i>1</i>=64.46%</td></tr> </table>	<i>0</i> =35.54%	<i>1</i> =64.46%						
El hogar cuenta con lavadora																
<i>No</i>																
<i>Si</i>																
Valores Posibles: 2																
<i>0</i>																
<i>1</i>																
<i>0</i> =35.54%																
<i>1</i> =64.46%																
<table border="1"> <tr><td>Trabaja actualmente</td></tr> <tr><td><i>No</i></td></tr> <tr><td><i>Si, con remuneración</i></td></tr> <tr><td><i>Si, sin remuneración</i></td></tr> <tr><td><i>Si, por ser práctica de estudios</i></td></tr> </table>	Trabaja actualmente	<i>No</i>	<i>Si, con remuneración</i>	<i>Si, sin remuneración</i>	<i>Si, por ser práctica de estudios</i>	<table border="1"> <tr><td>Valores Posibles: 4</td></tr> <tr><td><i>0</i></td></tr> <tr><td><i>1</i></td></tr> <tr><td><i>2</i></td></tr> <tr><td><i>3</i></td></tr> </table>	Valores Posibles: 4	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<table border="1"> <tr><td><i>0</i>=51.22%</td></tr> <tr><td><i>1</i>=9.41%</td></tr> <tr><td><i>2</i>=39.02%</td></tr> <tr><td><i>3</i>=0.35%</td></tr> </table>	<i>0</i> =51.22%	<i>1</i> =9.41%	<i>2</i> =39.02%	<i>3</i> =0.35%
Trabaja actualmente																
<i>No</i>																
<i>Si, con remuneración</i>																
<i>Si, sin remuneración</i>																
<i>Si, por ser práctica de estudios</i>																
Valores Posibles: 4																
<i>0</i>																
<i>1</i>																
<i>2</i>																
<i>3</i>																
<i>0</i> =51.22%																
<i>1</i> =9.41%																
<i>2</i> =39.02%																
<i>3</i> =0.35%																

Se puede establecer que la mayor parte de los estudiantes tienen celular y lavadora, no obstante pocos cuentan con servicio de internet en sus hogares y no tienen teléfono fijo. Respecto a su situación laboral, un poco más de la mitad de los estudiantes no trabaja, y si lo hacen es por tener experiencia.

Continuación de los factores de estudio socioeconómicos.

FACTORES DE ESTUDIO SOCIOECONÓMICOS																									
DESCRIPCIÓN	RANGO	RESUMEN																							
<table border="1"> <tr><td>Horas que trabaja a la semana</td></tr> <tr><td><i>No trabaja</i></td></tr> <tr><td><i>De 1 a 5 horas</i></td></tr> <tr><td><i>De 6 a 10 horas</i></td></tr> <tr><td><i>De 11 a 15 horas</i></td></tr> <tr><td><i>De 16 a 20 horas</i></td></tr> </table>	Horas que trabaja a la semana	<i>No trabaja</i>	<i>De 1 a 5 horas</i>	<i>De 6 a 10 horas</i>	<i>De 11 a 15 horas</i>	<i>De 16 a 20 horas</i>	<table border="1"> <tr><td>Valores Posibles: 5</td></tr> <tr><td>1</td></tr> <tr><td>2</td></tr> <tr><td>3</td></tr> <tr><td>4</td></tr> <tr><td>5</td></tr> </table>	Valores Posibles: 5	1	2	3	4	5	<table border="1"> <tr><td>1=51.57%</td></tr> <tr><td>2=2.44%</td></tr> <tr><td>3=10.80%</td></tr> <tr><td>4=8.01%</td></tr> <tr><td>5=27.18%</td></tr> </table>	1=51.57%	2=2.44%	3=10.80%	4=8.01%	5=27.18%						
Horas que trabaja a la semana																									
<i>No trabaja</i>																									
<i>De 1 a 5 horas</i>																									
<i>De 6 a 10 horas</i>																									
<i>De 11 a 15 horas</i>																									
<i>De 16 a 20 horas</i>																									
Valores Posibles: 5																									
1																									
2																									
3																									
4																									
5																									
1=51.57%																									
2=2.44%																									
3=10.80%																									
4=8.01%																									
5=27.18%																									
Número de dormitorios de la residencia	Rango: [1,10]	<table border="1"> <tr><td>1=5.92%</td></tr> <tr><td>2=24.74%</td></tr> <tr><td>3=44.25%</td></tr> <tr><td>4=18.47%</td></tr> <tr><td>5=4.53%</td></tr> <tr><td>6=1.74%</td></tr> <tr><td>7=1.14%</td></tr> <tr><td>8=0%</td></tr> <tr><td>9=0%</td></tr> <tr><td>10=0%</td></tr> </table>	1=5.92%	2=24.74%	3=44.25%	4=18.47%	5=4.53%	6=1.74%	7=1.14%	8=0%	9=0%	10=0%													
1=5.92%																									
2=24.74%																									
3=44.25%																									
4=18.47%																									
5=4.53%																									
6=1.74%																									
7=1.14%																									
8=0%																									
9=0%																									
10=0%																									
<table border="1"> <tr><td>Ingresos mensuales en salarios mínimos legales vigentes</td></tr> <tr><td><i>Menos de 1 SM</i></td></tr> <tr><td><i>Entre 1 y Menos de 2 SM</i></td></tr> <tr><td><i>Entre 2 y Menos de 3 SM</i></td></tr> <tr><td><i>Entre 3 y Menos de 5 SM</i></td></tr> <tr><td><i>Entre 5 y Menos de 7 SM</i></td></tr> <tr><td><i>Entre 7 y Menos de 10 SM</i></td></tr> <tr><td><i>10 o más SM</i></td></tr> </table>	Ingresos mensuales en salarios mínimos legales vigentes	<i>Menos de 1 SM</i>	<i>Entre 1 y Menos de 2 SM</i>	<i>Entre 2 y Menos de 3 SM</i>	<i>Entre 3 y Menos de 5 SM</i>	<i>Entre 5 y Menos de 7 SM</i>	<i>Entre 7 y Menos de 10 SM</i>	<i>10 o más SM</i>	<table border="1"> <tr><td>Valores Posibles: 7</td></tr> <tr><td>1</td></tr> <tr><td>2</td></tr> <tr><td>3</td></tr> <tr><td>4</td></tr> <tr><td>5</td></tr> <tr><td>6</td></tr> <tr><td>7</td></tr> </table>	Valores Posibles: 7	1	2	3	4	5	6	7	<table border="1"> <tr><td>1=20.56%</td></tr> <tr><td>2=52.61%</td></tr> <tr><td>3=18.47%</td></tr> <tr><td>4=5.23%</td></tr> <tr><td>5=2.79%</td></tr> <tr><td>6=0%</td></tr> <tr><td>7=0.35%</td></tr> </table>	1=20.56%	2=52.61%	3=18.47%	4=5.23%	5=2.79%	6=0%	7=0.35%
Ingresos mensuales en salarios mínimos legales vigentes																									
<i>Menos de 1 SM</i>																									
<i>Entre 1 y Menos de 2 SM</i>																									
<i>Entre 2 y Menos de 3 SM</i>																									
<i>Entre 3 y Menos de 5 SM</i>																									
<i>Entre 5 y Menos de 7 SM</i>																									
<i>Entre 7 y Menos de 10 SM</i>																									
<i>10 o más SM</i>																									
Valores Posibles: 7																									
1																									
2																									
3																									
4																									
5																									
6																									
7																									
1=20.56%																									
2=52.61%																									
3=18.47%																									
4=5.23%																									
5=2.79%																									
6=0%																									
7=0.35%																									

