

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

**CASO: DISEÑO INDUSTRIAL DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA
DE COLOMBIA FACULTAD SECCIONAL DUITAMA**

John Alexander Bello Vega

Código:201121833

Trabajo de grado presentado como requisito parcial 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

2017

Nota de aceptación

Firma jurado 1

Firma jurado 2

Firma presidente del jurado

Duitama Noviembre de 2017

*A mi madre Nohora Ilma Vega Sánchez. A mi padre
Jesus Bello Bello, a mis hermanos Jesus Armando
Bello Vega, Fabio Andrés Vega. A mis compañeros
de clase, a mis profesores.*

Agradecimientos

Mis más sinceros agradecimientos a:

- A Dios, Mis padres, mis hermanos quienes me dieron motivos para seguir adelante, A mis compañeros.
- A mi directora de monografía docente de la Universidad Pedagógica y Tecnológica de Colombia quien me orientó para llevar a cabo los propósitos de éste trabajo con la mayor disposición y paciencia, y también porque me compartió muchos de sus conocimientos estadísticos.

FACULTAD SECCIONAL DUTAMA L.M.E

E.E RAE No

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

CASO: DISEÑO INDUSTRIAL DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA

AUTOR: John Alexander Bello Vega

DIRECTORA: Carmen Helena Cepeda Araque

PUBLICACIÓN: Duitama, Universidad Pedagógica y Tecnológica de Colombia, 2017.

UNIDAD PATROCINANTE: Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, Licenciatura en Matemáticas y Estadística.

PALABRAS CLAVE: Prueba Saber Pro, Modelos regresión logística, GAMLSS.

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 Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama.

DESCRIPCIÓN: En el documento se presenta un estudio que tiene como eje principal modelar los resultados del componente genérico de las pruebas Saber Pro en el programa de Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama, por medio de modelos de Regresión Logística y Modelos Aditivos Generalizados para Localización, Escala y Forma (GAMLSS), en función de variables demográficas, socioeconómicas y académicas. Se contó con una población de 199 estudiantes del programa presentadas entre el segundo semestre de 2011 a 2015.

FUENTES: Para el desarrollo de este trabajo se consultaron 8 libros de estadística, específicamente de Regresión Logística Ordinal y GAMLSS, además se consultaron 15 revistas y artículos, en los cuales se describían estudios referentes a la prueba Saber Pro y los factores que influyen sobre esta. En el caso de la obtención de los datos se ejecutó a partir de un diseño bibliográfico; se tomó una base de datos suministradas por el ICFES, a través del taller prepa-

ratorio 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 199 filas por 58 columnas: <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados>.

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 último discusión y conclusiones.

METODOLOGÍA: Las faces que permitieron determinar la incidencia de los factores demográficos, socioeconómicos y académicos en el componente genérico para estudiantes del programa de Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama, 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 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.

CONCLUSIONES: Es importante recalcar que los resultados aquí presentados solo hacen referencia a los estudiantes de Diseño Industrial de la UPTC objeto de este estudio y que debido a las características de esta, no hace generalizaciones para la población colombiana.

Respecto al puntaje de razonamiento cuantitativo se determinó, que las variables que afectan dicho puntaje son realizar un pago de matrícula entre 3 y 5 millones de pesos, tener conexión a internet, ser soltero, unión libre, tener un hogar temporal, edad, género, tener madre con nivel educativo universitario y tipo de bachillerato técnico. lo cual esta de acuerdo con los planteamiento de (Gil, F., Sepúlveda, L., Rondón, M., Gómez, C 2013).

En cuanto a lectura crítica las variables que influyeron son la edad, tener personas a cargo, tener madre con nivel educativo universitario y tipo de bachillerato técnico. obteniendo más puntos en dichas pruebas. Por otra parte el factor socioeconómico no explica el puntaje obtenido en lectura crítica.

En cuanto a inglés las variables que influyeron son tener conexión a internet, tener madre con nivel educativo universitario y tipo de bachillerato técnico. lo cual esta de acuerdo con los planteamiento de (Arias 2014). Por otra parte el factor demográfico no explica el puntaje obtenido en inglés.

En cuanto a comunicación escrita las variables que influyeron son trabajar con remuneración en dinero, ocupación del padre como administrador, tener madre con nivel educativo universitario, tomar algún curso de preparación al examen Saber Pro, ser soltero, unión libre, tener un hogar temporal, edad y género. Dada la complejidad de esta investigación no se encontró referentes teóricos que sustentaran algunas de las variables influyentes, siendo este trabajo monográfico innovador debido a que explica como afecta cada factor para cada puntaje en específico.

Adicionalmente, se sugiere que el programa realice simulacros en cuanto a las pruebas del componente genérico ya que como se dio a conocer en las conclusiones el tomar algún curso de preparación al examen aumenta la probabilidad de obtener buenos resultados, además ayudará a concientizar al programa si lo que evalúa es realmente lo que se está evaluando en las pruebas Saber Pro.

Dado que en los aspectos demográficos la Universidad no podría actuar, si se recomienda que desde el aspecto socioeconómico como lo es tener conexión a internet se aumente la cobertura en este campo, ya que este servicio influye sobre el aumento de la probabilidad de obtener buenos resultados.

CONTENIDO

1. PRESENTACIÓN DEL PROYECTO	1
2. REFERENTES TEÓRICOS	4
2.1. PRUEBAS SABER PRO	4
2.2. MODELOS LINEALES (LM)	9
2.3. MODELOS LINEALES GENERALIZADOS (GLM)	10
2.4. MODELOS ADITIVOS GENERALIZADOS (GAM)	12
2.5. MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALA Y FORMA (GAMLSS)	13
2.6. REGRESIÓN LOGÍSTICA ORDINAL (RLO)	21
3. DISEÑO METODOLÓGICO	25
3.1. VARIABLES DE ESTUDIO	25
3.2. METODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS	29
4. DESCRIPCIÓN DE LOS DATOS	30
4.1. DESCRIPCIÓN UNIVARIADA POR COMPONENTE	32
4.2. RESULTADOS SABER PRO AÑO 2016	35
4.3. DESCRIPCIÓN POBLACIONAL DE ESTUDIO	37
5. FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO	43
5.1. ESPECIFICACIÓN DE LOS MODELOS	43
5.2. FACTORES DEMOGRÁFICOS ASOCIABLES AL COMPONENTE GENÉRICO	48
5.3. FACTORES ACADÉMICOS ASOCIABLES AL COMPONENTE GENÉRICO	61
5.4. FACTORES SOCIOECONÓMICOS ASOCIABLES AL COMPONENTE GENÉRICO	77

6. DISCUSIÓN	89
7. CONCLUSIONES	93
8. ANEXO A	95
9. ANEXO B	96
REFERENCIAS BIBLIOGRÁFICAS	97

Lista de Figuras

2-1. Función plot.gamlss()	17
2-2. Función wp()	18
4-1. Resultados promedio por año y competencia	31
5-1. Ajuste datos razonamiento cuantitativo	43
5-2. Prueba gráfica razonamiento cuantitativo	44
5-3. Ajuste datos lectura crítica	44
5-4. Prueba gráfica lectura crítica	45
5-5. Ajuste datos comunicación escrita	45
5-6. Prueba gráfica comunicación escrita	46
5-7. Ajuste datos inglés	46
5-8. Prueba gráfica inglés	47
5-9. Modelo óptimo razonamiento cuantitativo	49
5-10. Prueba ajuste modelo demográfico razonamiento cuantitativo	50
5-11. Diagnóstico modelo puntaje razonamiento cuantitativo factor demográfico . .	51
5-12. Ajuste de la distribución modelo razonamiento cuantitativo factor demográfi- co	51
5-13. Modelo óptimo lectura crítica	52
5-14. Prueba ajuste modelo demográfico lectura crítica	53
5-15. Diagnóstico modelo puntaje lectura crítica factor demográfico	53
5-16. Ajuste de la distribución modelo lectura crítica factor demográfico	54
5-17. Modelo óptimo comunicación escrita	55
5-18. Prueba ajuste modelo demográfico comunicación escrita	56
5-19. Diagnóstico modelo puntaje comunicación escrita factor demográfico	56
5-20. Ajuste de la distribución modelo comunicación escrita factor demográfico . .	57
5-21. Modelo óptimo inglés	58
5-22. Prueba ajuste modelo demográfico inglés	59
5-23. Diagnóstico modelo puntaje inglés factor demográfico	59

5-24. Ajuste de la distribución modelo inglés factor demográfico	60
5-25. Modelo óptimo razonamiento cuantitativo	62
5-26. Prueba ajuste modelo académico razonamiento cuantitativo	63
5-27. Diagnóstico modelo puntaje razonamiento cuantitativo factor académico	64
5-28. Ajuste de la distribución modelo razonamiento cuantitativo factor académico	64
5-29. Modelo óptimo lectura crítica	65
5-30. Prueba ajuste modelo académico lectura crítica	67
5-31. Diagnóstico modelo puntaje lectura crítica factor académico	67
5-32. Ajuste de la distribución modelo lectura crítica factor académico	68
5-33. Modelo óptimo comunicación escrita	69
5-34. Prueba ajuste modelo académico comunicación escrita	70
5-35. Diagnóstico modelo puntaje comunicación escrita factor académico	71
5-36. Ajuste de la distribución modelo comunicación escrita factor académico	71
5-37. Modelo óptimo inglés	73
5-38. Prueba ajuste modelo académico inglés	74
5-39. Diagnóstico modelo puntaje inglés factor académico	75
5-40. Ajuste de la distribución modelo inglés factor académico	75
5-41. Modelo óptimo razonamiento cuantitativo	78
5-42. Prueba ajuste modelo socioeconómico razonamiento cuantitativo	79
5-43. Diagnóstico modelo puntaje razonamiento cuantitativo factor socioeconómico	80
5-44. Ajuste de la distribución modelo razonamiento cuantitativo factor socioeconómico	80
5-45. Modelo óptimo lectura crítica	81
5-46. Prueba ajuste modelo socioeconómico lectura crítica	81
5-47. diagnóstico modelo puntaje lectura crítica factor socioeconómico	82
5-48. Ajuste de la distribución modelo lectura crítica factor socioeconómico	82
5-49. Modelo óptimo comunicación escrita	83
5-50. Prueba ajuste modelo socioeconómico comunicación escrita	84
5-51. Diagnóstico modelo puntaje comunicación escrita factor socioeconómico	84
5-52. Ajuste de la distribución modelo puntaje comunicación escrita factor socioeconómico	85
5-53. Modelo óptimo inglés	86
5-54. Prueba ajuste modelo socioeconómico inglés	87
5-55. Diagnóstico modelo puntaje inglés factor socioeconómico	87
5-56. Ajuste de la distribución modelo inglés factor socioeconómico	88

1 PRESENTACIÓN 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 Diseño Industrial 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 Diseño Industrial, 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 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 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 2012), además de medir el desarrollo de las compe-

tencias 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 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 pertenecía a la familia exponencial de densidades. Para dar solución a este problema se emplearon los GAMLSS, los cuales incluyen distribuciones que presentan una fuerte asimetría y/o curtosis. En cuanto a los desempeños se aplicó regresión logística ordinal.

Es importante recalcar que los resultados aquí presentados solo hacen referencia a los estudiantes de Diseño Industrial de la UPTC objeto de este estudio y que debido a las características de esta, no hace generalizaciones para la población colombiana.

Respecto a el puntaje de razonamiento cuantitativo se determinó, que las variables que afectan

dicho puntaje son realizar un pago de matrícula entre 3 y 5 millones de pesos, tener conexión a internet, ser soltero, unión libre, tener un hogar temporal, edad, género, tener madre con nivel educativo universitario y tipo de bachillerato técnico. lo cual esta de acuerdo con los planteamiento de (Gil, F., Sepúlveda, L., Rondón, M., Gómez, C 2013).

En cuanto a lectura crítica las variables que influyeron son la edad, tener personas a cargo, tener madre con nivel educativo universitario y tipo de bachillerato técnico. obteniendo más puntos en dichas pruebas. Por otra parte el factor socioeconómico no explica el puntaje obtenido en lectura crítica.

En cuanto a inglés las variables que influyeron son tener conexión a internet, tener madre con nivel educativo universitario y tipo de bachillerato técnico. lo cual esta de acuerdo con los planteamiento de (Arias 2014). Por otra parte el factor demográfico no explica el puntaje obtenido en inglés.

En cuanto a comunicación escrita las variables que influyeron son trabajar con remuneración en dinero, ocupación del padre como administrador, tener madre con nivel educativo universitario, tomar algún curso de preparación al examen Saber Pro, ser soltero, unión libre, tener un hogar temporal, edad y género. Dada la complejidad de esta investigación no se encontró referentes teóricos que sustentaran algunas de las variables influyentes, siendo este trabajo monográfico innovador debido a que explica como afecta cada factor para cada puntaje en específico.

Adicionalmente, se sugiere que el programa realice simulacros en cuanto a las pruebas del componente genérico ya que como se dio a conocer en las conclusiones el tomar algún curso de preparación al examen aumenta la probabilidad de obtener buenos resultados, además ayudará a concientizar al programa si lo que evalúa es realmente lo que se esta evaluando en las pruebas Saber Pro.

Dado que en los aspectos demográficos la universidad no podría actuar, si se recomienda que desde el aspecto socioeconómico como lo es tener conexión a internet se aumente la cobertura en este campo, ya que este servicio influye sobre el aumento de la probabilidad de obtener buenos resultados.

2 REFERENTES TEÓRICOS

En los siguientes párrafos se abordan las teorías que dan soporte a esta investigación en cuanto a temas de: evaluación en educación, pruebas genéricas, pruebas Saber Pro, variables demográficas, socioeconómicas, académicas y aspectos del modelamiento estadístico.

2.1 PRUEBAS SABER PRO

La universidad colombiana y en general las instituciones de educación superior se encuentran reguladas bajo la Ley 30 de 1992, la cual define la educación superior como “un servicio público cultural, inherente a la finalidad social del Estado” (MEN, 1992). Esta definición identifica cómo la educación universitaria no es concebida desde el Estado como un bien público otorgado a todos los ciudadanos en términos de equidad, sino como un servicio accesible a quienes demuestren poseer las capacidades requeridas y cumplan con las condiciones académicas exigidas en cada caso (MEN, 1992).

“Una educación de buena calidad juega un importante papel a la hora de crear capital humano y dotar a los profesionales de conocimientos, habilidades y aptitudes que les permita participar en la economía y en la sociedad” (Mundial, 2012). Esta afirmación demuestra el papel tan importante que juegan las instituciones educativas en el desarrollo integral de la sociedad, por tal razón, se puede decir que la preocupación frente a los resultados de las pruebas de Estado es grande, ya que ven reflejado su trabajo como formadores de los futuros líderes del país.

“El sistema de evaluación de las instituciones de educación superior se constituyó a partir de dos ejes: la autoevaluación institucional y las pruebas estandarizadas estatales llamadas inicialmente Examen de calidad de la Educación Superior (ECAES), hoy Pruebas Saber Pro” (Bogoya, 2013).

La primera vez que se definió el examen ECAES como una prueba académica de carácter oficial y obligatorio que forma parte del conjunto de instrumentos del que el Gobierno Nacio-

nal dispone para evaluar la calidad del servicio público educativo fue en el decreto 1781 de 2003. El Examen se divide en dos grandes bloques. El primero es el módulo de competencias genéricas, que son necesarias para el adecuado desempeño profesional, independientemente del programa cursado. Estas evalúan comunicación escrita, razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés. El segundo bloque es el módulo de competencias específicas comunes. Este se ajusta al programa y al enfoque de cada carrera y se construye a partir de un grupo de 35 competencias que de una u otra manera son relevantes para todos y cada uno de los programas de pregrado del país. Las competencias específicas son definidas por el MEN con participación de la comunidad académica y profesional y del sector productivo. Cada programa debe seleccionar los módulos que responderán sus estudiantes para la evaluación de aspectos fundamentales de su formación, pues no se pretende que todos los programas tomen los mismos módulos.

Se aplica a estudiantes de programas de pregrado que estén próximos a culminar su plan de estudios, esto es, que hayan aprobado por lo menos el 75% de los créditos académicos del programa correspondiente o que tengan previsto graduarse en el año siguiente a la fecha de aplicación del examen. Para estos estudiantes, la presentación de los exámenes es obligatoria como requisito de grado, además de los requisitos que cada institución educativa tenga establecidos. Esta prueba evalúa competencias consideradas fundamentales para los futuros egresados de programas de formación de educación superior, y son desarrollados con la participación permanente de las comunidades académicas, redes y asociaciones de facultades y programas. Algunos módulos evalúan competencias genéricas para los estudiantes de cualquier programa de formación, mientras que otros evalúan competencias específicas, comunes a grupos de programas. Estas competencias también son evaluadas en otros países, como es el caso del proyecto de evaluación internacional de competencias en educación superior (AHELO). De igual manera, la prueba Saber Pro puede ser presentada de manera independiente y voluntaria por quienes ya se hayan graduado de los programas académicos, siguiendo los parámetros establecidos por el ICFES para estos casos (Icfes 2012).

De acuerdo con el Decreto 3963 de octubre de 2009, son objetivos del SABER PRO los siguientes:

- Comprobar el grado de desarrollo de las competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior.
- Producir indicadores de valor agregado de la educación superior en relación con el

nivel de competencias de quienes ingresan a este nivel; proporcionar información para la comparación entre programas, instituciones y metodologías, y mostrar su evolución en el tiempo.

- Servir de fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e instituciones de educación superior y del servicio público educativo, que fomenten la cualificación de los procesos institucionales y la formulación de políticas, y soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

El componente genérico de la prueba Saber Pro consta de:

COMUNICACIÓN ESCRITA: “La prueba se enfoca en la competencia para comunicar ideas por escrito con base en el análisis de la información suministrada” (Icfes 2012).

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” (Icfes, 2012).

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” (Icfes, 2012).

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” (Icfes, 2012).

DESEMPEÑO EN 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 cuatro 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 comunicación 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 (Sánchez, 2013).

DESEMPEÑO EN 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 un fin, 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 adecua 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 diálogo 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 (Sánchez, 2013)

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: (Sánchez, 2010).

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

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 mas variables explicativas (modelo), la explicación del modelo en general no puede ser perfecta, existe un error o residual” (Díaz, 2012).

Las etapas del modelamiento estadístico se pueden resumir en:

Especificación: A través del análisis exploratorio de los datos se determinan los supuestos del componente aleatorio. Se establece la función del componente sistemático y se establece 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: Se desea que el modelo represente la realidad lo más aproximada posible.

El modelo más sencillo 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 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).

Un modelo LM se expresa de la siguiente forma:

$$Y_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{nxri} + e_i \quad (2-1)$$

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:

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

$$\mu_i = \beta_0 + \beta_{1x1i} + \dots + \beta_{nxri} + e_i \quad (2-3)$$

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. Según (Pérez, 2016), *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.

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). 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). Tres componentes especifican a un modelo lineal generalizado, (Díaz, 2012):

- ★ 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 la componente aleatoria 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 sustituida por 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:

$$Y \sim \text{ExpF}(\mu, \Phi), \quad (2-4)$$

$$g(\mu) = X\beta, \quad (2-5)$$

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:

$$f_y(y; \mu, \sigma) = \exp \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \quad (2-6)$$

Donde $E(Y) = \mu = b'(\theta)$ y $Var(Y) = \phi v$ siendo $v(\mu) = b''[\theta(\mu)]$.

Según (Pérez, 2016), 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, el 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.

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 (2006) citados por (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:

$$Y \sim \text{ExpF}(\mu, \Phi), \quad (2-7)$$

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j, \quad (2-8)$$

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.

Según (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, aparecieron los modelos GAMLSS.

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

Según (Pérez, 2016), estos modelos hacen referencia a un grupo de modelos estadísticos realizados por Rigby y Stasinopoulos (2001, 2005), 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 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, ver Nelder e Wedderburn (1972) y Hastie y Tibshirani (1990) respectivamente citados por (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 $Var(Y) = \phi v(\mu)$, 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.

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

Según (Pérez 2010), *en los modelos GAMLSS el supuesto que la variables respuesta Y pertenece a la familia exponencial y flexibilizada, permitiendo si trabajar con distribuciones con mayor asimetría o curtosis. Por lo tanto, la distribución de la variable respuesta Y puede ser seleccionada 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.*

Un modelo GAMLSS puede escribirse de la siguiente forma:

$$g(\mu) = X\beta + \sum_{j=1}^J h_j X_j \quad (2-9)$$

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:

$$g_1(\mu) = X_1\beta_1 + \sum_{j=1}^{J_1} h_{j1} X_{j1} \quad (2-10)$$

$$g_2(\sigma) = X_2\beta_2 + \sum_{j=1}^{J_2} h_{j2} X_{j2} \quad (2-11)$$

$$g_3(\nu) = X_3\beta_3 + \sum_{j=1}^{J_3} h_{j3}X_{j3} \quad (2-12)$$

$$g_4(\tau) = X_4\beta_4 + \sum_{j=1}^{J_4} h_{j4}X_{j4} \quad (2-13)$$

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, ν, τ respectivamente. Si se quiere generalizar un poco la notación utilizada podemos escribir:

$$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{J_k} h_{jk}X_{jk} \quad (2-14)$$

donde θ_k es el vector de parámetros de la distribución.