Se observa que los estudiantes que trabajan en su mayoría lo hacen entre 16 y 20 horas a la semana, con base en el dato anterior se puede afirmar que los estudiantes de Ingeniería Electromecánica trabajan entre 3 y 4 horas diarias por ganar experiencia o ayudar a sus padres. Por otra parte se observa

que los hogares cuentan con tres dormitorios por residencia en su gran mayoría contrastando con la información que en los hogares el promedio de personas en el hogar es de 4.

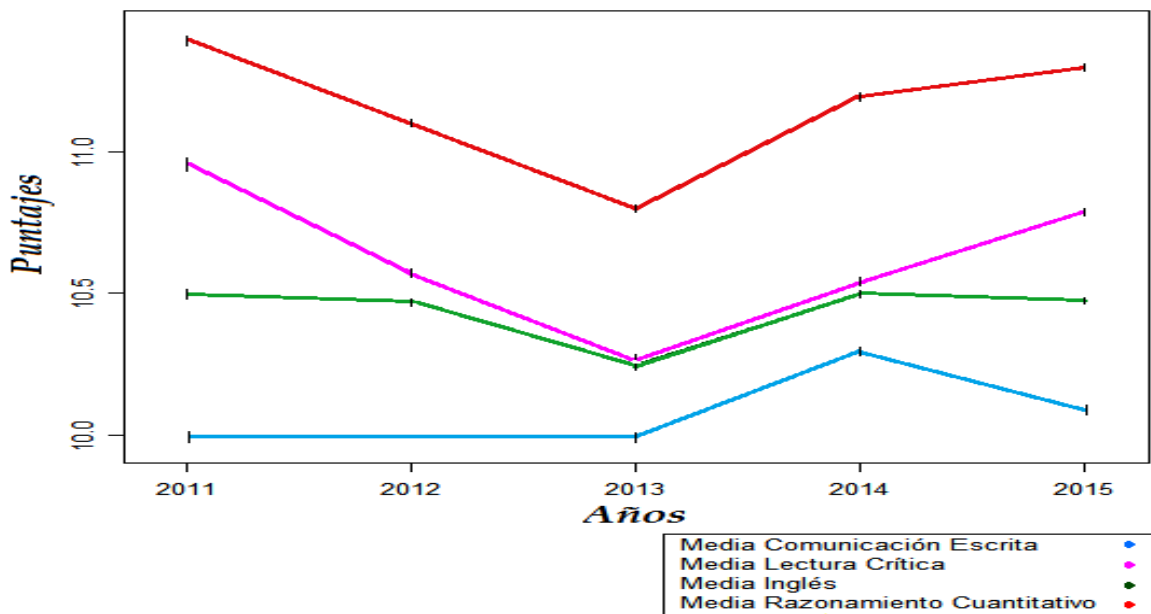
En cuanto a los ingresos familiares mensuales se observa que en gran parte de los hogares llegan entre 1 y menos de 2 salarios, acorde a un hogar de estrato 2, el cual es al que la mayoría de estudiantes de Ingeniería Electromecánica pertenecen.

FACTORES DE ESTUDIO ACADÉMICOS																	
DESCRIPCIÓN	RANGO	RESUMEN															
<table border="1"> <tr> <td>Tipo de bachillerato obtenido</td> </tr> <tr> <td><i>Académico</i></td> </tr> <tr> <td><i>Técnico</i></td> </tr> <tr> <td><i>Normalista</i></td> </tr> <tr> <td><i>Desconocido</i></td> </tr> </table>	Tipo de bachillerato obtenido	<i>Académico</i>	<i>Técnico</i>	<i>Normalista</i>	<i>Desconocido</i>	<table border="1"> <tr> <td>Valores Posibles: 4</td> </tr> <tr> <td><i>A</i></td> </tr> <tr> <td><i>T</i></td> </tr> <tr> <td><i>N</i></td> </tr> <tr> <td><i>D</i></td> </tr> </table>	Valores Posibles: 4	<i>A</i>	<i>T</i>	<i>N</i>	<i>D</i>	<table border="1"> <tr> <td><i>A=43.9%</i></td> </tr> <tr> <td><i>T=55.4%</i></td> </tr> <tr> <td><i>N=0.7%</i></td> </tr> <tr> <td><i>D=0%</i></td> </tr> </table>		<i>A=43.9%</i>	<i>T=55.4%</i>	<i>N=0.7%</i>	<i>D=0%</i>
Tipo de bachillerato obtenido																	
<i>Académico</i>																	
<i>Técnico</i>																	
<i>Normalista</i>																	
<i>Desconocido</i>																	
Valores Posibles: 4																	
<i>A</i>																	
<i>T</i>																	
<i>N</i>																	
<i>D</i>																	
<i>A=43.9%</i>																	
<i>T=55.4%</i>																	
<i>N=0.7%</i>																	
<i>D=0%</i>																	
Nivel educativo del padre; madre	Rango:[0,99]																
<i>Ninguno</i>	0	0=4.53%	0=1.39%														
<i>No tuvo escuela</i>	1	1=0%	1=0%														
<i>Preescolar</i>	2	2=0%	2=0%														
<i>Básica primaria</i>	3	3=0%	3=0%														
<i>Básica secundaria</i>	4	4=0%	4=0%														
<i>Media vocacional</i>	5	5=0%	5=0%														
<i>Tecnológico o técnico</i>	6	6=0%	6=0%														
<i>Universitario</i>	7	7=0%	7=0%														
<i>Postgrado</i>	8	8=0%	8=0%														
<i>Primaria incompleta</i>	9	9=17.07%	9=14.98%														
<i>Primaria completa</i>	10	10=21.95%	10=16.02%														
<i>Secundaria (bachillerato) incompleto</i>	11	11=15.67%	11=25.08%														
<i>Secundaria (bachillerato) completo</i>	12	12=19.5%	12=21.6%														
<i>Educación técnica o tecnológica incompleta</i>	13	13=2.09%	13=1.04%														
<i>Educación técnica o tecnológica completa</i>	14	14=7.31%	14=5.92%														
<i>Educación profesional incompleta</i>	15	15=2.09%	15=3.48%														
<i>Educación profesional completa</i>	16	16=6.27%	16=6.62%														
<i>Postgrado</i>	17	17=3.48%	17=3.83%														
<i>No sabe</i>	99	99=0%	99=0%														

La mayor parte de los estudiantes de Ingeniería electromecánica se graduó del colegio con un tipo de bachillerato técnico. En cuanto a la educación promedio alcanzada por el padre el mayor nivel se sitúa en primaria completa, seguido por bachillerato completo. Ahora al observar el máximo nivel de educación logrado por la madre que fue bachillerato incompleto, seguido por bachillerato completo al igual que el del padre, se puede determinar que:

- El nivel educativo de padres y madres es bajo, al considerar que apenas logran superar el nivel de bachillerato intermedio, es decir hasta la básica (noveno grado).
- Comparando el nivel de los padres, se alcanza a notar cierta diferencia a favor del entorno académico de la madre, es decir que alcanzan un nivel educativo más alto.

La siguiente figura muestra el resultado promedio por año para las competencias de razonamiento cuantitativo, inglés, comunicación escrita y lectura crítica.



Podemos observar que los puntajes de los estudiantes de Ingeniería Electromecánica en los niveles del componente genérico han mostrado un descenso a partir del año 2011 y hasta el 2013, para en los dos años siguientes marcar una paulatina mejora, con excepción de comunicación escrita que demuestra un nivel constante aunque siempre en el rango de regular.