El modelo GAMLSS queda descrito de la siguiente forma:

$$y \sim D(\theta_k) \quad (2-15)$$

$$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\gamma_{jk} \quad (2-16)$$

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_k es un vector de parámetros de longitud J_k y γ_{jk} es una variable aleatoria q_{jk} dimensional. La estimación de los parámetros se lleva a cabo maximizando la siguiente función de verosimilitud:

$$L_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} G_{jk} \gamma_{jk} \quad (2-17)$$

donde $l = \sum_{i=1}^n \log f(y_i | \theta_i)$ es un logaritmo 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 $M = D, G, T, \wedge$

un modelo GAMLSS donde:

- * D = específica a distribución de la variable respuesta
- * G = específica el conjunto de funciones enlace (g_1, \dots, g_p) para los parámetros $(1, \dots, p)$
- * T = engloba a las variables utilizadas en los predictores correspondientes a μ, σ, ν y τ
- * \wedge = 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 . Ahora 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 quedándose con aquella que proporcione un modelo con menor criterio de informacin de Akaike generalizado (GAIC). 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 (se escogera el de 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 \wedge : Para cada termino suavizado se requiere 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 debemos tratar de escoger uno entre todos los propuestos. A la hora de escoger el modelo más acertado podemos utilizar el criterio criterio de informacin de Akaike (AIC).

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

Los residuos cuantiles aleatorizados vienen dados por la expresión $r_i = \Phi^{-1}(u_i)$ donde Φ^{-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.

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

La función `plot.gamlss()` representa graficamente 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.

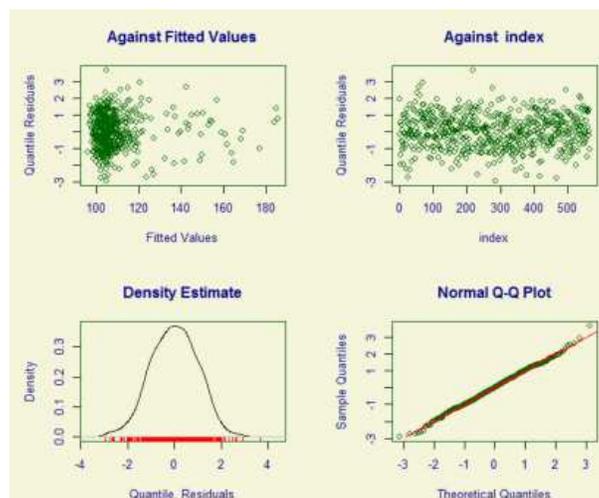


Figura 2-1: Función `plot.gamlss()`

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 y Fredrils (2001) citados por (Pérez 2010) 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 se puede ver que tan lejos están unos de los otros.

- 2: las líneas discontinuas negras marcan el intervalo de confianza de el 95%, por lo tanto para considerar que el modelo es correcto, sólo un 5% de los puntos podrían quedar fuera de este intervalo. Si esto no es así se considera el modelo inadecuado para explicar la variable respuesta.
- 3: La curva continua horizontal es un ajuste cúbico de los puntos y puede indicarnos diferentes problemas en el modelo tomado de Stasinopoulos et al. (2015c) citado por (Pérez 2010).

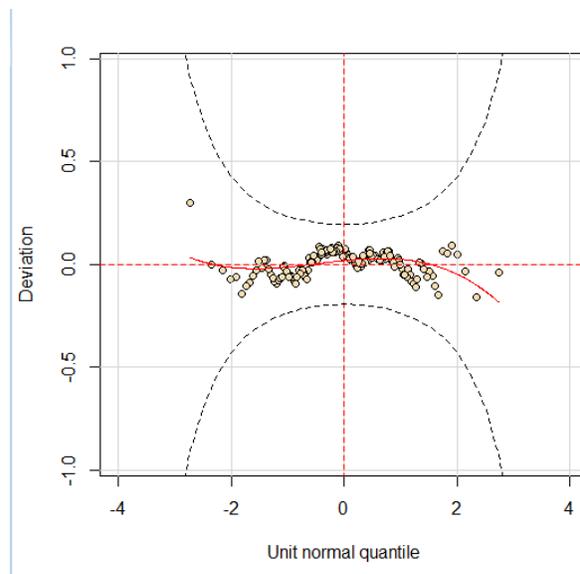


Figura 2-2: Función $w_p()$

SELECCIÓN DEL MODELO

Una vez ajustado el modelo debemos hacer una buena selección, una de las funciones de ayuda para una buena selección es la siguiente:

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 , ν y τ .
2. Dado el modelo para μ obtenido en 1 y para ν y τ fijas, úsese 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 utilizó un proceso de selección hacia adelante para escoger un modelo apropiado para ν .
- 4. Dados los modelos para μ , δ y ν obtenidos en 1, 2 y 3 respectivamente, se utilizó 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 utilizó 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 e 4 respectivamente, se utilizó 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 utilizó un proceso de selección hacia atrás para escoger un modelo apropiado para μ y finalizamos el proceso.

ALGUNAS DISTRIBUCIONES CONTINUAS PARA LOS GAMLSS

Los ajustes de la variable respuesta dieron como resultado lo que se presenta en la tabla 1, la cual muestra la distribución y sus correspondientes funciones de enlace (Pérez, 2016):

Tabla 1 Distribuciones y funciones de enlace

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	δ	ν	τ
Logistic LO()	Razonamiento Cuantitativo	identity	log	-	-
Gamma GA()	Lectura Crítica	log	log	-	-
Gamma Generalizada GG()	Comunicación Escrita	log	log	identity	-
skew t type 3 ST3()	Inglés	identity	log	log	log

Distribución Logistic (LO)

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

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

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 Gamma (GA)

La distribución Gamma, denotada por $GA(\mu, \delta)$, está definida por:

$$f_Y(y|\mu, \delta) = \frac{1}{(\delta^2\mu)^{\frac{1}{\delta^2}}} \frac{y^{1-\frac{1}{\delta^2}} \exp\left(\frac{-y}{\delta^2\mu}\right)}{\Gamma\left(\frac{1}{\delta^2}\right)} \quad (2-19)$$

para $y > 0$, donde $\mu > 0$ y $\delta > 0$. Su $E(Y) = \mu$ y $Var(Y) = \delta^2\mu^2$. Esta es una reparametrización de Johnson et al. (1994) p 343 ecuación (17.23) obtenida a través de $\delta^2 = \frac{1}{\alpha}$ y $\mu = \alpha\beta$. Con función de enlace $\mu = e^{X\beta}$ y $\delta^2 = e^{X\beta}$

Distribución Gamma Generalizada (GG)

La distribución Gamma Generalizada, denotada por $GG(\mu, \delta, \nu)$, fue usada por Lopatzidis y Green (2000), y está definida como:

$$f_Y(y|\mu, \delta, \nu) = \frac{|\nu|\theta^\theta z^\theta \exp^{-\theta z}}{\Gamma(\theta)y} \quad (2-20)$$

para $y > 0$, donde $\mu > 0$, $\delta > 0$ y $-\infty < \nu < \infty$ y donde $z = \frac{y}{\mu} \nu$ y $\theta = \frac{1}{\delta^2\nu^2}$. La media y la varianza de Y están dadas por $E(Y) = \frac{\mu\Gamma(\theta + \frac{1}{\nu})}{[\theta^{\frac{1}{\nu}}\Gamma(\theta)]}$ y $Var(Y) = \frac{\mu^2\{\Gamma(\theta)\Gamma(\theta + \frac{2}{\nu}) - [\Gamma(\theta + \frac{1}{\nu})]^2\}}{[\theta^{\frac{2}{\nu}}\Gamma(\theta)]^2}$

Con función de enlace $\mu = \frac{[\theta^{\frac{1}{\nu}}\Gamma(\theta)]e^{X\beta}}{\Gamma(\theta + \frac{1}{\nu})}$

Distribución Skew t type 3 (ST3)

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

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

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 $Var(Y) = \delta^2 V(Z)$. 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 REGRESIÓN LOGÍSTICA ORDINAL (RLO)

“La Regresión Logística es una técnica estadística multivariante que nos permite estimar la relación existente entre una variable dependiente no métrica, en particular dicotómica y un conjunto de variables independientes métricas o no métricas. 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 (Nelder and Wedderburn [1972]), al igual que los modelos de regresión lineal o el análisis de la varianza (Iglesias, 2013).

El modelo que se usa comúnmente cuando se tienen categorías ordinales es el modelo de odds proporcionales (Dobson 2002). 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 citados por (Agresti 2002).

En la RLO las funciones de enlace más empleadas son la Logit y la Cloglog (Agresti 2002). 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. 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 el

enlace Logit como la más satisfactoria para este caso.

La expresión de la función Logit para la RLO es la siguiente:

$$\ln(O_i) = \alpha_i + \beta X, \quad (2-22)$$

en esta ecuación, O_i es la razón de probabilidad (odds) asociada a la categoría i de la variable dependiente, siendo la expresión de esta razón:

$$O_i = \frac{\frac{P(\text{valor sea} \leq \text{categoría } i)}{\text{valores de } X}}{1 - \frac{P(\text{valor sea} < \text{categoría } i)}{\text{valores de } X}} \quad (2-23)$$

Con el término valor en la ecuación (2-23) se hace referencia a cualquier valor de la variable dependiente. Como se observa, las probabilidades de la ecuación (2-23) es condicional, es decir, dados los valores de la (s) variable(s) independiente(s). En la ecuación (2-22), 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 ésta igual a uno. El modelo de regresión ordinal descrito anteriormente es denominado modelo logit acumulado, ya que es construido basándose en las probabilidades acumuladas de la variable respuesta, dados los valores de las variables explicativas (Hosmer-Lemeshow, 2000). 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 (Agresti, 2002). Por tanto existen ecuaciones $k - 1$ con los mismos coeficientes acompañando a las variables explicativas y que sólo se diferencian en el valor del intercepto.

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

$$G = \frac{-2 \ln(\text{probabilidad de la muestra sin variables})}{\text{probabilidad de la muestra con variables}} \quad (2-24)$$

Este estadístico sigue distribución Ji-cuadrado, con tantos grados de libertad como variables independientes existan (McCullagh, 1990) citado por Agresti. Cuando las predicciones de la variable dependiente que se hacen con el modelo que incluye todas las variables independientes superan las predicciones que se realizan sin considerar éstas, el valor de G tiende a ser grande y conlleva a concluir que al menos una de las variables regresoras tiene efecto sobre la variable dependiente, y, por lo tanto, que la probabilidad de ocurrencia de los valores que representan esta variable varía para alguna de las combinaciones de valores de las variables independientes (Agresti, 2002). 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 (Agresti, 2002). El estadístico utilizado es el siguiente:

$$ZWald = b_j ES(b_j) \quad (2-25)$$

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 [(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. Para comparar estas frecuencias generalmente se calcula el estadístico Ji-cuadrado de Pearson según la fórmula (Agresti, 2002)

$$X^2 = \sum_{i=1}^k \sum_{l=1}^m \frac{(y_{il} - m p_{il})^2}{m p_{il}(1 - p_{il})} \quad (2-26)$$

En la expresión (2-26) 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.

ml 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:

$$\frac{\frac{P(\text{valor sea } \leq \text{categoría } i)}{\text{valores de } X}}{1 - \frac{P(\text{valor sea } < \text{categoría } i)}{\text{valores de } X}} = \ln(O_i) = \alpha_i + \beta X \quad (2-27)$$

Con el término “valor” en la ecuación (2-27) se hace referencia a cualquier valor de la variable dependiente. De (2-27) se deriva que:

$$\frac{P(\text{valor sea } \leq \text{categoría } i)}{\text{valores de } X} = \frac{e^{\alpha_i + \beta X}}{1 + e^{\alpha_i + \beta X}} \quad (2-28)$$

La expresión (2-28) 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 la razón de probabilidad (odds ratio) que provoca el cambio en cada una de las variables independientes. La razón de odds 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:

$$\text{odds-ratio} = \frac{\frac{P(Y \leq ij | X = X_2)}{P(Y \geq ij | X = X_2)}}{\frac{P(Y \leq ij | X = X_1)}{P(Y \geq ij | X = X_1)}} \quad (2-29)$$

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 una sola razón de odds. La razón de odds es utilizada para interpretar el efecto de las variables explicativas sobre la variable objeto de estudio. Si éste es igual a uno, indica que la variable predictora no tiene efecto. Si es menor que uno, lo cual sucede cuando el coeficiente de la variable regresora es negativo, indica que, si las otras variables explicativas permanecen constantes, los cambios en la variable explicativa analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable objeto de estudio (Agresti, 2002). 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 aplicada es de carácter descriptivo, inferencial para los resultados de la prueba saber Pro, con enfoque cuantitativo. Se contó con una población de 199 estudiantes del programa Diseño Industrial UPTC 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 199 filas por 58 columnas: <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados>.

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” (Camarero, 2013). 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 distintas categorías clasificatorias); 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, lugar de residencia, estado civil etc.

Académicas: A esta categoría pertenecen aquellas variables asociadas al rendimiento académico previo y al esfuerzo mostrado por el estudiante. Así, abarcan aquellos factores educativos que rodean directamente la intervención pedagógica o docente.

Socioeconómicas: A esta categoría pertenecen aquellas variables como estrato socioeconómico, forma de pago de la matrícula, ocupación de los padres entre otras. A continuación se exponen las variables que se consideraron en este estudio.

Tabla 2 Variables demográficas

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

Tabla 3 Variables socioeconómicas

VARIABLES		CATEGORÍAS	ESCALA DE MEDICIÓN
Ocupación padre, Ocupación 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 tabla 3

VARIABLE		CATEGORÍAS	ESCALA DE MEDICIÓN
El hogar cuenta con servicio de teléfono fijo	No	Valores posibles	Cualitativa Nominal
	Si	0	
		1	
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	
		6	
	10 o ms SM	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	
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

Tabla 4 Variables académicas

VARIABLE	CATEGORÍAS	ESCALA DE MEDICIÓN
Máximo nivel educativo alcanzado 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 METODOS DE ANÁLISIS Y PROCESAMIENTO DE LOS DATOS

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 3.3.1, 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 en las Pruebas Saber Pro, las cuales se dividen en 2 bloques: competencias específicas que evalúan los aspectos fundamentales de su formación profesional, y las competencias genéricas son las que se consideran debe desarrollar todo estudiante, independiente de su formación profesional, se evalúa inglés, razonamiento cuantitativo, lectura crítica, comunicación escrita y competencias ciudadanas. En primer lugar, se describen 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 4.1, 4.2, 4.4 y 4.5 fue tomada de la página <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados>. La información que sustenta la sección 4.5 se presenta en el Anexo A- Resultados Diseño Industrial y la información que sustenta la sección 4.3 se extrajo de las bases de datos que proporciona el ICFES para desarrollar investigaciones.

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

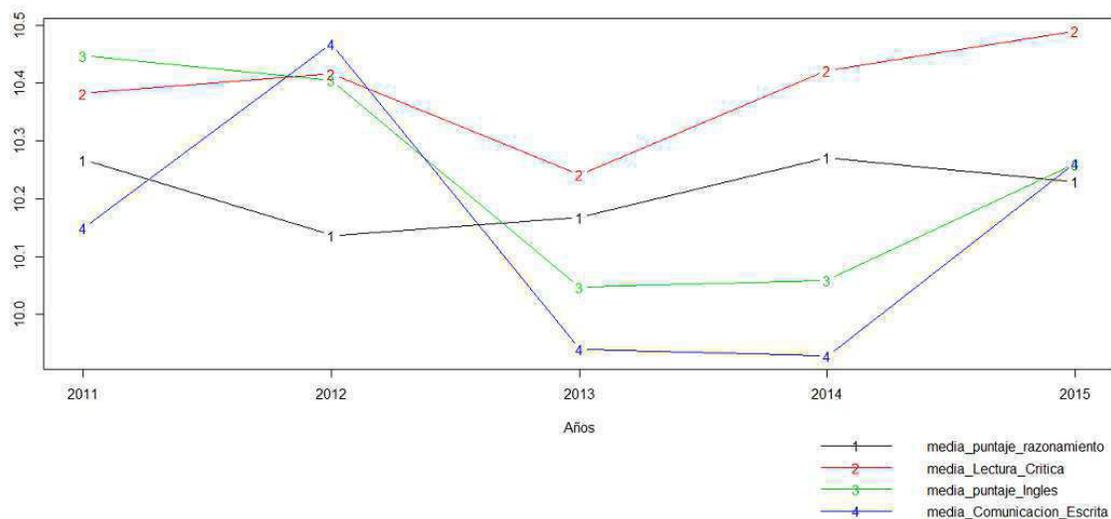


Figura 4-1: Resultados promedio por año y competencia

Para la descripción se tendrá en cuenta la guía de interpretación de resultados del Ministerio de Educación Nacional y del Icfes, como se estableció en la sección 2.2.

Podemos observar que durante los años 2011 a 2015 el puntaje para razonamiento cuantitativo ha permanecido estable, con un ligero cambio en 2014 mostrando un incremento con relación a los demás años. En cuanto a lectura crítica es una de los componentes que obtuvo mayores puntajes promedio a través de los años, lo cual nos indica que los estudiantes tienen la capacidad para leer de manera analítica y reflexiva, comprenden los planteamientos expuestos en un texto identifican y recuperan información, construyen un sentido global y establecen relaciones entre enunciados. Por otra parte, 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 la información suministrada, así como también el puntaje en inglés, que con el pasar de los años fue descendiendo, salvo el año 2015 donde hubo un incremento importante.

4.1 DESCRIPCIÓN UNIVARIADA POR COMPONENTE

Los resultados del componente genérico correspondiente a Diseño Industrial de la UPTC se presentan en una tabla, la cual incluye los resultados correspondientes al Programa, n , el número de estudiantes que presentaron la prueba, \bar{X}_n , puntaje promedio individual obtenido, 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. Para el análisis no se tuvo en cuenta la observación 157 debido a que el estudiante no presenta resultados en el componente genérico.

La Tabla 5 muestra la descripción por año de los resultados obtenidos por los estudiantes de Diseño Industrial para el componente en comunicación escrita.

Tabla 5 Resultados comunicación escrita

Comunicación escrita	Resumen numérico									
	Año	n	\bar{X}_n	CA	CC	CV	Q_1	Q_2	Q_3	NA
	2011	34	10.15	0.127	-0.615	0.098	9.4	10	10	0
	2012	19	10.46	-0.066	-1.131	0.104	9.45	10.7	11.25	0
	2013	57	9.94	-0.016	-0.366	0.107	9.3	9.9	10.6	0
	2014	42	9.93	0.0482	-0.328	0.107	9.25	9.95	10.5	0
	2015	44	10.26	-0.397	2.351	0.088	9.8	10.3	10.9	3

Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Diseño Industrial son muy homogéneos respecto a los puntajes (coeficientes de variación, del orden del 10%). A juzgar por los bajos valores de los coeficientes de curtosis los puntajes en comunicación escrita indican que no hay estudiantes que hayan obtenido puntajes muy diferentes, es decir, los resultados se concentran alrededor de su media para cada año y teniendo en cuenta que el promedio es de 10.2 podemos concluir que a través de los años se ubica en la categoría regular, lo cual apunta a que los estudiantes de Diseño Industrial 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 comunicación escrita, tabla 6.

Tabla 6 Resultados desempeño comunicación escrita

Desempeño comunicación escrita	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
2011	0%	14%	41%	21%	21%	3%
2012	0%	21%	16%	32%	21%	10%
2013	11%	11%	37%	30%	8%	3%
2014	2%	24%	26%	34%	12%	2%
2015	2%	5%	34%	48%	9%	2%

Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes se ubican en el nivel 5 de desempeño, indicando que el estudiante 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, existe buen uso del lenguaje, aunque pueden encontrarse errores en la aplicación de algunas reglas de ortografía y puntuación.

La siguiente Tabla muestra la descripción por año de los resultados obtenidos por los estudiantes de Diseño Industrial para el componente en inglés.

Tabla 7 resultados inglés

Inglés	Resumen numérico								
Año	n	\bar{X}_n	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	34	10.447	0.956	1.0681	0.090	9.8	10.25	10.8	0
2012	19	10.40	0.079	-0.774	0.079	9.85	10.3	10.9	0
2013	57	10.04	-4.259	26.93	0.160	9.5	10.1	10.5	0
2014	42	10.05	-3.579	19.47	0.188	9.6	10	10.82	0
2015	46	10.26	1.159	1.403	0.075	9.73	10.06	10.31	1

Se observa que, para los años 2011, 2012 y 2015 los estudiantes de Diseño Industrial son muy homogéneos respecto a los puntajes (coeficientes de variación, del orden del 10%), excepto en los años 2013 y 2014, donde se observa una gran heterogeneidad en los puntajes. Dados los valores de curtosis para los años 2011, 2012 y 2015, el puntaje en inglés indica que no hay estudiantes que hayan obtenido puntajes muy diferentes al resto del grupo. En cuanto a 2013 y 2014 se observa una curtosis mayor a 7, esto indica presencia de datos atípicos es decir que existen puntajes extraordinarios en comparación con el grupo. En general se puede concluir que, dados los valores de la mediana para cada año, el desempeño en inglés fue bueno. Los resultados anteriores se pueden complementar con los resultados cualitativos en inglés, Tabla 8.

Tabla 8 Resultados desempeño inglés

Desempeño Inglés	A-	A1	A2	B+	B1
2011	12 %	26 %	26 %	3 %	33 %
2012	11 %	21 %	26 %	0 %	42 %
2013	17 %	28 %	30 %	2 %	23 %
2014	17 %	36 %	19 %	9 %	19 %
2015	13 %	28 %	38 %	4 %	17 %

Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayoría de estudiantes se ubican en el nivel B1 de desempeño, indicando que el estudiante es capaz de comprender los puntos principales de textos si tratan sobre cuestiones que le son conocidas, ya sean en situaciones de trabajo, de estudio o de ocio.

La siguiente Tabla muestra la descripción por año de los resultados obtenidos por los estudiantes de Diseño Industrial para componente en lectura crítica.

Tabla 9 Resultados lectura crítica

Lectura crítica	Resumen numérico								
Año	n	\bar{X}_n	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	34	10.38	0.662	0.248	0.070	9.8	10.3	10.8	0
2012	10	10.41	-1.351	2.088	0.074	10.2	10.5	10.8	0
2013	57	10.24	0.135	-0.176	0.065	9.8	10.2	10.7	0
2014	42	10.42	-0.255	-0.607	0.068	9.82	10.6	10.8	0
2015	46	10.49	0.060	0.179	0.069	9.92	10.55	11	1

Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Diseño Industrial son muy homogéneos respecto a los puntajes (coeficientes de variación, menos del 10%). A juzgar por los bajos valores de los coeficientes de curtosis los puntajes en lectura crítica indican que no hay estudiantes que hayan obtenido puntajes muy diferentes al resto del grupo, es decir, los resultados se concentran alrededor de su media para cada año y teniendo en cuenta que el promedio es de 10.4 podemos concluir que a través de los años se ubica en la categoría bueno, lo cual apunta a que los estudiantes de Diseño Industrial tienen la capacidad para leer de manera analítica y reflexiva, además comprenden los planteamientos expuestos en un texto identificando y recuperando información presente en uno o varios textos, construyen un sentido global, y establecen relaciones entre enunciados y evalúan su intencionalidad.

La siguiente Tabla muestra la descripción por año de los resultados obtenidos por los estudiantes de Diseño industrial para el componente en razonamiento cuantitativo.

Tabla 10 Resultados razonamiento cuantitativo

Razonamiento cuantitativo	Resumen numérico								
Año	n	\bar{X}_n	CA	CC	CV	Q_1	Q_2	Q_3	NA
2011	34	10.26	-0.846	1.093	0.080	9.85	10.4	10.7	0
2012	19	10.13	-0.823	0.401	0.086	9.6	10.3	10.7	0
2013	57	10.16	-0.015	2.438	0.069	9.8	10.2	10.5	0
2014	42	10.27	0.370	-0.068	0.090	9.6	10.1	10.87	0
2015	46	10.23	0.389	0.189	0.088	9.6	10.15	10.8	1

Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Diseño Industrial son muy homogéneos respecto a los puntajes (coeficientes de variación, menos del 10%). A juzgar por los bajos valores de los coeficientes de curtosis los puntajes en razonamiento cuantitativo indican que no hay estudiantes que hayan obtenido puntajes muy diferentes al resto del grupo, es decir, los resultados se concentran alrededor de su media para cada año y teniendo en cuenta que el promedio es de 10.2 podemos concluir que a través de los años se ubican en la categoría regular, lo cual apunta a que los estudiantes no tengan competencias relacionadas o habilidades en la comprensión de conceptos básicos de las matemáticas para analizar, modelar y resolver problemas aplicando métodos y procedimientos cuantitativos y esquemáticos.

4.2 RESULTADOS SABER PRO AÑO 2016

Para el año 2016, el puntaje global del programa fue de 157 puntos con una desviación estándar de 17 puntos, esto implica que los estudiantes de Diseño Industrial fueron muy homogéneos (coeficiente de variación del 11%) en la prueba y que al menos el 75% de ellos obtuvo entre 123 y 191 puntos. La media en el grupo de referencia Bellas Artes del País fue de 157 puntos con una desviación de 23. Es decir, al menos el 75% de los estudiantes obtuvo entre 111 y 203 puntos. Lo cual implica que Diseño Industrial tiene un comportamiento similar al grupo de referencia. El puntaje global de Diseño Industrial fue igual con el de la Institución y el de la Sede.

Se tiene que en comunicación escrita el puntaje del Programa (140 (ds=24)) estuvo por debajo de los puntajes de la Institución, la Sede y el grupo de referencia, la mayoría de estudiantes del Programa (50%) se ubica en nivel de desempeño dos (Icfes, 2017), lo que indica que tan sólo superan las preguntas de menor complejidad del examen y los ubica en un uso aceptable del lenguaje, identificando un planteamiento o posición personal que va más allá de una opinión aislada sobre el tema, sin embargo el texto presentado puede carecer de unidad semántica o presencia de contradicciones, digresiones y repeticiones.

En razonamiento cuantitativo, el Programa fue superior (167(ds=23)) respecto al grupo de referencia y a la Institución, y similar con los programas de la Sede. La mayoría de estudiantes del Programa (69%) se ubica en nivel de desempeño tres (Icfes, 2017), lo que indica que muestran un desempeño adecuado en las competencias exigibles para el examen. Este es el nivel esperado que todos o la gran mayoría de los estudiantes deberían alcanzar. Indicando que los estudiantes 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. Esto indica que las acciones de mejoramiento del Programa se deben enfocar en identificar las dificultades a la hora de validar procedimientos y estrategias matemáticas para dar solución a problemas.

En Lectura crítica, competencias ciudadanas e Inglés el rendimiento del programa fue similar con los de la Institución, la Sede y el País. En lectura crítica la mayoría de estudiantes del Programa (53 %) se ubica en nivel de desempeño tres (Icfes, 2017), lo cual indica que interpretan el texto más allá de la forma y el contenido para comparar, asociar, relacionar, jerarquizar, analizar, comprobar, sintetizar y contextualizar la información local y general del texto, imaginan situaciones hipotéticas a partir de la información e idean situaciones o escritos a partir del texto.

En competencias ciudadanas la mayoría de estudiantes del Programa (57 %) se ubica en nivel de desempeño 3, es decir, muestran un desempeño adecuado en las competencias exigibles para el examen. Este es el nivel esperado que todos o la gran mayoría de los estudiantes deberían alcanzar y muestra que los estudiantes identifican argumentos implícitos en un enunciado o afirmación y establece la relación entre una afirmación y una cosmovisión, también analiza y compara propuestas de solución a un problema. Esto indica que las acciones de mejoramiento del Programa se deben enfocar en identificar las dificultades asociadas a conocer la organización del estado de acuerdo con la Constitución Política de Colombia.

En inglés la mayoría de estudiantes del Programa (46 %) se ubica en nivel de desempeño B1, es decir, el estudiante 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.

4.3 DESCRIPCIÓN POBLACIONAL DE ESTUDIO

A continuación se presenta la descripción de cada variable explicativa tenida en cuenta para este trabajo monográfico.

Tabla 11 Factores de estudio demográficos

VARIABLES		CATEGORÍAS	RESUMEN
sexo del inscrito	Femenino	F	F =44 %
	Masculino	M	M =56 %
Edad del estudiante		rango:[20-39]	n =198 $\bar{X}_n = 26.69$ sd =3.123 CV =0.116 CC = 1.941 $Q_1 = 24$ $Q_2 = 26$ $Q_3 = 28$ NA= 1
Estado civil del estudiante	Soltero	1	1 =94 %
	Casado	2	2 =3 %
	Viudo	3	3 =0 %
	Separado	4	4 =2 %
	Unión libre	5	5 =2 %
Situación de su hogar actual		Valores posibles	
Es habitual o permanente		1	1 =62 %
Es temporal por razones de estudio u otra razón		2	2 =38 %
Número de personas de las que usted se encuentra a cargo		rango: [0,5]	n =199 $\bar{X}_n = 0.150$ sd =0.548 CV =3.635 CC = 19.03 $Q_1 = 0$ $Q_2 = 0$ $Q_3 = 0$ NA= 0
Es cabeza de familia	No	0	0=94 %
	Si	1	1=5 %
Número de personas que conforman el hogar		rango: [1,12]	n =199 $\bar{X}_n = 4.065$ sd =1.735 CV =0.426 CC = 3.487 $Q_1 = 3$ $Q_2 = 4$ $Q_3 = 5$ NA= 0

En la población objeto de análisis se encontró que tienen un promedio de edad de 27 años de edad, cuyo estado civil predominante es soltero (94%), además de ser en su mayoría de

género masculino (56%). Otros detalles relevantes muestran que no son cabeza de familia (94%), y no tienen personas a cargo, su núcleo familiar está conformado en promedio por 4 personas, viviendo por lo general en un hogar permanente (62%).

Tabla 12 Factores de estudio socioeconómicos

VARIABLE	CATEGORÍAS	RESUMEN
Estrato de la residencia según factura de energía	Valores posibles	1=7%
Estrato 1	1	2=49%
Estrato 2	2	3=39%
Estrato 3	3	4=4%
Estrato 4	4	5=1%
Estrato 5	5	6=0%
Estrato 6	6	
Zona rural donde no hay estratificación económica		
Valor anual de la matrícula del año anterior	valores posibles	0=4%
No pagó matrícula	0	1=7%
Menos de 500 mil	1	2=1%
Entre 500 mil y menos de 1 millón	2	3=85%
Entre 1 millón y 3 millones	3	4=4%
Entre 3 millones y 5 millones	4	5=0%
Más de 5 millones	5	
Forma de pagó matrícula	valores posibles	0=49%
pagó por padres	0	1=19%
Varias formas de pagó	1	2=5%
No cancelo matrícula	2	3=15%
pagó por beca	3	4=7%
pagó por credito	4	5=6%
pagó propio	5	
El hogar cuenta con lavadora	Valores posibles	0=29%
No	0	1=71%
Si	1	
El hogar cuenta con celular	Valores posibles	0=21%
No	0	1=79%
Si	1	
El hogar cuenta con conexión a internet	Valores posibles	0=44%
No	0	1=56%
Si	1	
El hogar cuenta con servicio cerrado de televisión	Valores posibles	0=32%
No	0	1=68%
Si	1	
El hogar cuenta con servicio de teléfono fijo	Valores posibles	0=51%
No	0	1=49%
Si	1	
Trabaja actualmente	Valores posibles	0=50%
No	0	1=0%
Si, con remuneración en dinero o en especie	1	3=32%
Si, como ayudante sin remuneración	3	4=1%
Si, por ser práctica obligatoria del plan de estudios	4	5=17%
Si, por experiencia o recursos en gastos personales	5	

Continuación tabla 12.

VARIABLE	CATEGORÍAS	RESUMEN	
Ocupación del padre; madre	Valores posibles	1=0 %	1=0 %
Empresarios	1	2=0 %	2=0 %
Administradores o gerentes	2	3=0 %	3=0 %
Profesionales independientes	3	4=0 %	4=0 %
Profesionales empleados	4	5=0 %	5=0 %
Trabajadores independientes	5	6=0 %	6=0 %
Trabajadores empleados	6	7=0 %	7=0 %
Rentistas	7	8=0 %	8=0 %
Obreros	8	9=0 %	9=0 %
Jubilados	9	10=0 %	10=0 %
Hogar	10	11=0 %	11=0 %
Estudiantes	11	12=0 %	12=0 %
No devengan ingreso o buscan trabajo	12	13=2 %	13=0 %
Empresario	13	14=11 %	14=7 %
Pequeño empresario	14	15=3 %	15=3 %
Empleado con cargo como director o gerente	15	16=6 %	16=7 %
Empleado de nivel directivo	16	17=11 %	17=8 %
Empleado de nivel técnico o profesional	17	18=2 %	18=6 %
Empleado de nivel auxiliar o administrativo	18	19=13 %	19=6 %
Empleado obrero u operario	19	20=6 %	20=4 %
Profesional Independiente	20	21=20 %	21=16 %
Trabajador por cuenta propia	21	22=3 %	22=30 %
Hogar	22	23=14 %	23=7 %
Pensionado	23	24=0 %	24=0 %
Rentista	24	25=0 %	25=0 %
Estudiante	25	26=11 %	26=8 %
Otra actividad u ocupación	26	99=0 %	99=0 %
No sabe	99		
Número de dormitorios de la residencia	rango: [1,10]	n =199	
		$\bar{X}_n = 2.939$	
		sd =1.003	
		CV =0.341	
		CC = 2.795	
		$Q_1 = 2$	
		$Q_2 = 3$	
		$Q_3 = 3$	
		NA= 0	
Ingresos mensuales en salarios mínimos	Valores posibles	1=13 %	
Menos de 1 SM	1	2=36 %	
Entre 1 y Menos de 2 SM	2	3=30 %	
Entre 2 y Menos de 3 SM	3	4=17 %	
Entre 3 y Menos de 5 SM	4	5=5 %	
Entre 5 y Menos de 7 SM	5	6=1 %	
Entre 7 y Menos de 10 SM	6	7 =0 %	
10 o más SM	7		
Horas que trabaja a la semana	Valores posibles	1=50 %	
No trabaja	1	2=50 %	
De 1 a 5 horas	2	3=17 %	
De 6 a 10 horas	3	4=8 %	
De 11 a 15 horas	4	5=19 %	
De 16 a 20 horas	5		

En la población objeto de análisis se encontró que son de estrato dos (49%) y tres (39%), cuyo valor de matrícula predominante es entre 1 millón y 3 millones de pesos (85%) y en el 50% de los estudiantes la forma de pagó es a través de los padres. La ocupación del padre y la madre se presentan en la modalidad de trabajo por cuenta propia 20% y 16% respectivamente, seguido del hogar con un 30% para el caso de las mujeres, en el caso de los hombres la modalidad de pensión y empleado obrero u operario tienen un 14% y 13% respectivamente. La modalidad de empleado de nivel técnico o profesional y otra actividad u ocupación en ambos géneros presentan un 11% y 8% respectivamente, para las demás modalidades de ocupación su porcentaje es menor al 7%. Las condiciones socioeconómicas de los estudiantes en su mayoría tienen servicio cerrado de televisión (68%), lavadora (71%), internet (56%) y celular (79%), no cuentan con servicio de teléfono fijo (51%). Se puede determinar que un 50% de los estudiantes no trabajan pero a excepción de algunos estudiantes el tiempo de trabajo está entre 16 a 20 horas semanales, en promedio tiene 3 dormitorios en el hogar. En cuanto a los ingresos familiares mensuales se observa entre 1 y menos de 2 salarios mínimos con un (36%).

Tabla 13 Factores de estudio académicos

VARIABLES		CATEGORÍAS	RESUMEN	
Tomó algún curso de preparación al examen		valores posibles		
No		0	0=30%	
Si		1	1=70%	
Bachillerato del que se graduó		valores posibles		
Académico		A	A=57%	
Técnico		N	N=3%	
Normalista Superior		T	T=40%	
Desconocido		D	D=0%	
Nivel educativo del padre; madre		Valores posibles		
Ninguno		0	0=3%	0=1%
No tuvo escuela		1	1=0%	1=0%
Preescolar		2	2=0%	2=0%
Básica primaria		3	3=0%	3=0%
Básica secundaria		4	4=0%	4=0%
Media vocacional		5	5=0%	5=0%
Tecnológico o técnico		6	6=0%	6=0%
Universitario		7	7=0%	7=0%
Postgrado		8	8=0%	8=0%
Primaria incompleta		9	9=13%	9=12%
Primaria completa		10	10=10%	10=6%
Secundaria (bachillerato) incompleta		11	11=15%	11=15%
Secundaria (bachillerato) completa		12	12=14%	12=24%
Educación técnica o tecnológica incompleta		13	13=4%	13=3%
Educación técnica o tecnológica completa		14	14=13%	14=9%
Educación profesional incompleta		15	15=4%	15=3%
Educación profesional completa		16	16=15%	16=18%
Postgrado		17	17=9%	17=10%
No sabe		99	99=0%	99=1%

En la población objeto de análisis se encontró que en su mayoría presentan algún tipo de preparación previo para las pruebas Saber Pro (70%) y el tipo de bachillerato es académico (57%). El máximo nivel académico alcanzado por los padres de secundaria y educación profesional es el más común.

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 que sustentan las afirmaciones se discriminan en el Anexo A.

Se consideró rechazar hipótesis nulas teniendo en cuenta un p-valor menor o igual a 0.05. Al relacionar los puntajes obtenidos en los módulos del componente genérico y algunas características de los estudiantes se encontró correlación negativa entre el puntaje en inglés y el número de personas que conforman su hogar ($p=0.002$). De igual forma se observó una asociación en cuanto al desempeño en comunicación escrita con educación del padre ($p=0.025$).

Para variables de tipo cuantitativo, se realizó la prueba de normalidad de Shapiro Wilk prueba que se presenta en el Anexo A de cada variable respuesta, dando como resultado que los puntajes en razonamiento cuantitativo, lectura crítica y comunicación escrita se distribuyen como una normal. En cuanto al puntaje en inglés, este no se ajusta a una distribución normal, de este modo, para aquellas variables con distribución normal se realizó el test de varianzas de Levens. Se encontró que respecto a razonamiento cuantitativo hay diferencia entre los promedios según género ($p=0.002$), ser cabeza de familia ($p=0.007$) y tener conexión a internet ($p=0.006$) como se muestra en la Tabla 14.

Tabla 14 Razonamiento cuantitativo.

	MEDIA
VARIABLES	
Hombre	10.37
Mujer	10.02
acceso conexión internet (S)	10.36
acceso conexión internet (No)	10.03
Cabeza de familia (S)	9.92
Cabeza de familia (No)	10.23

Para el puntaje en inglés se realizó el test no paramétrico de Kruskal Wallis, encontrándose asociación entre el puntaje en inglés y tener conexión a internet ($p=0.014$) como se muestra en la Tabla 15.

Tabla 15 Inglés.

VARIABLES	MEDIA
Acceso conexión internet (Sí)	10.3
Acceso conexión internet (No)	10.0

Asimismo se observó diferencia entre los promedios en comunicación escrita según estado civil ($p=0.012$), ser cabeza de familia ($p=0.048$), tener conexión a internet ($p=0.056$), educación del padre ($p=0.015$) y educación de la madre ($p=0.001$) como se muestra en la Tabla 16.

Tabla 16 Comunicación escrita.

VARIABLES	MEDIA
Acceso conexión internet (Sí)	10.22
Acceso conexión internet (No)	9.94
Cabeza de familia (Sí)	9.47
Cabeza de familia (No)	10.13
Primaria incompleta (padre)	9.39
Primaria completa (padre)	10.00
Secundaria incompleta (padre)	9.68
Secundaria completa (padre)	10.26
Educación técnica o tecnológica incompleta (padre)	9.98
Educación técnica o tecnológica completa (padre)	10.37
Educación profesional incompleta (padre)	11.05
Educación profesional completa (padre)	10.15
Postgrado (padre)	10.21
Ninguno (padre)	9.05
Educación técnica o tecnológica completa (madre)	9.84
Educación técnica o tecnológica incompleta (madre)	10.00
Educación profesional completa (madre)	10.76
Educación profesional incompleta (madre)	11.45
Ninguno (madre)	8.90
Postgrado (madre)	9.70
Secundaria completa (madre)	10.14
Secundaria incompleta (madre)	10.19
Primaria completa (madre)	9.43
Primaria incompleta (madre)	9.43
Soltero	10.07
Casado	8.70
unión libre (No)	7.60
Separado	9.70

5 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, según se presentó en la sección 2.2.

5.1 ESPECIFICACIÓN DE LOS MODELOS

La Tabla 1 muestra la distribución para cada variable respuesta y sus correspondientes funciones de enlace.

Tabla 1 Distribuciones y funciones de enlace

DISTRIBUCIÓN	VARIABLE RESPUESTA	PARÁMETROS			
		μ	δ	ν	τ
Logistic LO()	razonamiento cuantitativo	identity	log	-	-
Gamma GA()	lectura crítica	log	log	-	-
Gamma Generalizada GG()	comunicación escrita	log	log	identity	-
skew t type 3 ST3()	inglés	identity	log	log	log

Para **razonamiento cuantitativo**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución Logistic, con un AIC de 489.3649. Distribución cuyos parámetros se describieron en la sección 2.5.

```
> ajuste$fit
      LO      PE      TF      GA      SEP3      SEP2      SEP1
489.3649 490.9942 491.2547 492.3614 492.9214 492.9322 492.9425
      ST5      ST1      BCTo      ST4      SN1      SN2      BCCGo
493.2098 493.2113 493.2124 493.2431 493.3397 493.4041 493.4432
> |
```

Figura 5-1: Ajuste datos razonamiento cuantitativo

La prueba gráfica que respalda el ajuste de la distribución Logistic sobre el puntaje en razonamiento cuantitativo se presenta en la figura 5.2.

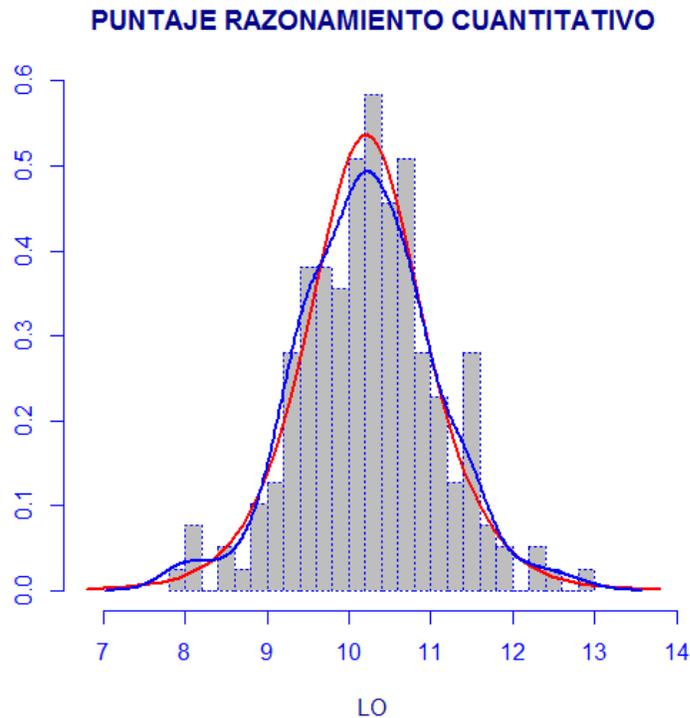


Figura 5-2: Prueba gráfica razonamiento cuantitativo

En cuanto a la variable respuesta **lectura crítica**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución Gamma, con un AIC de 431.9864. Distribución cuyos parámetros se describieron en la sección 2.5.

```
> ajuste$fit
      GA      LOGNO      PE      IG      SN2      GG      BCCGo
431.9864 432.7889 432.7905 432.7909 433.2477 433.2889 433.3044
      BCPEo      SHASHo      ST3      BCTo      ST2      JSU      ST4
434.7890 434.9213 435.2477 435.3044 435.3050 435.3064 435.3064
> |
```

Figura 5-3: Ajuste datos lectura crítica

La prueba gráfica que respalda el ajuste de la distribución Gamma sobre el puntaje en lectura crítica se presenta en la figura 5.4.