Comunicación escrita obtuvo en estos años los puntajes promedio más bajos, indicando que los estudiantes presentan dificultades a la hora de comunicar ideas por escrito con base en el análisis de información suministrada, mientras que razonamiento cuantitativo es el componente que en promedio demuestra un nivel de muy bueno y excelente a través de este periodo de tiempo.

Por otra parte el desempeño en Lectura crítica permanece en un rango de puntajes entre bueno y muy bueno, mientras inglés se sitúa en el rango de bueno, es decir que los estudiantes comprenden y utilizan expresiones cotidianas de uso frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato.

Descripción univariada por componente

Los resultados de cada módulo referentes a la UPTC se presentan en una tabla, la cual incluye de los resultados correspondientes al Programa, siendo n , el número de estudiantes que presentaron la prueba, \bar{x} : puntaje promedio individual obtenido, DE: desviación estándar, CA: coeficiente de asimetría, CC: coeficiente de curtosis, CV: coeficiente de variación, Q_1 , Q_2 y Q_3 representan los cuartiles uno, dos y tres respectivamente y NA se refiere al número de datos faltantes.

Tabla 1. Resultados Comunicación Escrita

Año	n	\bar{x}	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	35	10.0	0.24	0.02	0.08	9.4	9.8	10.4	0
2012	66	10.0	-0.7	1.4	0.11	9.6	10.0	10.7	0
2013	67	10.0	-1.01	3.0	0.08	9.5	9.9	10.5	1
2014	60	10.3	0.01	-0.61	0.08	9.7	10.2	11.3	1
2015	57	10.1	-4.28	26.9	0.16	9.5	10.1	10.9	0

Se observa que, para cada uno de los años que ha presentado la prueba, los estudiantes de Ingeniería Electromecánica presentan puntajes similares (coeficiente de variación del orden del 11%), con excepción del año 2015. A juzgar por los valores de los coeficientes de curtosis los puntajes en **COMUNICACIÓN ESCRITA** indican que no hay estudiantes que hayan obtenido puntajes muy diferentes entre los años 2011 y 2014, es decir los resultados se concentran alrededor de su media para cada año, exceptuando el año 2015 donde se indica la presencia de datos atípicos; por lo anterior y teniendo en cuenta que el promedio es de 10.08 para esta prueba, se puede concluir que a través de los años el promedio de puntajes se ubica en la categoría regular, lo cual apunta a que los estudiantes de Ingeniería Electromecánica presentan dificultad para comunicar ideas por escrito con base en el análisis de la información suministrada. Los resultados anteriores se pueden complementar con los resultados cualitativos en la prueba de **DESEMPEÑO EN COMUNICACIÓN ESCRITA**, Tabla 2.

Tabla 2. Desempeño en Comunicación Escrita

Año	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7	Nivel 8	n	NA
2011	0%	2.8%	17.1%	48.6%	20%	8.6%	2.8%	0%	35	0
2012	4.5%	1.5%	10.6%	39.4%	24.2%	16.7%	3.03%	0%	66	0
2013	2.9%	1.5%	1.5%	52.2%	32.8%	9%	0%	0%	67	0
2014	0%	0%	8.3%	40%	25%	25%	1.7%	0%	60	1
2015	0%	0%	7.1%	44.6%	35.7%	7.1%	5.4%	0%	56	1

Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes (44.72%) se ubican en el nivel 4 de desempeño, indicando que los estudiantes encuentran la idea central que se desarrolla de acuerdo con una intención comunicativa, hallando también la estructura básica, en otras palabras, identifican la introducción al tema que se abordará, un desarrollo y una conclusión. Sin embargo, como el texto no incluye toda la información necesaria (progresión temática), la organización no es completamente efectiva, o rompen 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 28.1% de los estudiantes se ubican en el nivel 5 y pocos (13.3%) en el nivel 6.

En la Tabla 3 se presentan los resultados obtenidos en **INGLÉS**. Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Ingeniería Electromecánica han sido heterogéneos respecto a los puntajes (los coeficientes de variación, oscilan entre el 8% y el 16%). Además los coeficientes de curtosis indican que existen estudiantes que obtuvieron puntajes muy diferentes a los del grupo con los que presentaron la prueba, exceptuando el año 2014.

Por lo tanto sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 10.43, se puede indicar que los puntajes a través de los años los ubica en la categoría de bueno.

Tabla 3. Resultados Módulo Inglés

<i>Año</i>	<i>n</i>	\bar{x}	DE	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	35	10.5	1.08	2.2	7.15	0.10	9.8	10.3	10.9	0
2012	66	10.47	1.14	1.6	3.12	0.11	9.9	10.1	10.8	0
2013	68	10.25	1.20	1.9	4.3	0.11	9.5	10	10.5	0
2014	61	10.5	0.93	0.8	-0.04	0.08	9.9	10.2	11.1	0
2015	57	10.47	1.77	-3.4	21.9	0.16	9.9	10.3	11.2	0

Ahora bien, sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes se ubican en el nivel A2 de **DESEMPEÑO EN INGLÉS**, indicando que los estudiante son capaces de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes como información básica sobre ellos mismos y su familia, compras, lugares de interés, ocupaciones, etc., como se muestra en la Tabla 4.