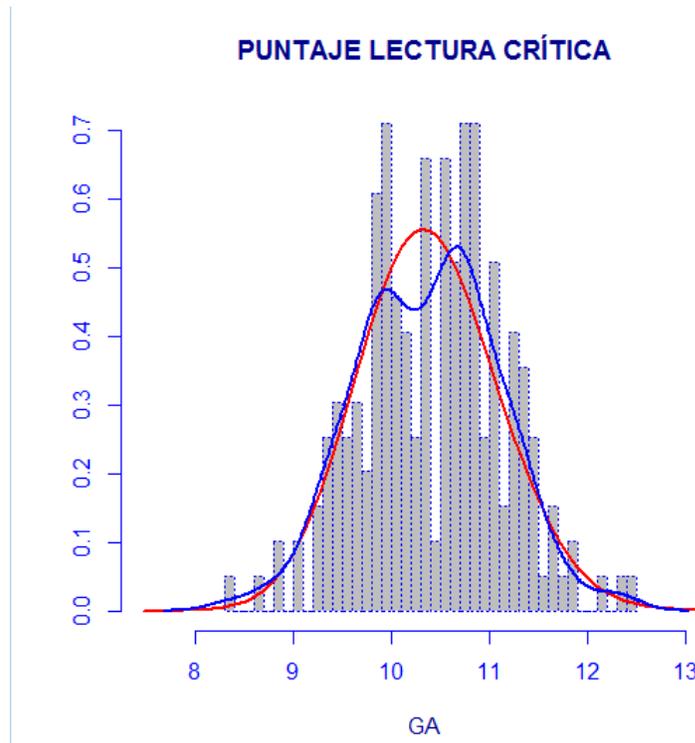


Figura 5-4: Prueba gráfica lectura crítica

En cuanto a la variable respuesta **comunicación escrita**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución Gamma Generalizada, con un AIC de 567.1342. Distribución cuyos parámetros se describieron en la sección 2.5.

```
> ajuste$fit
      GG      SN2      BCCGo      PE      TF      SN1      exGAUS
567.1342 567.2265 567.2710 567.2861 567.4810 567.4810 567.4825
      BCTo      GT      ST1      ST2      EGB2      ST5      ST4
569.2710 569.2862 569.2864 569.2864 569.4774 569.4808 569.4810
      PARETO2
1295.3744
> |
```

Figura 5-5: Ajuste datos comunicación escrita

La prueba gráfica que respalda el ajuste de la distribución Gamma Generalizada sobre el puntaje en comunicación escrita se presenta en la figura 5.6.

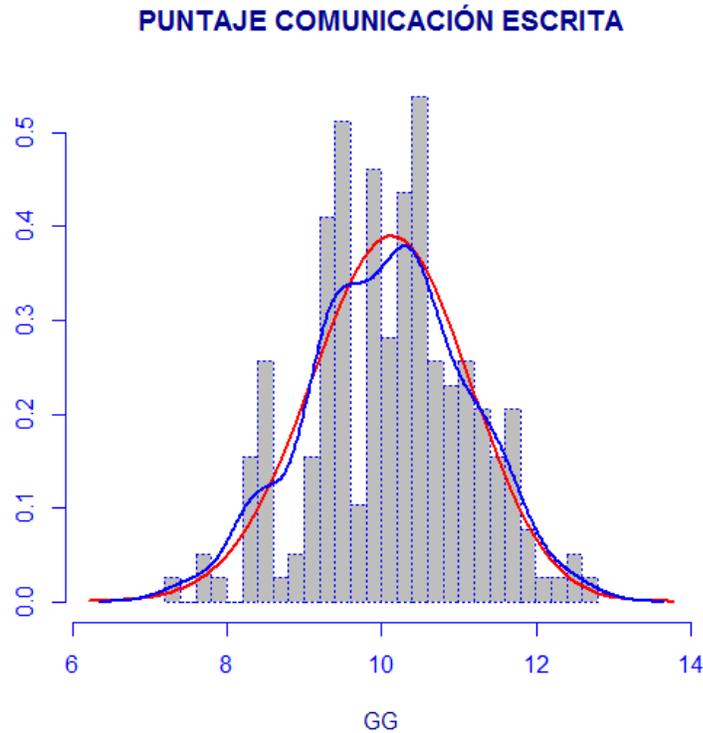


Figura 5-6: Prueba gráfica comunicación escrita

Para la variable respuesta **inglés**, el mejor ajuste que se presenta en los datos teniendo en cuenta el criterio AIC es la distribución skew t type 3, con un AIC de 547.6214. Distribución cuyos parámetros se describieron en la sección 2.5.

```
> ajuste$fit
      ST3      ST2      ST5      JSU      TF      ST4      ST1
547.6214 548.5855 552.7162 553.0192 555.2197 556.6939 556.9778
      SN2      SN1      exGAUS      RG      EXP      PARETO2
662.4654 687.1015 687.1517 932.3432 1310.8423 1312.8423
> |
```

Figura 5-7: Ajuste datos inglés

La prueba gráfica que respalda el ajuste de la distribución skew t type 3 sobre el puntaje en inglés se presenta en la figura 5.8.

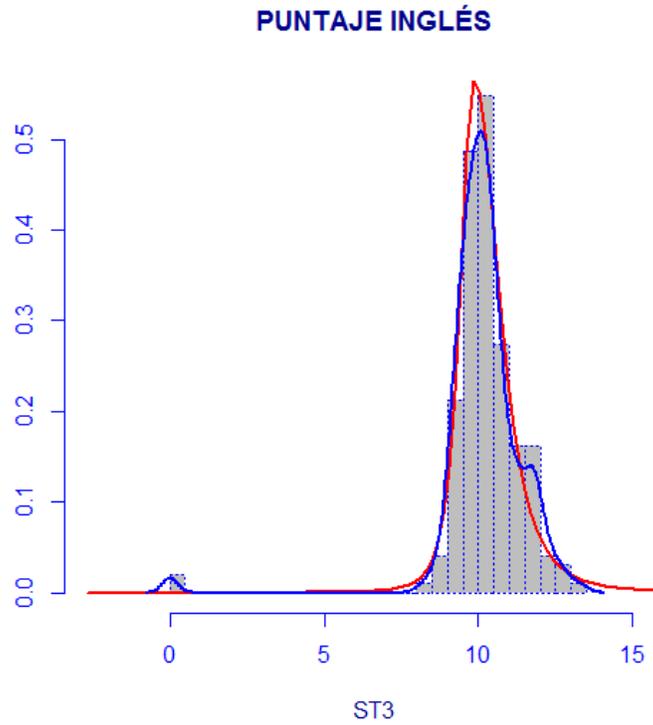


Figura 5-8: Prueba gráfica inglés

5.2 FACTORES DEMOGRÁFICOS ASOCIABLES AL COMPONENTE GENÉRICO

La Tabla 17 muestra las variables explicativas a tener en cuenta para los modelos demográficos.

Tabla 17 Componente sistemático demográfico

Notación	Nombre
X_1	Edad
X_2	Número de personas que conforman su hogar
X_3	Número de personas de las que usted se encuentra a cargo
X_4	Hogar actual
X_5	Sexo
X_6	Estado civil

El modelo lineal queda determinado por el siguiente predictor lineal

$$\eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_j X_4 + \alpha_j X_5 + \tau_k X_6 \quad (5-1)$$

Donde:

β_0 = Intercepto

β_1 = Efecto de la edad de estudiante.

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

β_3 = Efecto del número de personas de las que usted se encuentra a cargo.

γ_j = Efecto de la situación de su 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 en cuanto a **RAZONAMIENTO CUANTITATIVO** se utilizó el criterio StepGAICALL.A descrito en la sección 2.5

```

Mu link function: identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.72996    0.48764  22.003 < 2e-16 ***
Hogar_actual[T.temporal]  0.20417    0.08821   2.314 0.021770 *
FAMI_NUM_PERS_CARGO      0.22388    0.07787   2.875 0.004529 **
ESTU_GENERO[T.M]        0.33044    0.09338   3.539 0.000512 ***
Estado_civil_cod[T.Separado]  1.14347    1.00685   1.136 0.257598
Estado_civil_cod[T.Soltero]  0.68928    0.17524   3.933 0.000119 ***
Estado_civil_cod[T.Unión libre]  0.74224    0.16555   4.483 1.3e-05 ***
EDAD              -0.05608    0.01647  -3.404 0.000818 ***
---
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)   -1.599219    0.838507  -1.907 0.05808 .
Hogar_actual[T.temporal] -0.030540    0.131140  -0.233 0.81612
FAMI_NUM_PERS_GRUP_FAM   -0.019984    0.036565  -0.547 0.58538
FAMI_NUM_PERS_CARGO     -0.247872    0.220205  -1.126 0.26182
ESTU_GENERO[T.M]        0.049140    0.127338   0.386 0.70002
Estado_civil_cod[T.Separado]  2.067471    0.650042   3.181 0.00173 **
Estado_civil_cod[T.Soltero]  1.041735    0.574480   1.813 0.07144 .
Estado_civil_cod[T.Unión libre]  0.296793    0.720886   0.412 0.68104
EDAD                 -0.006264    0.020352  -0.308 0.75862
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 197
Degrees of Freedom for the fit: 17
Residual Deg. of Freedom: 180
                        at cycle: 8

Global Deviance: 448.8833
AIC: 482.8833
SBC: 538.6977
*****

```

Figura 5-9: Modelo óptimo razonamiento cuantitativo

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el estado civil se tiene que aquellos estudiantes solteros obtienen, 0.68 puntos más respecto a los estudiantes casados (nivel de significancia 1%). Del mismo modo aquellos estudiantes en unión libre obtienen, 0.74 puntos más respecto a los estudiantes casados.

Ahora bien, se puede afirmar, con un nivel de significancia inferior al 1 %, que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si su hogar actual es permanente o temporal, se tiene que aquellos que cuentan con un hogar temporal obtienen 0.20 puntos más que los que cuentan con un hogar permanente. Por otra parte si comparamos dos estudiantes que tan sólo difieren en que uno es mayor un año que el otro, se obtiene 0.05 puntos menos en razonamiento cuantitativo. Así como también si comparamos dos estudiantes de Diseño Industrial que tan sólo difieren en una persona de más a cargo, se obtiene 0.22 puntos más en razonamiento cuantitativo. También se puede afirmar, con un nivel de significancia inferior al 1 %, que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su género, se tiene que los hombres obtienen 0.33 puntos más que las mujeres.

Respecto a la variación en el puntaje de razonamiento cuantitativo se tiene que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el estado civil se tiene que aquellos estudiantes separados presentan una mayor variabilidad, de 2.38 puntos respecto a un estudiante casado (nivel de significancia del 1 %). Se puede afirmar, con un nivel de significancia inferior al 10 %, que aquellos estudiantes solteros presentan una mayor variabilidad, de 0.85 puntos, respecto a un estudiante casado.

De este modo podemos decir que estar soltero o en unión libre, además de ser hombre, tener un hogar temporal, tener personas a cargo y la edad, influyen sobre el valor esperado del puntaje en razonamiento cuantitativo.

Una vez ajustado el modelo GAMLSS, se utilizarán los residuales cuantiles aleatorizados descritos en la sección 2.5 para comprobar la adecuación del modelo y, más específicamente, la distribución escogida para la variable respuesta. Las siguientes figuras 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.

```
> plot(MO)
*****
      Summary of the Quantile Residuals
      mean      = 0.005991738
      variance  = 0.9906421
      coef. of skewness = -0.009402142
      coef. of kurtosis  = 2.618108
      Filliben correlation coefficient = 0.9977839
*****
> |
```

Figura 5-10: Prueba ajuste modelo demográfico razonamiento cuantitativo

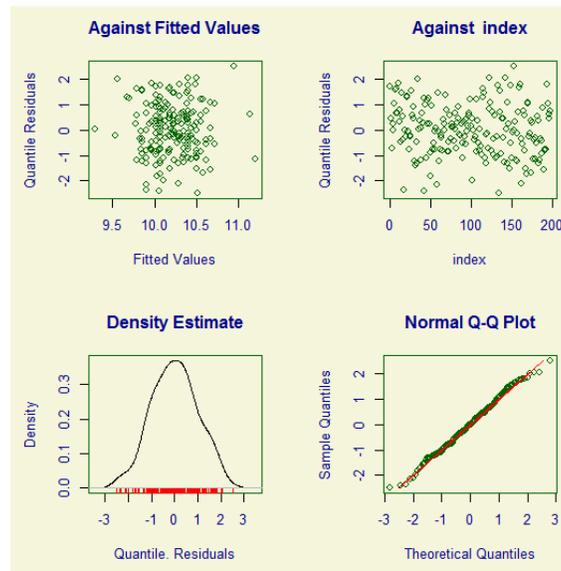


Figura 5-11: Diagnóstico modelo puntaje razonamiento cuantitativo factor demográfico

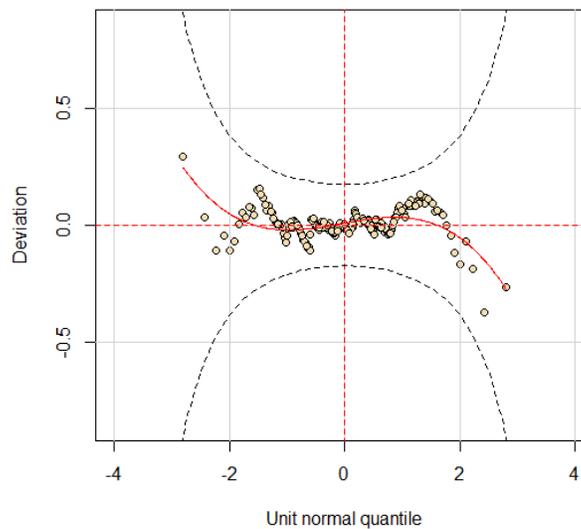


Figura 5-12: Ajuste de la distribución modelo razonamiento cuantitativo factor demográfico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-9) presentan un buen ajuste a una normal, esto quiere decir que la distribución Logística es la adecuada para describir el comportamiento del puntaje de razonamiento cuantitativo en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **LECTURA CRÍTICA** se tuvo en cuenta el criterio StepGAICALL.A descrito en la sección 2.5.

```

Mu link function:  log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.434320   0.053751  45.289  <2e-16 ***
Hogar_actual[T.temporal] 0.015895   0.009788   1.624   0.1061
FAMI_NUM_PERS_CARGO    0.013332   0.007531   1.770   0.0783 .
EDAD              -0.003804   0.002050  -1.855   0.0651 .
---
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)   -2.515459   0.723564  -3.476 0.000634 ***
Hogar_actual[T.temporal] 0.006231   0.108428   0.057 0.954237
FAMI_NUM_PERS_GRP_FAM   -0.043176   0.034301  -1.259 0.209719
FAMI_NUM_PERS_CARGO   -0.151528   0.197003  -0.769 0.442783
ESTU_GENERO[T.M]      -0.085197   0.110642  -0.770 0.442273
Estado_civil_cod[T.Separado] 0.038023   0.553961   0.069 0.945352
Estado_civil_cod[T.Soltero] -0.276360   0.531879  -0.520 0.603972
Estado_civil_cod[T.Unión libre] -1.282730   0.590212  -2.173 0.031032 *
EDAD               0.012921   0.018463   0.700 0.484899
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

No. of observations in the fit: 197
Degrees of Freedom for the fit: 13
  Residual Deg. of Freedom: 184
                        at cycle: 8

Global Deviance:  416.8195
      AIC:        442.8195
      SBC:        485.5011

```

Figura 5-13: Modelo óptimo lectura crítica

A partir de la información del modelo óptimo se puede afirmar respecto al valor esperado del puntaje en lectura crítica que si comparamos dos estudiantes que tan sólo difieren en que uno es mayor un año que el otro, se obtiene 0.99 puntos menos en lectura crítica (nivel de significancia del 10%). Por otra parte si comparamos dos estudiantes de Diseño Industrial que tan sólo difieren en una persona de más a cargo, se obtiene en promedio 1.01 puntos más en lectura crítica.

Respecto a la variación del puntaje en lectura crítica se encontró que al comparar dos estu-

diantes donde el segundo difiere del primero en tan sólo el estado civil se tiene que con un nivel de significancia inferior al 5%, aquellos estudiantes en unión libre presentan una menor variabilidad, de 0.27 puntos respecto a un estudiante casado.

De este modo podemos decir que estar en unión libre, tener personas a cargo y la edad, influyen sobre el valor esperado del puntaje en lectura crítica.

Las siguientes figuras 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.

```
> plot(MO)
*****
      Summary of the Quantile Residuals
      mean      = -0.01142456
      variance   = 1.004255
      coef. of skewness = -0.09739603
      coef. of kurtosis  = 2.631358
      Filliben correlation coefficient = 0.9978462
*****
```

Figura 5-14: Prueba ajuste modelo demográfico lectura crítica

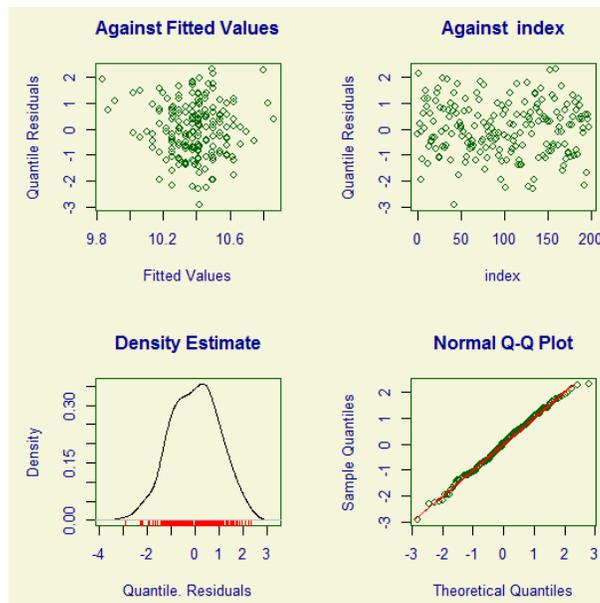


Figura 5-15: Diagnóstico modelo puntaje lectura crítica factor demográfico

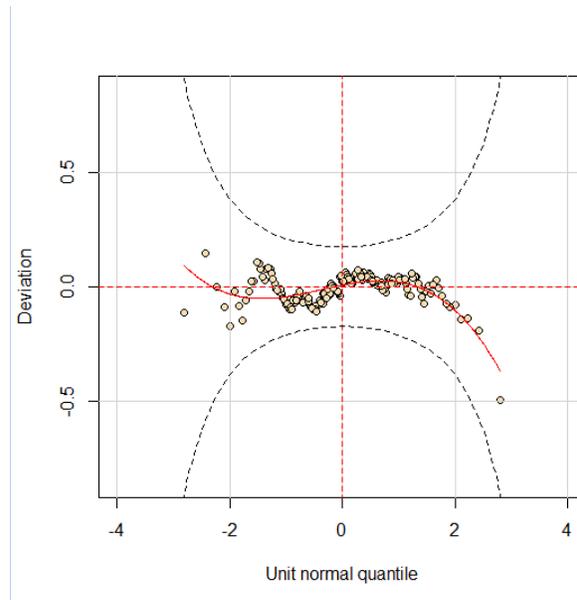


Figura 5-16: Ajuste de la distribución modelo lectura crítica factor demográfico

En las figuras anteriores podemos observar que los residuales aleatorizados para el modelo óptimo (figura 5-13) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma es la adecuada para describir el comportamiento del puntaje en lectura crítica en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **COMUNICACIÓN ESCRITA** se tuvo en cuenta el criterio StepGAICALL.A descrito en la sección 2.5.

```
-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.1606332	0.0794984	27.178	< 2e-16 ***
Hogar_actual[T.temporal]	0.0413877	0.0150568	2.749	0.00661 **
FAMI_NUM_PERS_GRP FAM	-0.0034306	0.0036208	-0.947	0.34470
FAMI_NUM_PERS_CARGO	0.0150789	0.0131865	1.144	0.25438
ESTU_GENERO[T.M]	-0.0278564	0.0140307	-1.985	0.04865 *
Estado_civil_cod[T.Separado]	0.1181806	0.0546880	2.161	0.03205 *
Estado_civil_cod[T.Soltero]	0.2146191	0.0429084	5.002	1.37e-06 ***
Estado_civil_cod[T.Unión libre]	0.1502864	0.0514858	2.919	0.00397 **
EDAD	-0.0008767	0.0025948	-0.338	0.73586

```
---
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)	-2.267932	0.939615	-2.414	0.0168 *
Hogar_actual[T.temporal]	0.069283	0.107857	0.642	0.5215
FAMI_NUM_PERS_GRP FAM	-0.055717	0.034643	-1.608	0.1096
FAMI_NUM_PERS_CARGO	-0.184510	0.274213	-0.673	0.5019
ESTU_GENERO[T.M]	-0.084130	0.107182	-0.785	0.4336
Estado_civil_cod[T.Separado]	0.430272	0.606556	0.709	0.4790
Estado_civil_cod[T.Soltero]	0.274684	0.802138	0.342	0.7324
Estado_civil_cod[T.Unión libre]	0.307642	0.870975	0.353	0.7243
EDAD	-0.004922	0.020301	-0.242	0.8087

```
---
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)	5.664	2.389	2.371	0.0188 *

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

-----
No. of observations in the fit: 195
Degrees of Freedom for the fit: 19
Residual Deg. of Freedom: 176
at cycle: 13

Global Deviance: 532.7612
AIC: 570.7612
SBC: 632.9482
```

Figura 5-17: Modelo óptimo comunicación escrita

A partir de información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el estado civil se tiene que con un nivel de significancia inferior al 1 %, aquellos estudiantes solteros obtienen, 0.61 puntos más respecto a los estudiantes casados. Ahora bien, aquellos estudiantes separados obtienen, 1.11 puntos más respecto a los estudiantes casados (nivel de significancia de 5 %). Del mismo modo aquellos estudiantes en unión libre obtienen, 1.16 puntos más respecto a los estudiantes casados. 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 tan sólo si su hogar actual es permanente o temporal, se tiene que aquellos que cuentan con un hogar temporal obtienen

1.04 puntos más que los que cuentan con un hogar permanente. Con un nivel de significancia inferior al 5 %, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su género, se tiene que los hombres obtienen 0.98 puntos más que las mujeres.