Tabla 4. Resultados de Desempeño en Inglés

DESEMPEÑO EN INGLÉS	A-	A1	A2	B+	B1	<i>n</i>	NA
2011	11.4%	25.7%	28.6%	5.7%	28.57%	35	0
2012	10.6%	30.3%	33.3%	9.1%	16.7%	66	0

2013	22.1%	29.4%	26.5%	7.3%	14.7%	68	0
2014	4.9%	31.1%	27.9%	9.8%	26.2%	61	0
2015	5.3%	24.6%	29.8%	7.0%	33.3%	57	0

La siguiente tabla muestra la descripción por año de los resultados obtenidos por los estudiantes de Ingeniería Electromecánica para el componente en **LECTURA CRÍTICA**, en la cual se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Ingeniería Electromecánica son muy homogéneos (coeficientes de variación, son inferiores al 10%); excepto en el año 2015, donde se observa heterogeneidad en los puntajes. 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. Los valores indican que entre el año 2011 y el 2015 el puntaje promedio de los estudiantes ha estado entre 10.26 y 10.96, lo cual los ubica en las categorías de bueno y muy bueno en la escala, por lo tanto, los estudiantes de Ingeniería Electromecánica tienen la capacidad para leer de manera analítica y reflexiva, además de comprender los planteamientos expuestos en un texto exigiendo que el lector identifique y recupere información presente en uno o varios textos, construya un sentido global, establezca relaciones entre enunciados y evalúe su intencionalidad.

Tabla 5. Resultados de Lectura Crítica

<i>Año</i>	<i>n</i>	\bar{x}	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	<i>NA</i>
2011	35	10.96	0.91	1.41	0.07	10.5	10.8	11.4	0
2012	66	10.57	0.16	-0.47	0.06	10.2	10.6	10.9	0
2013	68	10.26	0.61	0.72	0.07	9.8	10.2	10.7	0
2014	61	10.54	1.22	3.54	0.08	9.8	10.4	11.0	0
2015	57	10.79	-5.52	37.5	0.14	10.5	10.9	11.4	0

La siguiente tabla muestra la descripción por año de los resultados obtenidos por los estudiantes de Ingeniería Electromecánica para el componente en **RAZONAMIENTO CUANTITATIVO**, en la cual se tiene que, para el periodo comprendido entre los años 2011 y 2014, los puntajes de los estudiantes de Ingeniería Electromecánica son muy homogéneos (coeficientes de variación, del orden del 10%), además al juzgar los valores de los coeficientes de curtosis los puntajes en **RAZONAMIENTO CUANTITATIVO** indican que no hay estudiantes que hayan obtenido puntajes muy diferentes, es decir los resultados se concentran alrededor de su media para cada año. Ahora bien, para el año 2015 se presentó un alto coeficiente de variación y de curtosis, indicando que existieron puntajes que superaron la media (cerca del 50%), confirmado por los cuartiles. Por lo anterior se tiene que el promedio general de la prueba es de 11.1 concluyendo que a través de los años se ubica en la categoría de excelente, lo cual apunta a que los estudiantes desarrollan en buena medida competencias relacionadas con las habilidades en la comprensión de conceptos básicos de las matemáticas para analizar, modelar y resolver problemas aplicando métodos y procedimientos cuantitativos y esquemáticos.

Tabla 6. Resultados Razonamiento Cuantitativo

<i>Año</i>	<i>n</i>	\bar{x}	<i>CA</i>	<i>CC</i>	<i>CV</i>	Q_1	Q_2	Q_3	<i>NA</i>
2011	35	11.4	-0.05	-0.27	0.06	10.8	11.4	12.1	0
2012	66	11.1	0.38	0.8	0.08	10.5	10.9	11.6	0

2013	68	10.8	0.41	0.16	0.08	10.2	10.8	11.5	0
2014	61	11.2	0.57	-0.6	0.08	10.6	10.9	11.9	0
2015	57	11.3	-4.4	28.3	0.16	10.8	11.4	11.7	1

Para las siguientes conclusiones se aplicó la prueba de correlación de Pearson sobre los puntajes numéricos del componente genérico de las pruebas, de este modo al relacionar los puntajes obtenidos y algunas características de los estudiantes se encontró correlación entre la variable edad y los puntajes en razonamiento cuantitativo, lectura crítica e inglés, además indicando que mientras mayor sea la edad menor será el puntaje en dichas categorías (correlación negativa).

En cuanto a los desempeños del componente genérico de la prueba se obtuvo los siguientes resultados a través del test ji-cuadrado de Pearson indicando que para el desempeño en comunicación escrita se encuentra asociado el sexo del estudiante y el número de horas que trabaja a la semana, de igual manera para el desempeño en inglés se encontró dependencia con el nivel educativo del padre y el salario mensual del grupo familiar en el hogar.