Respecto a la variación en el puntaje de comunicación escrita podemos afirmar que las variables demográficas consideradas en este estudio no explican la variabilidad del puntaje en comunicación escrita.

De esta forma podemos decir que estar soltero, separado o en unión libre, además de ser hombre, tener un hogar temporal, influyen sobre el valor esperado del puntaje en comunicación escrita.

Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean = -0.001742645
      variance = 1.002251
      coef. of skewness = -0.01302183
      coef. of kurtosis = 2.927349
Filliben correlation coefficient = 0.99764
*****
```

Figura 5-18: Prueba ajuste modelo demográfico comunicación escrita

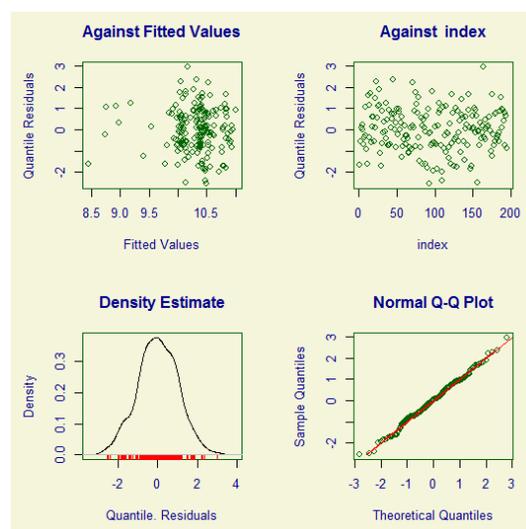


Figura 5-19: Diagnóstico modelo puntaje comunicación escrita factor demográfico

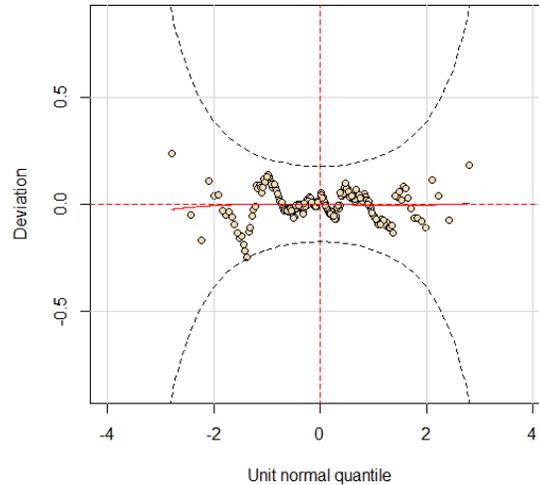


Figura 5-20: Ajuste de la distribución modelo comunicación escrita factor demográfico

En las figuras anteriores podemos observar que los residuales aleatorizados para el modelo óptimo (figura 5-17) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma Generalizada es la adecuada para describir el comportamiento del puntaje de comunicación escrita en el programa de Diseño Industrial.

En cuanto al **DESEMPEÑO EN COMUNICACIÓN ESCRITA** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo, tabla 18.

Tabla 18 Resumen modelo óptimo

	Coef estimado	Error	Wald	P-valor	OR	I.C 95 %	
Estado civil (separado)	1.6303	1.2288	1.327	0.18	5.066	0.4593	56.747
Estado civil (soltero)	2.8841	0.8091	3.564	0.0004	17.88	3.6629	87.352
Estado civil (unión libre)	2.1314	1.2343	1.727	0.085	8.379	0.7499	94.676
Hogar actual (temporal)	0.4954	0.2774	1.786	0.075	1.632	0.9527	2.826
Nive 12 — Nivel 3	-0.4048	0.8091	-0.5004				
Nivel 3 — Nivel 4	1.3491	0.8089	1.6678				
Nivel 4 — Nivel 5	3.0050	0.8309	3.6165				
Nivel 5 — Nivel 6	4.6927	0.8532	5.5000				
Nivel 6 — Nivel 7	6.4920	0.9311	6.9720				

Para la variable estado civil se observa que su razón de probabilidad y su intervalo de confianza al 95 % es mayor que uno, lo que significa que la probabilidad acumulada, comenzando del Nivel de desempeño 1 hasta el 8, aumenta en todos los valores que puede tomar comunicación escrita, así un estudiante soltero es más probable que obtenga valores altos en el desempeño en comunicación escrita respecto a uno casado.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **INGLÉS** se tuvo en cuenta el criterio StepGAICALL.A descrito en la sección 2.5.

```

Mu link function: identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.40538    1.02658  10.136 <2e-16 ***
Hogar_actual[T.temporal]  0.11647    0.12553   0.928  0.355
FAMI_NUM_PERS_GRUP_FAM   -0.04875    0.03032  -1.608  0.110
FAMI_NUM_PERS_CARGO      0.05189    0.23013   0.225  0.822
ESTU_GENERO[T.M]        -0.12063    0.10978  -1.099  0.273
Estado_civil_cod[T.Separado]  0.45652    0.54110   0.844  0.400
Estado_civil_cod[T.Soltero]  0.19960    0.64037   0.312  0.756
Estado_civil_cod[T.Unión libre] 0.89658    0.66453   1.349  0.179
EDAD                 -0.02229    0.02335  -0.955  0.341
---
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)  -0.4764    0.1082  -4.405 1.8e-05 ***
---
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)   0.3906    0.1463   2.67 0.00826 **
---
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)   1.89930    0.73467   2.585 0.0105 *
Hogar_actual[T.temporal] -0.08485    0.38280  -0.222 0.8248
ESTU_GENERO[T.M]        -0.99075    0.70559  -1.404 0.1620
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 197
Degrees of Freedom for the fit: 14
Residual Deg. of Freedom: 183
                        at cycle: 20

Global Deviance: 525.6048
AIC: 553.6048
BIC: 599.5697
-----

```

Figura 5-21: Modelo óptimo inglés

Se tiene que las variables demográficas consideradas en el estudio, tales como hogar actual, número de personas de las que usted se encuentra a cargo, género, estado civil, número de personas que conforman el hogar y edad, no explican el puntaje obtenido en inglés. Es decir estos aspectos no influyen en el valor esperado ni la variabilidad del puntaje en inglés.

Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean = -0.03531054
      variance = 1.017545
      coef. of skewness = -0.2264989
      coef. of kurtosis = 3.362173
Filliben correlation coefficient = 0.9943945
*****
```

Figura 5-22: Prueba ajuste modelo demográfico inglés

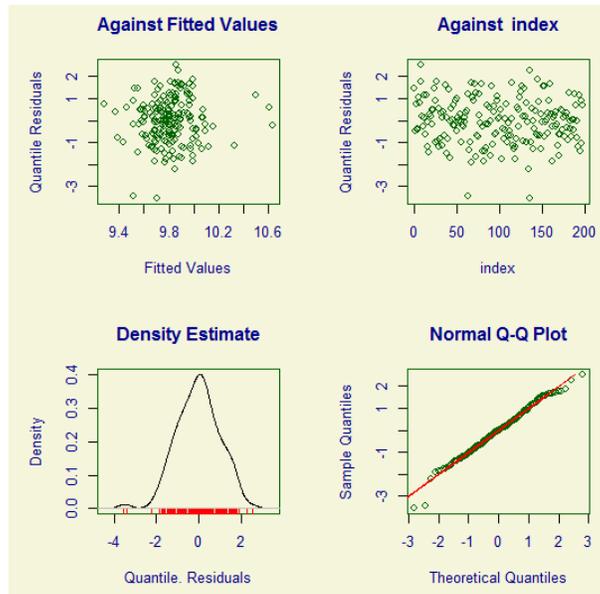


Figura 5-23: Diagnóstico modelo puntaje inglés factor demográfico

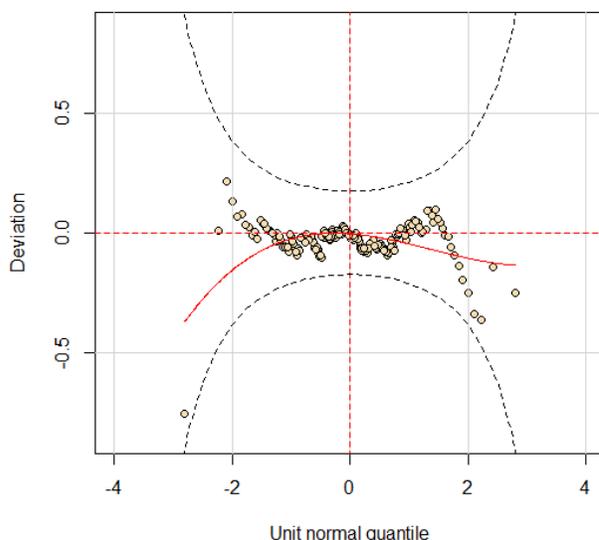


Figura 5-24: Ajuste de la distribución modelo inglés factor demográfico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-21) presentan un buen ajuste a una normal, esto quiere decir que la distribución skew t type 3 es la adecuada para describir el comportamiento del puntaje de inglés en el programa de Diseño Industrial.

En cuanto al DESEMPEÑO EN INGLÉS se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo, tabla 19.

Tabla 19 Resumen modelo óptimo

	Coef estimado	Error	Wald	P-valor	OR	I.C 95 %	
Núm personas grupo familiar	-0.1265	0.0751	-1.685	0.093	0.865	0.7605	1.020
A— A1	-2.3242	0.3770	-6.1654				
A1 — A2	-0.7879	0.3370	-2.3381				
A2 — B+	0.4460	0.3362	1.3266				
B+ — B1	0.6594	0.3394	1.9431				

Se tiene que la variable demográfica número de personas que conforman el hogar no explica el desempeño en inglés, debido a que la variable no presenta significancia estadística.

5.3 FACTORES ACADÉMICOS ASOCIABLES AL COMPONENTE GENÉRICO

La Tabla 20 muestra las variables explicativas a tener en cuenta para los modelos académicos.

Tabla 20 Componente sistemático académico

Notación	Nombre
X_1	Tomó algún curso de preparación al examen
X_2	Tipo de bachillerato
X_3	Nivel educativo del padre
X_4	Nivel educativo de la madre

El modelo lineal queda determinado por el siguiente predictor lineal

$$\eta_i = \beta_0 + \beta_1 X_1 + \sum_{j=2}^4 \tau_{jk} X_j \quad (5-2)$$

Donde:

β_0 = Intercepto

$\beta_1 k$ = Efecto de tomar algún curso de preparación al examen, con $k = 0, 1$

$\tau_2 k$ = Efecto del tipo de bachillerato, con $k = 1, 2, 3, 4$

$\tau_3 k$ = Efecto del nivel educativo del padre, con $k = 1, 2, 3, 4, 5$

$\tau_4 k$ = 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** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function: identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.51386	0.16678	63.042	< 2e-16 ***
Educación_madre[T.Primaria]	-0.61248	0.19970	-3.067	0.00260 **
Educación_madre[T.Secundaria]	-0.53076	0.04246	-12.500	< 2e-16 ***
Educación_madre[T.Sin estudio]	-2.03347	0.07634	-26.638	< 2e-16 ***
Educación_madre[T.Técnico]	-0.01158	0.26032	-0.044	0.96459
Educación_padre[T.Primaria]	-0.01638	0.22755	-0.072	0.94271
Educación_padre[T.Secundaria]	-0.10103	0.20387	-0.496	0.62100
Educación_padre[T.Sin estudio]	-0.47584	0.16564	-2.873	0.00471 **
Educación_padre[T.Técnico]	-0.22432	0.21612	-1.038	0.30111
ESTU_TITULO_BTO[T.N]	-0.44412	0.43153	-1.029	0.30519
ESTU_TITULO_BTO[T.T]	0.26077	0.03873	6.734	4.02e-10 ***

```

---
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)	-0.56836	0.19930	-2.852	0.00501 **
Educación_madre[T.Primaria]	0.15930	0.27535	0.579	0.56385
Educación_madre[T.Secundaria]	0.02629	0.18946	0.139	0.88983
Educación_madre[T.Sin estudio]	0.94276	0.63488	1.485	0.13983
Educación_madre[T.Técnico]	-0.01041	0.28468	-0.037	0.97087
Educación_padre[T.Primaria]	-0.24539	0.25014	-0.981	0.32828
Educación_padre[T.Secundaria]	-0.12152	0.20858	-0.583	0.56109
Educación_padre[T.Sin estudio]	-3.40543	0.43402	-7.846	1.03e-12 ***
Educación_padre[T.Técnico]	-0.18662	0.23519	-0.793	0.42885
ESTU_TITULO_BTO[T.N]	-0.19980	0.48495	-0.412	0.68097
ESTU_TITULO_BTO[T.T]	-0.27757	0.15850	-1.751	0.08211 .
Tomó_curso[T.Si]	-0.04771	0.15938	-0.299	0.76511

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
No. of observations in the fit: 162
Degrees of Freedom for the fit: 23
  Residual Deg. of Freedom: 139
                        at cycle: 7

Global Deviance:    358.4459
                   AIC:    404.4459
                   SBC:    475.4607
-----

```

Figura 5-25: Modelo óptimo razonamiento cuantitativo

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo de la madre se tiene que con un nivel de significancia inferior al 1 %, aquellos que tienen madres con nivel educativo primaria obtienen, 0.61 puntos menos respecto a los que tienen madres con nivel educativo universitario. Del mismo modo aquellos estudiantes que tienen madres con nivel

educativo secundaria obtienen, 0.53 puntos menos respecto a los que tienen madres con nivel educativo universitario. También aquellos que tienen madres sin estudios obtienen, 2.03 puntos menos respecto a los que tienen madres con nivel educativo universitario. Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo del padre se tiene que con un nivel de significancia inferior al 1 %, aquellos que tienen padres sin nivel educativo obtienen, 0.47 puntos menos respecto a los que tienen padres con nivel educativo universitario. Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el tipo de bachillerato del que se graduó se tiene que con un nivel de significancia inferior al 1 %, aquellos que tienen un tipo de bachillerato técnico obtienen, 0.26 puntos más respecto a los que tienen un tipo de bachillerato académico.

Respecto a la variabilidad en el puntaje de razonamiento cuantitativo se tiene que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo del padre se tiene que con un nivel de significancia inferior al 1 %, aquellos que tienen padres sin algún nivel educativo presentan una menor variabilidad, de 0.01 puntos respecto a los tienen padres con nivel educativo universitario. Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el tipo de bachillerato del que se graduó se tiene que con un nivel de significancia inferior al 10 %, aquellos que tienen un tipo de bachillerato técnico presentan una menor variabilidad, de 0.23 puntos respecto a los que tienen un tipo de bachillerato académico.

De este modo se puede decir que aspectos académicos como tener un nivel educativo universitario por parte del padre o de la madre, además de haberse graduado del colegio de tipo técnico influyen sobre el valor esperado del puntaje en razonamiento cuantitativo.

Una vez ajustado el modelo GAMLSS, se utilizarán los residuales cuantiles aleatorizados descritos en la sección 2.5 para comprobar la adecuación del modelo y, más específicamente, la distribución escogida para la variable respuesta. Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean = 0.04041464
  variance = 0.9935676
  coef. of skewness = 0.2317482
  coef. of kurtosis = 2.661979
Filliben correlation coefficient = 0.995747
*****
```

Figura 5-26: Prueba ajuste modelo académico razonamiento cuantitativo

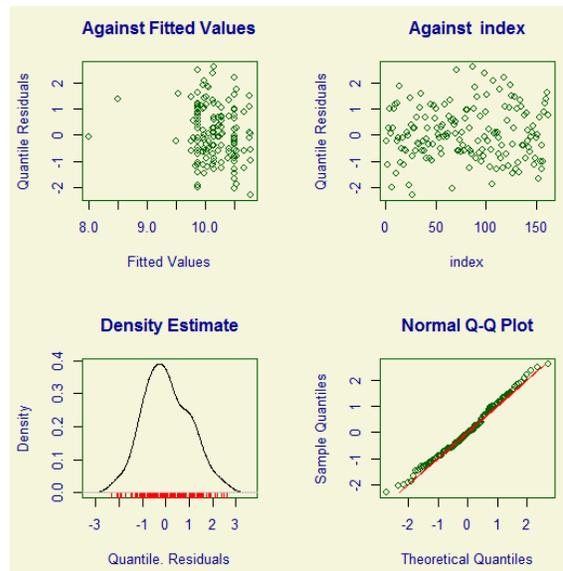


Figura 5-27: Diagnóstico modelo puntaje razonamiento cuantitativo factor académico

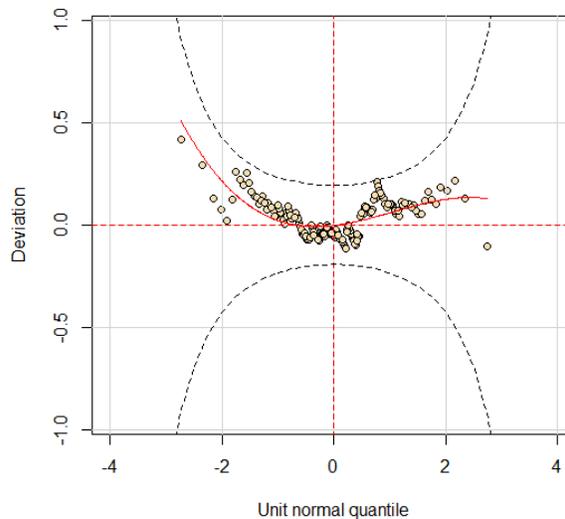


Figura 5-28: Ajuste de la distribución modelo razonamiento cuantitativo factor académico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-25) presentan un buen ajuste a una normal, esto quiere decir que la distribución logistic es la adecuada para describir el comportamiento del puntaje de razonamiento cuantitativo en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **LECTURA CRÍTICA** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function: log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.383600  0.013716 173.787 < 2e-16 ***
Educación_madrePrimaria -0.030652  0.016020  -1.913 0.057759 .
Educación_madreSecundaria -0.026997  0.007808  -3.458 0.000723 ***
Educación_madreSin estudio -0.035035  0.053375  -0.656 0.512657
Educación_madreTécnico -0.075834  0.015110  -5.019 1.57e-06 ***
Educación_padrePrimaria -0.020210  0.019240  -1.050 0.295362
Educación_padreSecundaria -0.025611  0.017084  -1.499 0.136101
Educación_padreSin estudio -0.081967  0.028056  -2.922 0.004066 **
Educación_padreTécnico -0.044377  0.015136  -2.932 0.003942 **
ESTU_TITULO_BT0N    0.030201  0.013488   2.239 0.026737 *
ESTU_TITULO_BT0T    0.021812  0.011836   1.843 0.067472 .
---
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)   -2.296445  0.173538 -13.233 < 2e-16 ***
Educación_madrePrimaria -0.155139  0.247080  -0.628 0.53111
Educación_madreSecundaria  0.148311  0.178609   0.830 0.40776
Educación_madreSin estudio  0.172931  0.604488   0.286 0.77524
Educación_madreTécnico  -0.318361  0.288466  -1.104 0.27166
Educación_padrePrimaria -0.497169  0.220696  -2.253 0.02584 *
Educación_padreSecundaria -0.450706  0.197966  -2.277 0.02433 *
Educación_padreSin estudio -0.826452  0.415301  -1.990 0.04855 *
Educación_padreTécnico  -0.536848  0.204860  -2.621 0.00975 **
ESTU_TITULO_BT0N   -2.674216  0.428265  -6.244 4.85e-09 ***
ESTU_TITULO_BT0T   -0.007979  0.136672  -0.058 0.95353
Tomó_cursoSí      -0.040730  0.141341  -0.288 0.77365
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 162
Degrees of Freedom for the fit: 23
  Residual Deg. of Freedom: 139
                        at cycle: 20

Global Deviance:    325.7441
                   AIC:    371.7441
                   SBC:    442.7588

```

Figura 5-29: Modelo óptimo lectura crítica

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo de la madre se tiene que con un nivel de significancia inferior al 10 %, aquellos que tienen madres con nivel educativo primaria obtienen, 0.97 puntos menos respecto a los que tienen madres con nivel educativo universitario. Con un nivel de significancia inferior al 1 %, aquellos que tienen madres con nivel educativo secundaria, 0.98 puntos menos respecto a los que tienen madres con

nivel educativo universitario. Del mismo modo, aquellos estudiantes que tienen madres con nivel educativo técnico obtienen, 0.93 puntos menos respecto a los que tienen madres con nivel educativo universitario.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo del padre se tiene que con un nivel de significancia inferior al 5%, aquellos que tienen padres sin nivel educativo obtienen, 0.92 puntos menos respecto a los que tienen padres con nivel educativo universitario. Del mismo modo aquellos estudiantes que tienen padres con nivel educativo técnico obtienen, 0.96 puntos menos respecto a los que tienen padres con nivel educativo universitario.

Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el tipo de bachillerato del que se graduó se tiene que, con un nivel de significancia inferior al 5%, aquellos que tienen un tipo de bachillerato normalista superior obtienen, 1.03 puntos más respecto a los que tienen un tipo de bachillerato académico. Con un nivel de significancia inferior al 10%, aquellos que tienen un tipo de bachillerato técnico obtienen, 1.02 puntos más respecto a los que tienen un tipo de bachillerato académico.

Respecto a la variabilidad en el puntaje de lectura crítica se tiene que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo del padre se tiene que, con un nivel de significancia inferior al 1%, aquellos que tienen padres con nivel educativo técnico presentan una menor variabilidad de 0.58 puntos respecto a los que tienen padres con nivel educativo universitario. Con un nivel de significancia inferior al 5%, aquellos que tienen padres con nivel educativo primaria presentan una menor variabilidad de 0.61 puntos respecto a los que tienen padres con nivel educativo universitario. Del mismo modo, aquellos estudiantes que tienen padres con nivel educativo secundaria presentan una menor variabilidad de 0.63 puntos respecto a los que tienen padres con nivel educativo universitario. También aquellos estudiantes que tienen padres sin nivel educativo presentan una menor variabilidad de 0.44 puntos respecto a los que tienen padres con nivel educativo universitario.

Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el tipo de bachillerato del que se graduó se tiene que, con un nivel de significancia inferior al 1%, que aquellos que tienen un tipo de bachillerato normalista superior presentan una menor variabilidad de 0.06 puntos respecto a los que tienen un tipo de bachillerato académico.

De este modo se concluye que tener un nivel educativo universitario por parte del padre o de la madre, además de haberse graduado del colegio de tipo técnico o normalista superior

influyen sobre el valor esperado del puntaje en lectura crítica.

Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean = -0.04329253
      variance = 0.9996004
      coef. of skewness = -0.04170858
      coef. of kurtosis = 2.75173
Filliben correlation coefficient = 0.9987586
*****
```

Figura 5-30: Prueba ajuste modelo académico lectura crítica

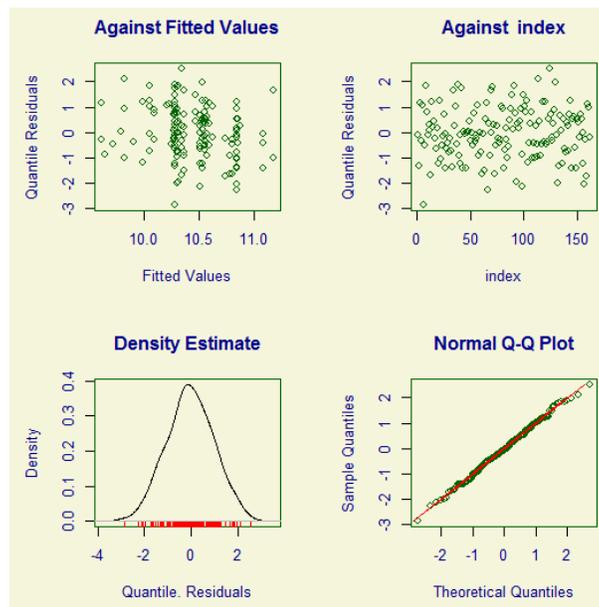


Figura 5-31: Diagnóstico modelo puntaje lectura crítica factor académico

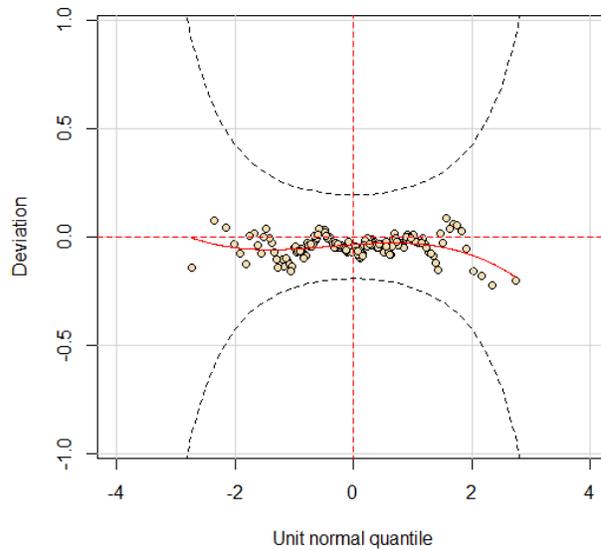


Figura 5-32: Ajuste de la distribución modelo lectura crítica factor académico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-29) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma es la adecuada para describir el comportamiento del puntaje de lectura crítica en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **COMUNICACIÓN ESCRITA** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function: log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.33693    0.02505  93.296 < 2e-16 ***
Educación_madre[T.Primaria] -0.11728    0.03354  -3.497 0.000631 ***
Educación_madre[T.Secundaria] -0.04613    0.02488  -1.854 0.065784 .
Educación_madre[T.Sin estudio] -0.15088    0.08770  -1.720 0.087572 .
Educación_madre[T.Técnico] -0.07071    0.03668  -1.928 0.055089 .
Tomó_curso[T.Si] 0.06487    0.02360   2.749 0.006762 **
---
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) -2.47632    0.07315 -33.85 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: identity
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.080     6.504   0.320  0.7496
Educación_madre[T.Primaria] -6.349     7.823  -0.801  0.4243
Educación_madre[T.Secundaria] -3.380     5.867  -0.576  0.5654
Educación_madre[T.Sin estudio] -18.775    33.806  -0.498  0.6205
Educación_madre[T.Técnico] -7.428     9.748  -0.762  0.4473
Educación_padre[T.Primaria]  1.832     5.056   0.362  0.7177
Educación_padre[T.Secundaria]  6.730     4.559   1.482  0.2018
Educación_padre[T.Sin estudio] 14.645    15.995   0.916  0.3614
Educación_padre[T.Técnico] -3.441     4.811  -0.715  0.4756
ESTU_TITULO_BIO[T.N] -1.754    13.034  -0.096  0.9235
ESTU_TITULO_BIO[T.T] -2.572     3.492  -0.737  0.4626
Tomó_curso[T.Si] 10.470     5.432   1.928  0.0559 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 160
Degrees of Freedom for the fit: 19
Residual Deg. of Freedom: 141
at cycle: 17

Global Deviance: 433.267
AIC: 471.267
SBC: 529.6953

```

Figura 5-33: Modelo óptimo comunicación escrita

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo de la madre se tiene que, con un nivel de significancia inferior al 1 %, aquellos que tienen madres con nivel educativo primaria obtienen, 0.88 puntos menos respecto a los que tienen madres con nivel educativo universitario. Con un nivel de significancia inferior al 10 %, aquellos que tienen madres con nivel educativo secundaria obtienen, 0.96 puntos menos respecto a los que tienen madres con nivel educativo universitario. Del mismo modo, aquellos estudiantes que tienen madres sin estudios obtienen, 0.86 puntos menos respecto a los que tienen madres con nivel educativo universitario. Aquellos que tienen madres con nivel educativo técnico obtienen, 0.92 puntos menos respecto a los que tienen madres con nivel educativo universitario.

Ahora bien, se puede afirmar, con un nivel de significancia inferior al 1 %, que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si tomó o no curso de preparación al examen Saber Pro, se tiene que aquellos que tomaron curso obtienen 0.94 puntos más que los que no lo tomaron.

Se puede decir que dadas las variables académicas consideradas en este estudio no explican la variabilidad del puntaje en comunicación escrita.

Podemos concluir que los aspectos académicos del estudiante como tener un nivel educativo universitario por parte de la madre, además tomar un curso de preparación al examen Saber Pro explican los resultados en comunicación escrita.

Las siguientes figuras 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.

```
> plot(MO)
*****
      Summary of the Quantile Residuals
      mean      = 0.006734587
      variance  = 1.00606
      coef. of skewness = -0.07249736
      coef. of kurtosis  = 2.722356
      Filliben correlation coefficient = 0.9982083
*****
```

Figura 5-34: Prueba ajuste modelo académico comunicación escrita

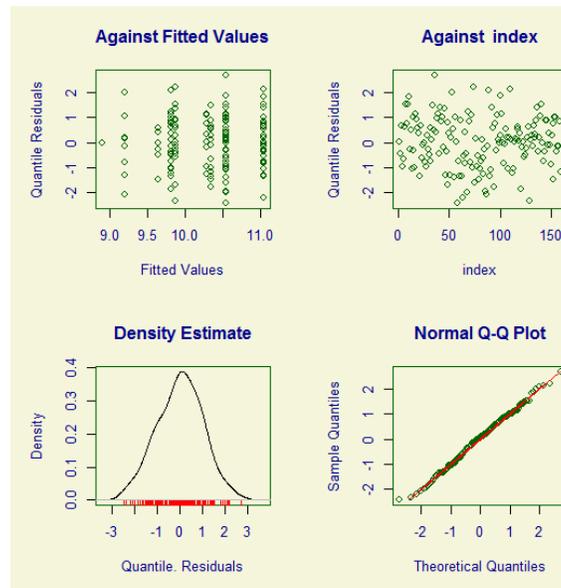


Figura 5-35: Diagnóstico modelo puntaje comunicación escrita factor académico

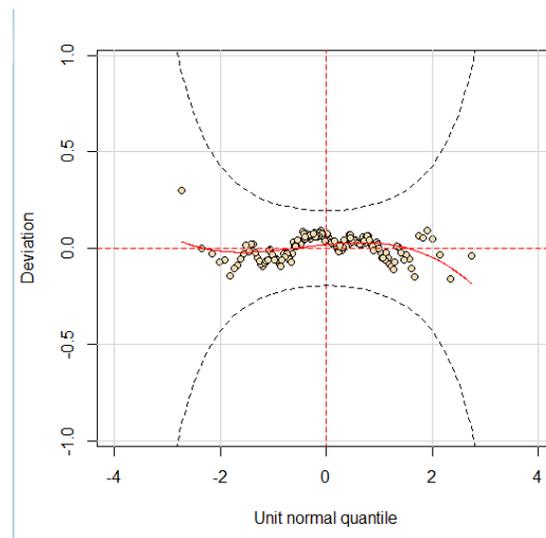


Figura 5-36: Ajuste de la distribución modelo comunicación escrita factor académico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-33) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma Generalizada es la adecuada para describir el comportamiento del puntaje de comunicación escrita en el programa de Diseño Industrial.

En cuanto al **DESEMPEÑO EN COMUNICACIÓN ESCRITA** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo, tabla 21.

Tabla 21 Resumen modelo óptimo socioeconómico

	Coef estimado	Error	Wald	P-valor	OR	I.C 95%	
Educación madre (Primaria)	-1.658	0.4384	-3.7825	0.0002	0.190	0.0806	0.4497
Educación madre (Secundaria)	-0.6688	0.3544	-1.8872	0.060	0.515	0.25579	1.0261
Educación madre (Sin estudio)	-2.5919	1.5765	-1.6441	0.101	0.064	0.00340	1.6453
Educación madre (Técnico)	-0.5255	0.5280	-0.9954	0.320	0.589	0.2100	1.6640
Tomó curso (Sí)	0.6625	0.3171	2.0890	0.037	1.907	1.0417	3.6113
Hogar actual (temporal)	0.4954	0.2774	1.786	0.075	1.632	0.9527	2.826
Nivel 2 — Nivel 3	-3.3766	0.4883	-6.9144				
Nivel 3 — Nivel 4	-1.8073	0.3745	-4.8256				
Nivel 4 — Nivel 5	-0.2268	0.3467	-0.6542				
Nivel 5 — Nivel 6	1.7056	0.3779	4.5129				
Nivel 6 — Nivel 7	3.2179	0.5123	6.2815				

Para la variable nivel educativo de la madre 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 del Nivel de desempeño 1 hasta el 8, decrece en todos los valores que puede tomar comunicación escrita, así un estudiante que tiene madre con nivel educativo primaria es menos probable que obtenga valores altos en el desempeño en comunicación escrita respecto a uno con madre de nivel educativo universitario.

Para la variable tomar algún curso de preparación al examen Saber Pro se observa que su razón de probabilidad y su intervalo de confianza al 95 % es mayor que uno, lo que significa que la probabilidad acumulada, comenzando del nivel de desempeño 1 hasta el 8, aumenta en todos los valores que puede tomar comunicación escrita, así un estudiante que tomó curso es más probable que obtenga valores altos en el desempeño de comunicación escrita respecto a uno que no lo tome.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **INGLÉS** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5..

```

-----
No link function: identity
No Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  10.21113    0.23271  43.879  <2e-16 ***
Educación_madre[T.Primaria] -0.45878    0.21600  -2.124  0.0354 *
Educación_madre[T.Secundaria] -0.42432    0.17107  -2.480  0.0143 *
Educación_madre[T.Sin estudio] -1.01965    0.49401  -2.064  0.0405 *
Educación_madre[T.Técnico] -0.05475    0.23930  -0.229  0.8193
Educación_padre[T.Primaria]  0.09235    0.20150  0.458  0.6474
Educación_padre[T.Secundaria]  0.05321    0.17365  0.463  0.6440
Educación_padre[T.Sin estudio] -0.83429    0.41168  -2.027  0.0446 *
Educación_padre[T.Técnico]  0.06259    0.19982  0.313  0.7546
ESTU_TITULO_BIO[T.B] -0.31825    0.32639  -0.975  0.3322
ESTU_TITULO_BIO[T.T] -0.24490    0.13375  -1.832  0.0691 .
Tomó_cursos[T.Si]  0.04789    0.15350  0.312  0.7555
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.5311    0.1288  -4.123  6.33e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No link function: log
No Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.2914    0.1331  2.159  0.0302 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Tau link function: log
Tau Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.18457    0.66452  1.783  0.0765 .
Educación_padre[T.Primaria] -0.64634    0.69436  -0.931  0.3535
Educación_padre[T.Secundaria] -0.85931    0.69716  -1.233  0.2198
Educación_padre[T.Sin estudio] -0.77926    1.18293  -0.659  0.5111
Educación_padre[T.Técnico]  0.09761    0.93528  0.104  0.9170
Tomó_cursos[T.Si]  0.46029    0.43853  1.050  0.2956
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 162
Degrees of Freedom for the fit: 20
Residual Deg. of Freedom: 142
          at cycle: 20

Global Deviance: 426.415
          AIC: 466.415
          SBC: 528.167

```

Figura 5-37: Modelo óptimo inglés

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo de la madre se tiene que, Con un nivel de significancia inferior al 1%, aquellos que tienen madres con nivel educativo secundaria obtienen, 0.42 puntos menos respecto a los que tienen madres con nivel educativo universitario. Con un nivel de significancia inferior al 5%, aquellos que tienen madres con nivel educativo primaria obtienen, 0.45 puntos menos respecto a los que tienen madres con nivel educativo universitario. Del mismo modo, aquellos estudiantes que tienen madres sin nivel educativo obtienen, 1.01 puntos menos respecto a los que tienen madres con nivel educativo universitario.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el nivel educativo del padre se tiene que, con un nivel de significancia inferior al 5%, aquellos que tienen padres sin nivel educativo obtienen, 0.83 puntos menos respecto a los que tienen padres con nivel educativo universitario.

Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el tipo de bachillerato del que se graduó se tiene que: Con un nivel de significancia inferior al 10%, que aquellos que tienen un tipo de bachillerato técnico obtienen, 0.24 puntos menos respecto a los que tienen un tipo de bachillerato académico.

Se puede decir que dadas los aspectos académicos considerados en este estudio no explican la variabilidad, asimetría y curtosis del puntaje en inglés.

De esta manera podemos decir que aspectos académicos como tener un nivel educativo universitario por parte del padre o de la madre, además de haberse graduado del colegio de tipo técnico influyen sobre el puntaje en inglés.

Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean = -0.02267662
      variance = 1.013525
      coef. of skewness = -0.2641611
      coef. of kurtosis = 3.170766
Filliben correlation coefficient = 0.9948758
-----
```

Figura 5-38: Prueba ajuste modelo académico inglés

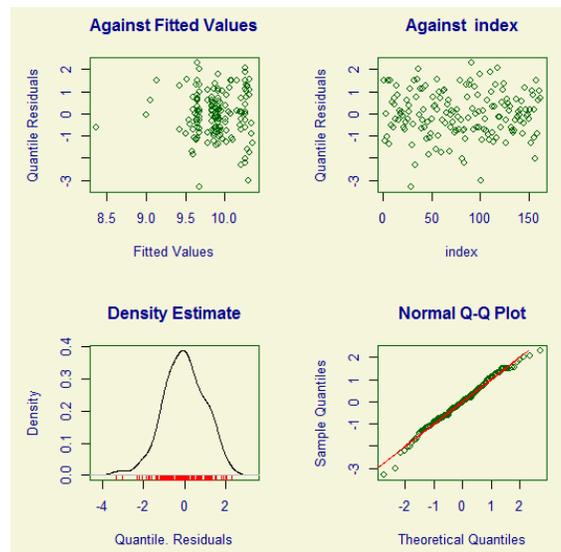


Figura 5-39: Diagnóstico modelo puntaje inglés factor académico

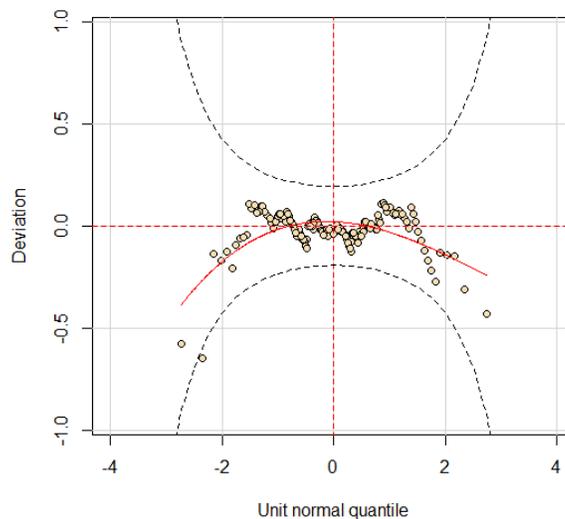


Figura 5-40: Ajuste de la distribución modelo inglés factor académico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-37) presentan un buen ajuste a una normal, esto quiere decir que la distribución skew t type 3 es la adecuada para describir el comportamiento del puntaje de inglés en el programa de Diseño Industrial.

En cuanto al **DESEMPEÑO EN INGLÉS** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo tabla 22.

Tabla 22 Resumen modelo óptimo académico

	Coef estimado	Error	Wald	P-valor	OR	I.C 95 %	
Educación madre (Primaria)	-0.9745	0.4186	-2.3281	0.020	0.372	0.1661	0.8571
Educación madre (Secundaria)	-0.6983	0.3480	-2.0069	0.046	0.495	0.2514	0.9837
Educación madre (Sin estudio)	-2.6148	1.3196	-1.9815	0.048	0.064	0.0055	0.9719
Educación madre (Técnico)	-0.0716	0.5186	-0.1381	0.890	0.927	0.3368	2.5723
A — A1	-2.3393	0.3345	-6.9940				
A1 — A2	-0.7728	0.2815	-2.7452				
A2 — B+	0.5261	0.2741	1.9195				
B+ — B1	0.7655	0.2782	2.7512				

Para la variable nivel educativo de la madre 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 del Nivel de desempeño A- hasta el B1, decrece en todos los valores que puede tomar inglés, así un estudiante que tiene madre con nivel educativo primaria es menos probable que obtenga valores altos en el desempeño en inglés respecto a uno con madre de nivel educativo universitario.

Para la variable nivel educativo de la madre 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 del Nivel de desempeño A- hasta el B1, decrece en todos los valores que puede tomar inglés, así un estudiante que tiene madre con nivel educativo secundaria es menos probable que obtenga valores altos en el desempeño en inglés respecto a uno con madre de nivel educativo universitario.

Para la variable nivel educativo de la madre 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 del Nivel de desempeño A- hasta el B1, decrece en todos los valores que puede tomar inglés, así un estudiante que tiene madre sin estudio es menos probable que obtenga valores altos en el desempeño en inglés respecto a uno con madre de nivel educativo universitario.

5.4 FACTORES SOCIOECONÓMICOS ASOCIABLES AL COMPONENTE GENÉRICO

La siguiente Tabla 23 muestra las variables explicativas a tener en cuenta para los modelos socioeconómicos.

Tabla 23 componente sistemático socioeconómico

Notación	Nombre
X_1	Número de dormitorios.
X_2	Estrato de la residencia según factura de energía
X_3	Valor anual de la matrícula
X_4	Formas de pagó matrícula
X_5	Ocupación del padre
X_6	Ocupación de la madre
X_7	El hogar cuenta con celular
X_8	El hogar cuenta con internet
X_9	El hogar cuenta con televisión
X_{10}	El hogar cuenta con lavadora
X_{11}	El hogar cuenta con teléfono fijo
X_{12}	Trabaja actualmente
X_{13}	Horas que trabaja a la semana
X_{14}	Ingresos mensuales en salarios mínimos

El modelo lineal queda determinado por el siguiente predictor lineal

$$\eta_i = \beta_0 + \beta_1 X_1 + \sum_{jp=2}^{14} \tau_{jk} X_p \quad (5-3)$$

Donde.

β_0 = Intercepto

β_1 = Efecto del número de dormitorios.

$\tau_2 k$ = Efecto del estrato de la residencia según factura de energía, con $k = 1, 2, 3, 4, 5, 6$

$\tau_3 k$ = Efecto del valor anual de la matrícula, con $k = 0, 1, 2, 3, 4, 5$

$\tau_4 k$ = Efecto de las formas de pagó matrícula, con $k = 0, 1, 2, 3, 4, 5$

$\tau_5 k$ = Efecto de la ocupación del padre, con $k = 1, 2, 3, 4, 5$

$\tau_6 k$ = Efecto de la ocupación de la madre, con $k = 1, 2, 3, 4, 5$

$\tau_7 k$ = Efecto del celular, con $k = 0, 1$

τ_8k = Efecto del internet, con $k = 0, 1$

τ_9k = Efecto de la televisión, con $k = 0, 1$

$\tau_{10}k$ = Efecto del la lavadora, con $k = 0, 1$

$\tau_{11}k$ = Efecto del teléfono fijo, con $k = 0, 1$

$\tau_{12}k$ = Efecto del trabajo, con $k = 0, 1, 2, 3, 4, 5$

$\tau_{13}k$ = Efecto del número de horas que trabaja a la semana, con $k = 0, 1, 2, 3, 4$

$\tau_{14}k$ = Efecto de los Ingresos mensuales representado en salarios mínimos, con $k = 1, 2, 3, 4, 5, 6$

Para la selección y mejor ajuste del modelo óptimo en cuanto a **RAZONAMIENTO CUANTITATIVO** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function: identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.98260	0.22848	43.692	< 2e-16 ***
Valor_anual_matricula_cod[T.Entre 3 millones y 5 millones]	0.84433	0.39232	2.152	0.03274 *
Valor_anual_matricula_cod[T.Entre 500 mil y menos de de 1 millón]	0.32360	0.63676	0.508	0.61195
Valor_anual_matricula_cod[T.Menos de 500 mil]	-0.49889	0.21936	-2.274	0.02415 *
Valor_anual_matricula_cod[T.No pagó matricula]	-0.03509	0.34051	-0.103	0.91803
Forma_de_pagó_matricula_cod[T.Pagó por beca]	-0.11029	0.23516	-0.469	0.63965
Forma_de_pagó_matricula_cod[T.Pagó por credito]	0.24595	0.31487	0.781	0.43579
Forma_de_pagó_matricula_cod[T.Pagó propio]	0.24502	0.25601	0.957	0.33983
Forma_de_pagó_matricula_cod[T.Varias formas de pagó]	0.26748	0.26235	1.020	0.30933
Ocupación_madre[T.Obrero]	0.15448	0.17750	0.870	0.38532
Ocupación_madre[T.Ocupación no remunerada]	-0.05262	0.13663	-0.385	0.70062
Ocupación_madre[T.Otra actividad u ocupación]	-0.06003	0.22487	-0.267	0.78981
Ocupación_madre[T.Pensionado]	0.04000	0.22033	0.182	0.85616
Ocupación_padre[T.Obrero]	-0.09856	0.14422	-0.683	0.49524
Ocupación_padre[T.Ocupación no remunerada]	0.15110	0.35853	0.421	0.67394
Ocupación_padre[T.Otra actividad u ocupación]	-0.04769	0.20598	-0.232	0.81718
Ocupación_padre[T.Pensionado]	0.03079	0.17127	0.180	0.85754
El_hogar_cuenta_con_internet_cod[T.Si]	0.39335	0.11702	3.362	0.00095 ***