Se determinó realizar el test de normalidad de Shapiro Wilk (ver Anexo A) de cada variable respuesta numérica, obteniendo como resultado que ninguno de los puntajes del componente genérico de la prueba tienden a distribuirse normalmente

En este orden de ideas al asumir la no normalidad de los puntajes numéricos del componente genérico de las pruebas Saber Pro se realizó la prueba no paramétrica de Kruskal Wallis con el fin de determinar asociación entre variables, obteniendo que:

El puntaje en **RAZONAMIENTO CUANTITATIVO** presenta diferencia entre los promedios según las variables género, horas de trabajo a la semana, nivel educativo de los padres, ocupación de la madre y valor anual de la matrícula.

Tabla 7. Razonamiento Cuantitativo

Variables	Media
Mujeres	10.7
Hombres	11.1
Trabaja entre 1 y 5 horas	11.5
Trabaja entre 6 y 10 horas	10.6
Nivel educativo universitario de los padres	11.6
Nivel educativo sin estudio de los padres	10.9
Ocupación administración en la madre	10.95
Ocupación Pensionada de la madre	11.7

Así mismo para el puntaje en **LECTURA CRÍTICA** existe nexo con las variables tener trabajo, horas de trabajo a la semana, nivel educativo de la madre, ocupación del padre y valor anual de la matrícula.

Tabla 8. Lectura critica

VARIABLES	Media
No Tiene trabajo	10.7
Trabaja como ayudante sin remuneracion	10.3
Ocupación padre: Administrador	10.5
Ocupación padre: Obrero	10.7
Matricula anual: No paga	10.8
Matricula anual: Entre 3 y 5 millones	9.9

Además se halló correlación entre los puntajes en **INGLÉS** y las variables estrato, horas de trabajo a la semana, nivel de educación de los padres, ocupación de la madre, salario mensual del grupo familiar en el hogar y tiene trabajo.

Tabla 9. Puntaje en Inglés

Variable	Media
Estrato 2	10.14
Estrato 1	10.4
Trabaja 0 horas	10.3
Trabaja entre 16 y 20 horas	10.0
Madre con estudios universitarios	10.55
Madre sin estudios	9.86
Salario mensual familiar más de 10 SMLV	13.94
Salario mensual familiar menos de 1 SMLV	10.18

Finalmente, para el puntaje en **COMUNICACIÓN ESCRITA** no se halló ningún tipo de asociación con otra variable.

FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO

La construcción de los modelos se llevó a cabo a través de la especificación, selección (modelo más parsimonioso), evaluación e interpretación de los resultados. Para el ajuste de los modelos se utilizó el criterio StepGAICALL.A.

El mejor ajuste que se presenta para razonamiento cuantitativo, lectura crítica, comunicación escrita e inglés teniendo en cuenta el criterio AIC es la Skew t type 3, Logistic y Skew t type 2 con un AIC de 804.1670, 737.1300, 772.2682 y 798.2387 respectivamente.

La tabla 11 muestra la distribución para cada variable respuesta y sus correspondientes funciones de enlace.

Tabla 10. Distribuciones, funciones de enlace

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	δ	ν	τ
Skew t type 3 (ST3)	Razonamiento cuantitativo Lectura crítica	identity	log	log	log
Logistic (LO)	Comunicación escrita	identity	log	---	---
Skew t type 2 (ST2)	Inglés	identity	log	identity	log

La Tabla 11, 12 y 13 muestra los factores que influyen aun nivel de significancia del 1%, 5% o 10% para cada componente, siendo RC: razonamiento cuantitativo, LC: lectura crítica, CE: comunicación escrita, IN: inglés, D.CE: desempeño comunicación escrita, D.IN: desempeño inglés, ahora observemos que:

“↑” indica que la variable influye a la hora de obtener puntajes altos.

“↓” indica la posibilidad de obtener valores bajos.

“—” indica que la variable no hace parte, o no resultado significativa en el modelo óptimo.

Tabla 11. Resumen variables demográficas influyentes

VARIABLES DEMOGRÁFICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Sexo (M)	↑	-	-	-	↓	-
Sexo (F)	↓	-	-	-	↑	-
Edad	↑	↑	↑	↑	↑	-
Estado civil (Casado)	-	-	-	↑	-	↑
Estado civil (Soltero)	-	-	-	↓	-	↓
Estado civil (Unión libre)	-	-	-	↓	-	↓
Estado civil (Separado)	-	-	-	↓	-	↓
Hogar actual (Permanente)	↓	-	-	-	-	-
Hogar actual (Temporal)	↑	-	-	-	-	-

De acuerdo a la Tabla 11, se observa que tener estado civil de casado genera mejores puntajes en los componentes de inglés y desempeño en inglés, en comparación a estudiantes con otro tipo de estado civil, además contar con un hogar temporal también aumenta la posibilidad de tener puntajes altos en comparación a quienes cuentan con un hogar permanente en razonamiento cuantitativo.

En cuanto a la edad para cada componente con excepción del desempeño en inglés se tiene que, el ser más joven aumenta la posibilidad de obtener puntajes más altos. Los hombres obtienen mejor puntaje en el componente de razonamiento cuantitativo mientras las mujeres lo hacen en el componente de desempeño en comunicación.