```

---
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)	-0.84383	0.06004	-14.06	<2e-16 ***

```

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

-----
No. of observations in the fit: 196
Degrees of Freedom for the fit: 19
Residual Deg. of Freedom: 177
at cycle: 3

Global Deviance: 454.5232
AIC: 492.5232
SBC: 554.8074

```

Figura 5-41: Modelo óptimo razomaniento cuantitativo

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en el valor anual de la matrícula del año anterior se tiene que con un nivel de significancia inferior al 5 %, aquellos que pagan entre 3 y 5 millones de pesos obtienen, 0.84 puntos más que los que pagan entre 1 y 3 millones de pesos. Del mismo modo, aquellos estudiantes que pagan menos de 500 mil pesos obtienen, 0.49 puntos menos que los que pagan entre 1 y 3 millones de pesos.

Ahora bien, con un nivel de significancia inferior al 1 %, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si cuenta o no cuenta con internet, se tiene que aquellos que cuentan con internet obtienen 0.39 puntos más que los que no cuentan con internet.

Se puede decir que dados los aspectos socioeconómicos considerados en este estudio no explican la variabilidad del puntaje en razonamiento cuantitativo.

De este modo podemos decir que aspectos socioeconómicos como pagar entre 3 y 5 millones de pesos y menos de 500 mil pesos, además de contar con internet, influyen sobre el valor esperado del puntaje en razonamiento cuantitativo.

Una vez ajustado el modelo GAMLSS, se utilizaran los residuales cuantiles aleatorizados descritos en la sección 2.5 para comprobar la adecuación del modelo y, más específicamente, la distribución escogida para la variable respuesta.

Las siguientes figuras 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.

```
> plot(MO)
*****
Summary of the Quantile Residuals
      mean      = -0.003304185
      variance   =  1.011686
      coef. of skewness = -0.09985602
      coef. of kurtosis  =  3.016073
Filliben correlation coefficient =  0.9971549
*****
```

Figura 5-42: Prueba ajuste modelo socioeconómico razonamiento cuantitativo

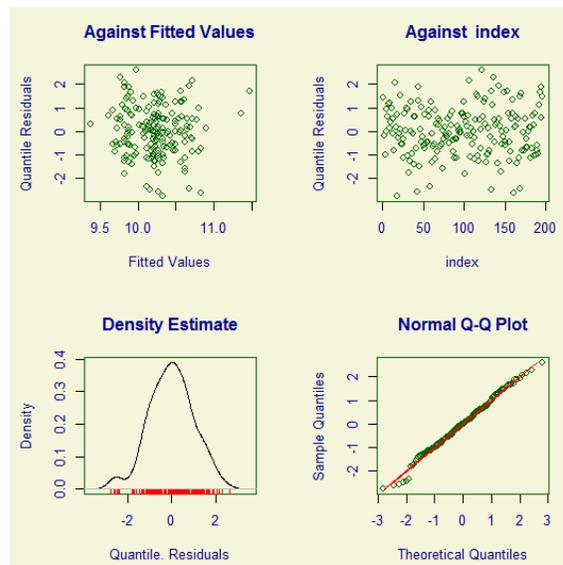


Figura 5-43: Diagnóstico modelo puntaje razonamiento cuantitativo factor socioeconómico

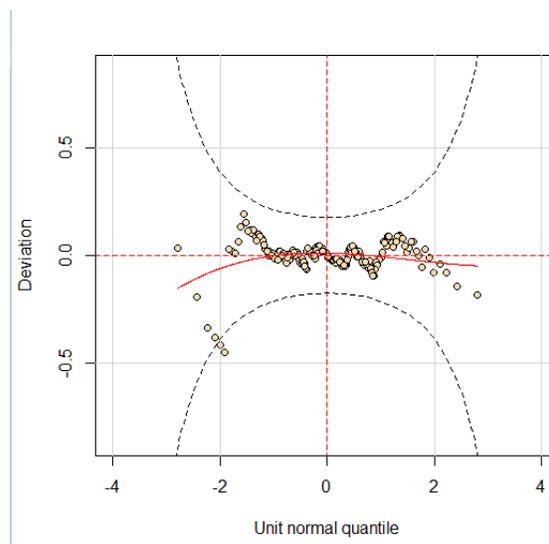


Figura 5-44: Ajuste de la distribución modelo razonamiento cuantitativo factor socioeconómico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-41) presentan un buen ajuste a una normal, esto quiere decir que la distribución Logistic es la adecuada para describir el comportamiento del puntaje de razonamiento cuantitativo en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **LECTURA CRÍTICA** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function:  log
Mu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.340186  0.004948  472.9  <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) -2.66967  0.05047  -52.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  196
Degrees of Freedom for the fit:  2
  Residual Deg. of Freedom:  194
                        at cycle:  2

Global Deviance:  426.4393
      AIC:  430.4393
      SBC:  436.9955
*****

```

Figura 5-45: Modelo óptimo lectura crítica

Se tiene que las variables socioeconómicas, consideradas en este estudio tales como ocupación de los padres, tener internet, celular, forma de pagó de la matrícula entre otras, no explican el puntaje obtenido en lectura crítica. Es decir estos aspectos no influyen en el valor esperado ni en la variabilidad del puntaje en lectura crítica.

Las siguientes figuras 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.

```

> plot(M0)
Warning in plot.window(...) :
  amplitud relativa de valores = 43 * EPS, es pequeño (eixo 1)
*****
      Summary of the Quantile Residuals
      mean      = 2.016958e-05
      variance  = 1.005129
      coef. of skewness = -0.1504689
      coef. of kurtosis  = 2.920274
      Filliben correlation coefficient = 0.9963503
*****

```

Figura 5-46: Prueba ajuste modelo socioeconómico lectura crítica

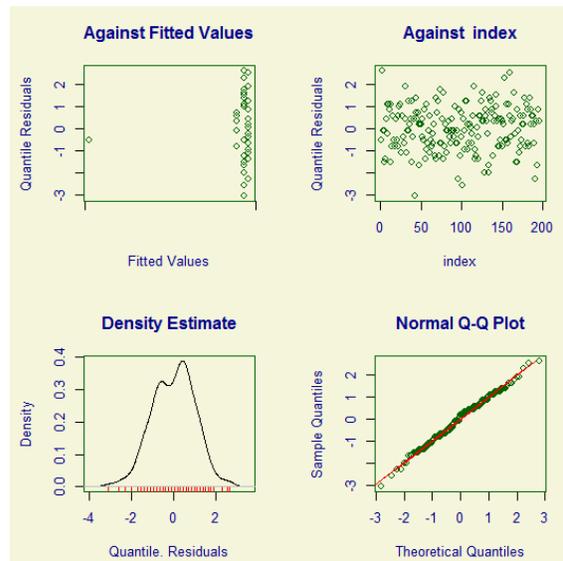


Figura 5-47: diagnóstico modelo puntaje lectura crítica factor socioeconómico

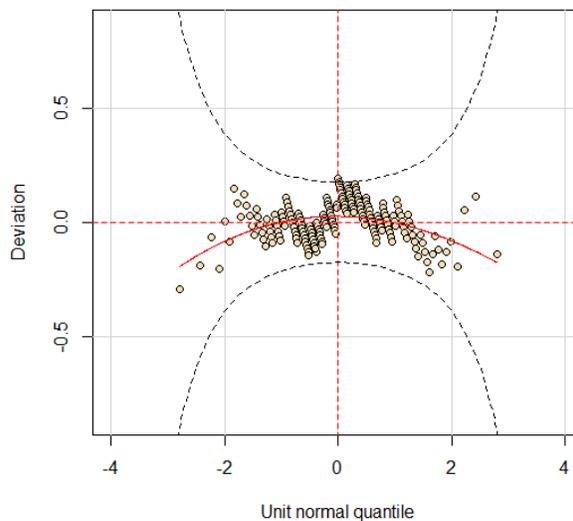


Figura 5-48: Ajuste de la distribución modelo lectura crítica factor socioeconómico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-45) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma es la adecuada para describir el comportamiento del puntaje de lectura crítica en el programa de Diseño Industrial.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **COMUNICACIÓN ESCRITA** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```

Mu link function: log
Mu Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    2.318351    0.025845  89.700 < 2e-16 ***
Ocupación_padre[T.Obrero]      -0.033229    0.016188  -2.053 0.04154 *
Ocupación_padre[T.Ocupación no remunerada] -0.060500    0.037695  -1.605 0.11023
Ocupación_padre[T.Otra actividad u ocupación] 0.033437    0.022596   1.480 0.14066
Ocupación_padre[T.Pensionado]  -0.027612    0.019021  -1.452 0.14832
El_hogar_cuenta_con_internet_cod[T.Si]      0.033136    0.013073   2.535 0.01210 *
Estudiante_trabaja_cod[T.Si, como ayudante sin remuneración] -0.014590    0.015063  -0.969 0.33403
Estudiante_trabaja_cod[T.Si, por experiencia o recursos en gastos personales] -0.054234    0.018213  -2.979 0.00329 **
Estudiante_trabaja_cod[T.Si, por ser práctica obligatoria del plan de estudios] -0.183816    0.064212  -2.863 0.00469 **
ESTU_DORMITORIOS                0.009801    0.006715   1.459 0.14618
---
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)                    -2.43291    0.06466  -37.63 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: identity
Nu Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     7.890     2.617   2.992 0.00315 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 194
Degrees of Freedom for the fit: 12
Residual Deg. of Freedom: 182
as cycle: 16

Global Deviance: 531.6013
AIC: 555.6013
SBC: 594.8156

```

Figura 5-49: Modelo óptimo comunicación escrita

A partir de la información del modelo óptimo se puede afirmar que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la ocupación del padre se tiene que, con un nivel de significancia inferior al 5 %, aquellos padres que se ocupan como obrero obtienen, 1.02 puntos menos que los que se ocupan como administradores.

Ahora bien, con un nivel de significancia inferior al 1 %, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si cuenta o no cuenta con internet, se tiene que aquellos que cuentan con internet obtienen 1.02 puntos más que los que no cuentan con internet. Por otra parte, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si el estudiante Trabaja actualmente se tiene que, con un nivel de significancia inferior al 5 %, aquellos que trabajan por experiencia o recursos en gastos personales obtienen, 0.94 puntos menos que los que trabajan con remuneración en dinero. Del mismo modo aquellos que trabajan por ser práctica obligatoria del plan de estudios obtienen, 0.82 puntos menos que los que trabajan con remuneración en dinero.

Se puede decir que dados los aspectos socioeconómicos considerados en este estudio no explican la variabilidad del puntaje en comunicación escrita.

De esta manera podemos decir que aspectos socioeconómicos como tener una ocupación de administrador por parte del padre, además de contar con internet y tener un trabajo con remuneración influyen sobre el valor esperado del puntaje en comunicación escrita.

Las siguientes figuras 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.

```
> plot(M0)
*****
Summary of the Quantile Residuals
      mean = 0.0006191243
      variance = 1.004471
      coef. of skewness = 0.005877601
      coef. of kurtosis = 2.655292
Filliben correlation coefficient = 0.9976593
.....
```

Figura 5-50: Prueba ajuste modelo socioeconómico comunicación escrita

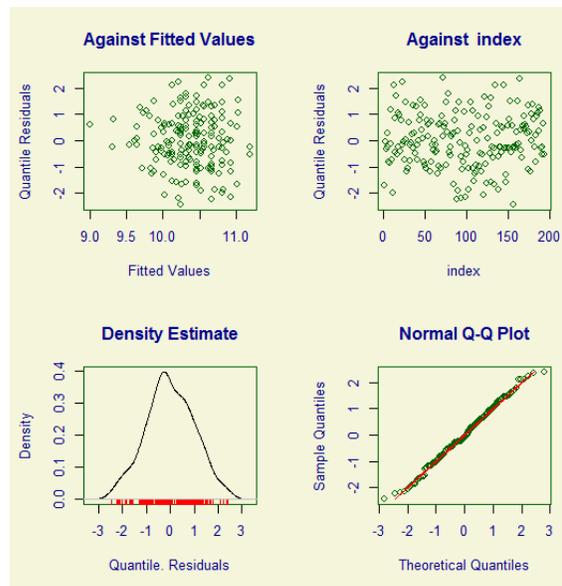


Figura 5-51: Diagnóstico modelo puntaje comunicación escrita factor socioeconómico

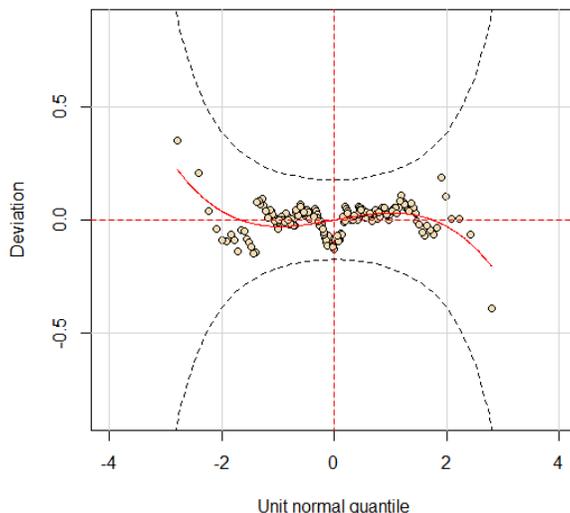


Figura 5-52: Ajuste de la distribución modelo puntaje comunicación escrita factor socio-económico

En las figuras anteriores podemos observar que los residuales aleatorizados para el modelo óptimo (figura 5-49) presentan un buen ajuste a una normal, esto quiere decir que la distribución Gamma Generalizada es la adecuada para describir el comportamiento del puntaje de comunicación escrita en el programa de Diseño Industrial.

En cuanto al **DESEMPEÑO EN COMUNICACIÓN ESCRITA** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo, tabla 24.

Tabla 24 Resumen modelo óptimo

	Coef estimado	Error	Wald	P-valor	OR	I.C 95 %	
Estudiante trabaja (Sin remuneración)	-0.2508	0.3008	-0.8338	0.405	0.764	0.4316	1.4030
Estudiante trabaja (Por experiencia)	-0.7194	0.3533	-2.0365	0.043	0.477	0.2437	0.9733
Estudiante trabaja (práctica obligatoria)	-1.8624	1.1594	-1.6064	0.109	0.150	0.0160	1.5067
Núm. dormitorios de la residencia	0.2709	0.1313	2.0633	0.040	1.293	1.0136	1.6958
Nivel 2 — Nivel 3	-2.6761	0.5360	-4.9927				
Nivel 3 — Nivel 4	-1.0509	0.4437	-2.3684				
Nivel 4 — Nivel 5	0.5802	0.4357	1.3318				
Nivel 5 — Nivel 6	2.2731	0.4654	4.8838				
Nivel 6 — Nivel 7	4.0811	0.5945	6.8650				

Para la variable estudiante trabaja 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 del Nivel de desempeño 1 hasta el 8, decrece en todos los valores que puede tomar

comunicación escrita, así un estudiante que trabaja por experiencia es menos probable que obtenga valores altos en el desempeño en comunicación escrita respecto a uno que no trabaje.

Para la variable número de dormitorios de la residencia se observa que su razón de probabilidad y su intervalo de confianza al 95 % es mayor que uno, lo que significa que la probabilidad acumulada, comenzando del Nivel de desempeño 1 hasta el 8, aumenta en todos los valores que puede tomar comunicación escrita, así un estudiante por cada dormitorio de más tiene mayor probabilidad de obtener valores altos en el desempeño en comunicación escrita.

Para la selección y mejor ajuste del modelo óptimo en cuanto a **INGLÉS** se tuvo en cuenta el criterio StepGAICALL.A descritos en la sección 2.5.

```
-----
Mu link function: identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.7619    0.1025  95.214  <2e-16 ***
El_hoger_cuenta_con_internet_cod[T.Si]  0.2524    0.1063   2.375  0.0185 *
---
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)  -0.50274    0.09715  -5.175  5.74e-07 ***
---
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)   0.28625    0.09431   3.035  0.00274 **
---
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)   0.9724    0.2122   4.582  8.3e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 196
Degrees of Freedom for the fit: 5
Residual Deg. of Freedom: 191
                        at cycle: 16

Global Deviance:    533.4674
AIC:                543.4674
SBC:                559.858
```

Figura 5-53: Modelo óptimo inglés

A partir de la información del modelo óptimo se puede afirmar respecto al valor esperado del puntaje en inglés que con un nivel de significancia inferior al 1 %, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo si cuenta o no cuenta con internet, se tiene que aquellos que cuentan con internet obtienen 0.25 puntos más que los que no cuentan con internet.

Se puede decir que dados los aspectos socioeconómicos considerados en este estudio tales como ocupación de los padres, tener coexión a internet, forma de pagó de la matrícula entre otras, no explican la variabilidad del puntaje en inglés.

Las siguientes figuras 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.

```
> plot(M0)
*****
Summary of the Quantile Residuals
      mean = -0.01760816
      variance = 1.008102
      coef. of skewness = -0.3328131
      coef. of kurtosis = 3.37138
Filliben correlation coefficient = 0.9900298
.....
```

Figura 5-54: Prueba ajuste modelo socioeconómico inglés

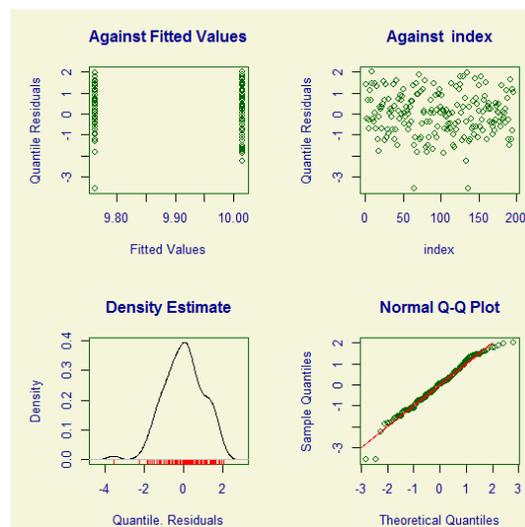


Figura 5-55: Diagnóstico modelo puntaje inglés factor socioeconómico

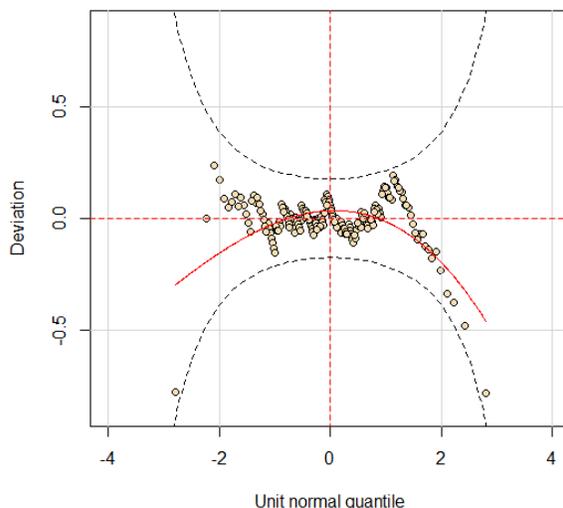


Figura 5-56: Ajuste de la distribución modelo inglés factor socioeconómico

En las figuras anteriores se observa que los residuales aleatorizados para el modelo óptimo (figura 5-53) presentan un buen ajuste a una normal, esto quiere decir que la distribución skew t type 3 es la adecuada para describir el comportamiento del puntaje de inglés en el programa de Diseño Industrial.