Tabla 12. Resumen variables académicas influyentes

VARIABLES ACADÉMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Tipo de bachillerato (Técnico)	-	↑	-	-	-	-
Tipo de bachillerato (Académico)	-	↓	-	-	-	-
Tipo de bachillerato (Normalista superior)	-	↑	-	-	-	-
Nivel educativo padre (Universitario)	-	-	↑	↑	-	↑
Nivel educativo padre (primaria)	-	-	↓	↓	-	↓

Nivel educativo padre (Técnico)	-	-	↑	↑	-	↑
Nivel educativo madre (Universitario)	↑	-	-	-	-	-
Nivel educativo madre (Primaria)	↓	-	-	-	-	-
Nivel educativo madre (Secundaria)	↑	-	-	-	-	-
Nivel educativo madre (Sin estudio)	↓	-	-	-	-	-

Por otra parte, en cuanto al factor académico se tiene que, tener madre con nivel educativo universitario y secundaria mejora los puntajes en razonamiento cuantitativo, en comparación a tener niveles educativos de primaria y sin estudio, así mismo tener padre con nivel educativo universitario y técnico aumenta el puntaje en comunicación escrita e inglés, además obtener un tipo de bachillerato académico disminuye la posibilidad de lograr puntajes altos respecto a un bachiller de tipo técnico en lectura crítica.

Tabla 13. Resumen variables socioeconómicas influyentes

VARIABLES SOCIOECONÓMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Valor anual matrícula (Menos de 500 mil pesos)	-	-	-	-	-	↓
Valor anual matrícula (Entre 1 y 3 millones de pesos)	-	-	-	-	-	↓
Valor anual matrícula (Entre 3 y 5 millones de pesos)	-	-	-	-	-	↑
Ocupación madre (Administrador)	↓	↓	-	-	-	-
Ocupación madre (Obrera)	↑	↑	-	-	-	-
Conexión a internet (Sí)	-	-	↑	-	-	-
Conexión a internet (No)	-	-	↓	-	-	-
Trabaja actualmente (Sí, con remuneración en dinero)	-	-	-	-	-	↓
Trabaja actualmente (Sí, por experiencia)	-	-	-	-	-	↓
Trabaja actualmente (Sí, práctica obligatoria)	-	-	-	-	-	↓
Trabaja actualmente (No)	-	-	-	-	-	↑
Horas de trabajo a la semana (1-5)	↑	-	-	-	↑	-
Horas de trabajo a la semana (6-10)	↓	-	-	-	↓	-
Horas de trabajo a la semana (11-15)	↓	-	-	-	↑	-
Horas de trabajo a la semana (16-20)	↓	-	-	-	↓	-

Estrato 1	-	-	-	↑	-	↑
Estrato 2	-	-	-	↓	-	-
Estrato 3	-	-	-	↓	-	-
Estrato 4	-	-	-	↑	-	-
Salario mensual familiar(1-2)	-	-	-	↓	-	-
Salario mensual familiar(2-3)	-	-	-	↑	-	-
Salario mensual familiar(3-5)	-	-	-	↑	-	-
Salario mensual familiar(5-7)	-	-	-	↑	-	-
Salario mensual familiar(+10)	-	-	-	↑	-	-

De acuerdo a la Tabla 13, se observa que realizar pagos de matrícula entre 3 y 5 millones de pesos, aumenta la posibilidad de obtener puntajes altos en desempeño en inglés, en comparación a quienes realizan pagos de cualquier otro valor. Por otra parte, al tener madre con ocupación de administradora se obtiene puntajes más bajos en el componente de razonamiento cuantitativo y lectura crítica, comparado con cualquier otro tipo de ocupación.

En cuanto a la conexión a internet, obtiene más puntos en comunicación escrita quien posea este servicio, así también como un estudiante que no trabaje.

Ahora bien, un estudiante que trabaje semanalmente entre 1 y 5 horas obtendrá mejores puntajes respecto a cualquier categoría comparativa en razonamiento cuantitativo, igual caso en el desempeño en comunicación escrita, con excepción de la categoría de trabajar entre 11 y 15 horas a la semana donde también aumenta esta probabilidad.

En el puntaje en inglés, pertenecer a estrato 1 y 4 además de tener un salario mensual familiar superior a 2 salarios mínimos legales vigentes mejora este valor.

5. Conclusiones

Ahora bien, al determinar aspectos de correlación y asociación de variables se logró observar que la edad, el sexo, el número de horas de trabajo a la semana, así como la educación de los padres y el salario mensual familiar inciden en los puntajes y desempeños de los estudiantes de Ingeniería Electromecánica.

Se obtuvo una adecuada distribución para cada uno de los puntajes asociados al componente genérico, para finalmente asociar las variables respuesta cada tipo de factor de manera independiente, obteniendo como resultado 18 modelos (6 por cada factor).

Para Razonamiento Cuantitativo los factores demográficos que inciden en este componente son la edad del estudiante al momento de presentar la prueba, la situación o tipo de hogar en el que se encuentre y finalmente su sexo. El puntaje según variables académicas sólo es determinado por el nivel educativo de la madre, mientras que, para el factor socioeconómico se asocian el número de horas de trabajo en la semana, y la ocupación de la madre.

En Lectura Crítica la edad en el aspecto demográfico, el tipo de bachillerato realizado por el estudiante al finalizar la educación media por la parte académica, y en el aspecto socioeconómico la ocupación de la madre.

Para Comunicación Escrita el puntaje se explica por la edad, el estado civil y el sexo del estudiante en el aspecto demográfico. Para el factor académico el nivel educativo del padre, y en las variables socioeconómicas el tener servicio de Internet en la residencia.

En Inglés como variables demográficas la edad y el estado civil, en las académicas el nivel educativo del padre, y para el factor socioeconómico el estrato, además del salario mensual familiar.

Para los casos del desempeño en comunicación escrita y desempeño en inglés se realizó una regresión logística ordinal y se tuvo en cuenta el criterio Akaike (AIC) en la selección del modelo.

Para desempeño en comunicación escrita se encontró que la variable demográfica sexo del estudiante incide en el desempeño, además no existe factor académico que explique el nivel en comunicación escrita, por otra parte el tener servicio de internet si se relaciona con este módulo.

En el desempeño en inglés el estado civil como variable demográfica, el nivel educativo del padre en cuanto a factor académico, el estrato, tener trabajo y el valor anual de la matrícula como parte socioeconómica explica el nivel de desempeño de un estudiante.

Referencias Bibliográficas

- Agresti, A. (2002), Análisis Categórico de Datos, Universidad de la Florida, Gainesville, Florida, segunda edición, Publicaciones corporación John Wiley e hijo, Hoboken, Nueva Jersey.
- Arias I., Ávila 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 DC, Revista educación y desarrollo social, Bogotá.
- Bogoya, D. (2013), Benchmarking: elementos de calidad de la educación superior en Colombia, SCIMAGO: base de datos universidad nacional de Colombia.
- Bogoya, D. (2009), Evaluación de la Calidad de la Educación Superior y valor académico agregado, Revista: Anfibios académicos: pedagogías, docencia y evaluación en la educación superior, Bogotá, Colombia, Universidad Nacional de Colombia.
- Brunner, J. Gomes, C. Fordham, E. (2016), Revisión de políticas nacionales de educación: La educación en Colombia, Organización para la Cooperación y el Desarrollo Económicos (OCDE), paginas 15-75, 268-329.
- Cañadas, L. (2013), Regresión logística: tratamiento computacional con R, Universidad de Granada, España.
- Díaz, L. (2012), Análisis estadístico de datos categóricos, Universidad Nacional de Colombia UNAL, Bogotá.
- Egea J., Kessler M., Regresión lineal con R commander, Universidad Politécnica de Cartagena, Departamento de Matemática Aplicada y Estadística.

- Erazo, O. (2012). El rendimiento académico, un fenómeno de múltiples relaciones y complejidades. *Revista vanguardia psicológica, clínica teórica y práctica*, 2(2), 144-173.
- Escuela de ingeniería electromecánica, Universidad Pedagógica y Tecnológica de Colombia UPTC (2016), Acta de comité curricular sobre las pruebas saber pro en la escuela de ingeniería electromecánica de la UPTC seccional Duitama.
- Guzmán Tovar C., Serna, C. & Hoyos, D. (2012). Las pruebas ecaes en Colombia: una evaluación a la evaluación. *Panorama*, 6 (10), 33-54.
- Iglesias, T. (2013), *Métodos de bondad de ajuste en Regresión logística*, Universidad de Granada.
- Instituto Colombiano para la Evaluación de la Educación - ICFES. (2014). Bases de datos pruebas Saber pro. [On-line]. Disponible en: <http://www.icfesinteractivo.gov.co>.
- Instituto Colombiano para la Evaluación de la Educación - ICFES. (2011), *Orientaciones para el examen de Estado de calidad de la educación superior saber pro (ecaes), prueba de competencias comunes del área de educación*.
- Isáziga, C. Coello, J (2014), *Estudios sobre calidad de la educación en Colombia*, Instituto colombiano para la evaluación de la educación (ICFES).
- Mesa, H. (2009), *De los ECAES, la autonomía universitaria y el derecho*, Revista Electrónica.
- Ministerio de educación, cultura y deporte, subdirección general de cooperación internacional europea (2003), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*, Ministerio de educación, cultura y deporte, subdirección general de cooperación internacional europea, consejo de Europa para la publicación en inglés y francés.
- Mundial, (2012), "La educación superior en Colombia 2012", Organización para la Cooperación y el Desarrollo Económicos (OCDE).
- OCDE, (2016). *Revisión de políticas nacionales de educación: Educación en Colombia*.
- Salas, W. *formación por competencias en educación superior. Una aproximación conceptual a propósito del caso colombiano*, Revista Iberoamericana de Educación (ISSN: 1681-5653) Universidad de Antioquia, Colombia.
- Pérez, M. (2016), *Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS)*, Trabajo para optar al título de master, Universidad de Vigo, Pontevedra, España.
- Sancho, R. (2012), *Ministerio de Ciencia y Tecnología*, Madrid, España.
- 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.
- Torrado C., Teichler U. (2014), *Estudios sobre calidad de la educación en Colombia: 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*, Instituto Colombiano para la Evaluación de la Educación - ICFES, Bogotá.