En cuanto al **DESEMPEÑO EN INGLÉS** se aplicó regresión logística ordinal. A continuación se presenta el modelo óptimo, tabla 25.

Tabla 25 Resumen modelo óptimo

	Coef estimado	Error	Wald	P-valor	OR	I.C 95%	
El hogar cuenta con internet (sí)	0.516	0.2605	1.98	0.049	1.652	1.0053	2.7914
A— A1	-1.5269	0.2420	-6.3088				
A1 — A2	-0.0008	0.2045	-0.0038				
A2 — B+	1.2488	0.2246	5.5603				
B+ — B1	1.4634	0.2315	6.3221				

Para la variable el hogar cuenta con conexión a internet se observa que su razón de probabilidad y su intervalo de confianza al 95 % es mayor que uno, lo que significa que la probabilidad acumulada, comenzando del Nivel de desempeño A- hasta el B1, aumenta en todos los valores que puede tomar inglés, así un estudiante que tiene conexión a internet es más probable que obtenga valores altos en el desempeño de inglés respecto a uno que no tenga conexión.

6 DISCUSIÓN

La Tabla 26, 27 y 28 muestra los factores que influyen para cada componente, siendo RC, razonamiento cuantitativo, LC, lectura crítica, CE, comunicación escrita, IN, inglés, D.CE, desempeño comunicación escrita, D.IN, desempeño inglés, ↑, indica que la variable influye a la hora de obtener puntajes altos. Por otra parte ↓ indica la posibilidad de obtener valores bajos, -, indica que la variable no hace parte, o no resultado significativa en el modelo óptimo.

Tabla 26 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)	↑	-	↑	-	-	-
Núm personas a cargo	↑	↑	-	-	-	-

De acuerdo a la Tabla 26, se observa que ser soltero, estar en unión libre o separado se obtiene mejores puntajes en los componentes de razonamiento cuantitativo y comunicación escrita en comparación a estudiantes casados, lo cual esta de acuerdo con los planteamiento de (Ramírez 2014b) “para las mujeres, el estado civil se asocia positivamente, aunque de forma débil con su desempeño; ser soltera se relaciona con un mejor desempeño; esto se podría explicar así: las mujeres solteras tendrían más tiempo para dedicarle al estudio que el que le pueden dedicar las mujeres casadas, quienes tienen que alternar su papel como estudiante con el de esposa y madre”, en cuanto al estado civil de los hombres existe un desacuerdo con los

resultados de (Ramírez 2014b) quien determina que, “para los hombres el estado civil resulta no asociarse con su desempeño”. Además contar con un hogar temporal también aumenta la posibilidad de tener puntajes altos en comparación a quienes cuentan con un hogar permanente.

Por otra parte, a menor edad y tener personas a cargo aumenta la posibilidad de obtener puntajes altos en razonamiento cuantitativo y lectura crítica, estando en desacuerdo con los resultados de (Ramírez 2014b) quien determina que, “la edad se asocia débilmente, en forma negativa con su desempeño académico”. Los hombres obtienen mejores puntajes en los componentes de razonamiento cuantitativo y comunicación escrita en comparación a las mujeres, lo cual esta de acuerdo con los planteamientos de (Ramírez, 2014b) quien afirma que “para las mujeres, la asociación de acuerdo con el área del examen varía; por un lado, lenguaje y biología están asociadas fuertemente; mientras que matemáticas e idiomas se encuentran asociadas débilmente”. En cuanto al desempeño en comunicación escrita, ser soltero es más probable que obtenga valores altos respecto a uno casado.

Tabla 27 Resumen variables académicas influyentes

VARIABLES ACADÉMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Curso de preparación al examen (Sí)	-	-	↑	-	↑	-
Curso de preparación al examen (No)	-	-	↓	-	↓	-
Tipo de bachillerato (Técnico)	↑	↑	-	↓	-	-
Tipo de bachillerato (Académico)	↓	↓	-	↑	-	-
Tipo de bachillerato (Normalista superior)	-	↑	-	-	-	-
Nivel educativo padre (Universitario)	↑	↑	-	↑	-	-
Nivel educativo padre (Sin estudio)	↓	↓	-	↓	-	-
Nivel educativo padre (Técnico)	-	↓	-	-	-	-
Nivel educativo madre (Universitario)	↑	↑	↑	↑	↑	↑
Nivel educativo madre (Primaria)	↓	↓	↓	↓	↓	↓
Nivel educativo madre (Secundaria)	↓	↓	↓	↓	-	↓
Nivel educativo madre (Sin estudio)	↓	-	↓	↓	-	↓
Nivel educativo madre (Técnico)	-	↓	↓	-	-	-

De acuerdo a la Tabla 27, se observa que tomar algún curso de preparación al examen Saber Pro, aumenta la posibilidad de obtener puntajes altos en el componente de comunicación escrita y su desempeño.

Se sugiere que el programa realice simulacros en cuanto a las pruebas del componente genérico ya que como se dio a conocer el tomar algún curso de preparación al examen aumenta la probabilidad de obtener buenos resultados, además ayudará a concientizar al programa si lo que evalúa es realmente lo que se está evaluando en las pruebas Saber Pro.

Por otra parte, tener madre con nivel educativo universitario obtiene mejores puntajes en todo el componente genérico, en comparación a tener niveles educativos de primaria, secundaria, técnica o sin estudios, asimismo tener padre con nivel educativo universitario aumenta el puntaje en razonamiento cuantitativo, lectura crítica e inglés, lo cual está de acuerdo con los planteamientos de (Ramírez, 2014b) quien analizó algunas diferencias por nivel de formación evidenciando que, *en el nivel de formación técnico profesional, el nivel de estudios de la madre, específicamente el nivel técnico y tecnológico, tiene una asociación positiva y débil con el desempeño académico cuando se compara con estudiantes cuyas madres alcanzaron algún tipo de estudios de bachillerato. Por el contrario, para los estudiantes de programas profesionales, la educación de los padres resulta asociada débil y positivamente con su desempeño académico, particularmente en aquellos estudiantes cuyos padres alcanzaron algún nivel de posgrado en vez de algún tipo de estudios de bachillerato.* Además, obtener un tipo de bachillerato académico disminuye la posibilidad de lograr puntajes altos respecto a un bachiller de tipo técnico, lo cual está de acuerdo con lo planteado por (Ramírez, 2014b) quien determina que solo para los hombres, tener un bachillerato técnico en vez de uno académico se asocia positivamente y de manera débil con su desempeño, caso contrario se presenta en inglés.

Tabla 28 Resumen variables socioeconómicas influyentes

VARIABLES SOCIOECONÓMICAS	COMPONENTE GENÉRICO					
	RC	LC	CE	IN	D.CE	D.IN
Núm de dormitorios	-	-	-	-	↑	-
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 padre (Administrador)	-	-	↑	-	-	-
Ocupación padre (Obrero)	-	-	↓	-	-	-
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)	-	-	-	-	↑	-

De acuerdo a la Tabla 28, se observa que realizar pagos de matrícula entre 3 y 5 millones de pesos, aumenta la posibilidad de obtener puntajes altos en razonamiento cuantitativo, en comparación a quienes realizan pagos entre menos de 500 mil y 3 millones de pesos. Por otra parte, al tener conexión a internet se obtienen mejores puntajes en el componente de razonamiento cuantitativo, comunicación escrita, inglés y su desempeño.

En cuanto a la ocupación del padre como administrador obtiene más puntos en comunicación escrita respecto a un padre con ocupación como obrero, así también un estudiante que trabaje con remuneración en dinero obtiene mejores puntajes respecto a estudiantes que trabajen por experiencia o práctica obligatoria. Por cada dormitorio de más que tenga el estudiante, aumenta la probabilidad de obtener puntajes altos para el desempeño en comunicación escrita, asimismo un estudiante que no trabaje presenta mejores puntaje respecto a uno que si trabaje por experiencia.

Se recomienda, que el análisis efectuado con base en los resultados obtenidos en las pruebas Saber Pro, aplicadas a los estudiantes del programa de Diseño Industrial, Facultad seccional Duitama de la Universidad Pedagógica y Tecnológica de Colombia sirva para desarrollar estrategias al interior del programa, asimismo para autoevaluarse y tomar acciones de mejoramiento.

Dentro de las estrategias en que puede intervenir a corto plazo la universidad, para la mejora en los resultados del componente genérico se sugiere: realizar talleres para docentes y directivos sobre las características de las pruebas, así como, informar y concientizar a los estudiantes sobre lo pertinente e importancia de la prueba, la estructura de la misma y los beneficios que se pueden obtener de acuerdo a los resultados.

Dado que en los aspectos demográficos la Universidad no podría actuar, si se recomienda que desde el aspecto socioeconómico como lo es tener conexión a internet se aumente la cobertura en este campo, ya que este servicio influye sobre el aumento de la probabilidad de obtener buenos resultados.

7 CONCLUSIONES

Es importante recalcar que los resultados aquí presentados solo hacen referencia a los estudiantes de Diseño Industrial de la UPTC objeto de este estudio y que debido a las características de esta, no hace generalizaciones para la población colombiana.

En la población objeto de análisis se encontró que tienen un promedio de edad de 27 años de edad, cuyo estado civil predominante es soltero, además de ser en su mayoría de género masculino. Otros detalles relevantes muestran que no son cabeza de familia, y no tienen personas a cargo, su núcleo familiar esta conformado en promedio por 4 personas, viviendo por lo general en un hogar permanente, son de estrato dos y tres, cuyo valor de matrícula predominante es entre 1 millón y 3 millones de pesos y en general la forma de pagó es a través de los padres, en su mayoría presentan algún tipo de preparación previo para las pruebas Saber Pro y el tipo de bachillerato es académico.

El máximo nivel académico alcanzado por los padres es secundaria y educación profesional es el más común. La ocupación del padre y la madre se presenta en la modalidad de trabajo por cuenta propia, tienen servicio cerrado de televisión, lavadora, internet y celular, no cuentan con servicio de teléfono fijo, los ingresos familiares mensuales se encuentran entre 1 y menos de 2 salarios mínimos, el 50% de los estudiantes no trabaja, pero a excepción de algunos estudiantes el tiempo de trabajo está entre 16 a 20 horas semanales y cuentan con un promedio de 3 dormitorios en su hogar.

Se identificó correlación significativa entre el número de personas que conforman su hogar con el puntaje en inglés, también se encontró diferencia entre los promedios según género y ser cabeza de familia en razonamiento cuantitativo. Asimismo se observó diferencia entre los promedios en comunicación escrita por estado civil y ser cabeza de familia. Respecto a razonamiento cuantitativo y comunicación escrita se encontró diferencia entre los promedios al tener conexión a internet. Por otra parte, se determinó correlación entre el puntaje en inglés y tener conexión a internet. También se identificó diferencia entre los promedios en comunicación escrita con educación de sus padres.

Respecto a el puntaje de razonamiento cuantitativo se determinó, que las variables que afectan dicho puntaje son realizar un pago de matrícula entre 3 y 5 millones de pesos, tener conexión a internet, ser soltero, unión libre, tener un hogar temporal, edad, género, tener madre con nivel educativo universitario y tipo de bachillerato técnico. lo cual esta de acuerdo con los planteamiento de Gil, F., Sepúlveda, L., Rondón, M., Gómez, C 2013.

En cuanto a lectura crítica las variables que influyeron son la edad, tener personas a cargo, tener madre con nivel educativo Universitario y tipo de bachillerato técnico, obteniendo más puntos en dichas pruebas. Por otra parte el factor socioeconómico no explica el puntaje obtenido en lectura crítica.

En cuanto a inglés las variables que influyeron son tener conexión a internet, tener madre con nivel educativo universitario y tipo de bachillerato técnico, lo cual esta de acuerdo con los planteamiento de Arias 2014. Por otra parte, el factor demográfico no explica el puntaje obtenido en inglés.

En cuanto a comunicación escrita las variables que influyeron son trabajar con remuneración en dinero, ocupación del padre como administrador, tener madre con nivel educativo universitario, tomar algún curso de preparación al examen Saber Pro, ser soltero, unión libre, tener un hogar temporal, edad y género. Dada la complejidad de esta investigación no se encontró referentes teóricos que sustentaran algunas de las variables influyentes, siendo este trabajo monográfico innovador debido a que explica como afecta cada factor a cada puntaje en específico.

Considerando lo anterior se puede decir que el modelamiento es una herramienta que nos permitió explicar aquellas variables demográficas, académicas y socioeconómicas influyentes en los resultados del componente genérico de las pruebas Saber Pro, a través de diversas técnicas estadísticas como fueron los GAMLSS y RLO que van más allá de lo descriptivo.

8 ANEXO A

Resultados estadísticos análisis bivariado

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 la correlación entre algunas variables con un nivel de significancia del 5%.

1. Cruzando Edad vs Razonamiento cuantitativo

```
> with(Dataset, cor.test(EDAD, MOD_RAZO_CUANTITATIVO_PUNT, alternative="two.sided",  
+ method="pearson"))
```

Pearson's product-moment correlation

```
data: EDAD and MOD_RAZO_CUANTITATIVO_PUNT  
t = -1.6525, df = 195, p-value = 0.1001  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.25315192 0.02265285  
sample estimates:  
cor  
-0.117515
```

2. Cruzando Número de personas de las que usted se encuentra a cargo vs Razonamiento cuantitativo

```
> with(Dataset, cor.test(FAMI_NUM_PERS_CARGO, MOD_RAZONA_CUANTITATIVO_PUNT,  
+ alternative="two.sided", method="pearson"))
```

Pearson's product-moment correlation

```
data: FAMI_NUM_PERS_CARGO and MOD_RAZONA_CUANTITATIVO_PUNT  
t = 0.46671, df = 196, p-value = 0.6412  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1066190 0.1719604  
sample estimates:  
cor  
0.03331782
```

3. Cruzando Número de personas que conforman su hogar vs Razonamiento cuantitativo

```
> with(Dataset, cor.test(FAMI_NUM_PERS_GRUP_FAM, MOD_RAZONA_CUANTITATIVO_PUNT,  
+ alternative="two.sided", method="pearson"))
```

Pearson's product-moment correlation

```
data: FAMI_NUM_PERS_GRUP_FAM and MOD_RAZONA_CUANTITATIVO_PUNT  
t = 0.57539, df = 196, p-value = 0.5657  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.09894312 0.17947868  
sample estimates:  
cor  
0.04106493
```

4. Cruzando Número de dormitorios vs Desempeño en Inglés

```
> with(Dataset, cor.test(ESTU_DORMITORIOS, MOD_RAZONA_CUANTITATIVO_PUNT,  
+ alternative="two.sided", method="pearson"))
```

Pearson's product-moment correlation

```
data: ESTU_DORMITORIOS and MOD_RAZONA_CUANTITATIVO_PUNT  
t = 0.65577, df = 196, p-value = 0.5127  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.09326063 0.18502363  
sample estimates:  
cor  
0.04678931
```

5. Cruzando Edad vs Lectura crítica

```
> with(Dataset, cor.test(EDAD, MOD_LECTURA_CRITICA, alternative="two.sided",  
+ method="pearson"))
```

Pearson's product-moment correlation

```
data: EDAD and MOD_LECTURA_CRITICA  
t = -0.52027, df = 195, p-value = 0.6035  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1761106 0.1031008  
sample estimates:  
cor  
-0.03723153
```

6. Cruzando Número de personas de las que usted se encuentra a cargo vs Lectura crítica

```
> with(Dataset, cor.test(FAMI_NUM_PERS_CARGO, MOD_LECTURA_CRITICA, alternative="two.sided",  
+ method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: FAMI_NUM_PERS_CARGO and MOD_LECTURA_CRITICA  
t = 0.41421, df = 196, p-value = 0.6792  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1103232 0.1683206  
sample estimates:  
cor  
0.0295732
```

7. Cruzando Número de personas que conforman su hogar vs Lectura crítica

```
> with(Dataset, cor.test(FAMI_NUM_PERS_GRUP_FAM, MOD_LECTURA_CRITICA, alternative="two.sided",  
+ method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: FAMI_NUM_PERS_GRUP_FAM and MOD_LECTURA_CRITICA  
t = 0.20869, df = 196, p-value = 0.8349  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1247963 0.1540260  
sample estimates:  
cor  
0.01490459
```

8. Cruzando Número de dormitorios vs Lectura crítica

```
> with(Dataset, cor.test(ESTU_DORMITORIOS, MOD_LECTURA_CRITICA, alternative="two.sided",  
+ method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: ESTU_DORMITORIOS and MOD_LECTURA_CRITICA  
t = -0.43815, df = 196, p-value = 0.6618  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1699814 0.1086339  
sample estimates:  
cor  
-0.0312814
```

9. Cruzando Edad vs Comunicación escrita

```
> with(Dataset, cor.test(EDAD, MOD_COMUNICA_ESCRITA_PUNT, alternative="two.sided",  
+ method="pearson"))
```

Pearson's product-moment correlation

```
data: EDAD and MOD_COMUNICA_ESCRITA_PUNT  
t = -0.91065, df = 193, p-value = 0.3636  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.2040463 0.0757997  
sample estimates:  
cor  
-0.06540932
```

10. Cruzando Número de personas de las que usted se encuentra a cargo vs Comunicación escrita

```
> with(Dataset, cor.test(FAMI_NUM_PERS_CARGO, MOD_COMUNICA_ESCRITA_PUNT,  
+ alternative="two.sided", method="pearson"))
```

Pearson's product-moment correlation

```
data: FAMI_NUM_PERS_CARGO and MOD_COMUNICA_ESCRITA_PUNT  
t = -1.6699, df = 194, p-value = 0.09655  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.2549399 0.0214703  
sample estimates:  
cor  
-0.1190406
```

11. Cruzando Número de personas que conforman su hogar vs Comunicación escrita

```
> with(Dataset, cor.test(FAMI_NUM_PERS_GRUP_FAM, MOD_COMUNICA_ESCRITA_PUNT, alternative="two.sided",  
+ method="pearson"))
```

Pearson's product-moment correlation

```
data: FAMI_NUM_PERS_GRUP_FAM and MOD_COMUNICA_ESCRITA_PUNT  
t = -0.089879, df = 194, p-value = 0.9285  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1464730 0.1338209  
sample estimates:  
cor  
-0.006452816
```

12. Cruzando Número de dormitorios vs Comunicación escrita

```
> with(Dataset, cor.test(ESTU_DORMITORIOS, MOD_COMUNICA_ESCRITA_PUNT, alternative="two.sided",  
+ method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: ESTU_DORMITORIOS and MOD_COMUNICA_ESCRITA_PUNT  
t = 1.835, df = 194, p-value = 0.06803  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.009710779 0.265904488  
sample estimates:  
cor  
0.1306197
```

13. Cruzando Edad vs Inglés

```
> with(Dataset, cor.test(EDAD, MOD_INGLES_PUNT, alternative="two.sided", method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: EDAD and MOD_INGLES_PUNT  
t = -0.031262, df = 195, p-value = 0.9751  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1419900 0.1376001  
sample estimates:  
cor  
-0.002238696
```

14. Cruzando Número de personas de las que usted se encuentra a cargo vs Inglés

```
> with(Dataset, cor.test(FAMI_NUM_PERS_CARGO, MOD_INGLES_PUNT, alternative="two.sided", method="pearson"))
```

```
Pearson's product-moment correlation  
  
data: FAMI_NUM_PERS_CARGO and MOD_INGLES_PUNT  
t = 0.22679, df = 196, p-value = 0.8208  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.1235234 0.1552878  
sample estimates:  
cor  
0.01619702
```

15. Cruzando Número de personas que conforman su hogar vs Inglés

```
> with(Dataset, cor.test(FAMI_NUM_PERS_GRP_FAM, MOD_INGLES_PUNT, alternative="two.sided", method="pearson"))
```

```
Pearson's product-moment correlation

data: FAMI_NUM_PERS_GRP_FAM and MOD_INGLES_PUNT
t = -3.0966, df = 196, p-value = 0.002245
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.34501649 -0.07889954
sample estimates:
      cor
-0.215965
```

16. Cruzando Número de dormitorios vs Inglés

```
> with(Dataset, cor.test(ESTU_DORMITORIOS, MOD_INGLES_PUNT, alternative="two.sided",
+ method="pearson"))
```

```
Pearson's product-moment correlation

data: ESTU_DORMITORIOS and MOD_INGLES_PUNT
t = -0.85309, df = 196, p-value = 0.3947
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.19857953  0.07929174
sample estimates:
      cor
-0.06082223
```

Las siguientes tablas de contingencia muestran asociación entre algunas variables con los desempeños en inglés y comunicación escrita respectivamente.

1. Cruzando Sexo vs Desempeño en Inglés

```
Frequency table:
```

```
          MOD_INGLES_DESEM
ESTU_GENERO A- A1 A2 B+ B1
      F   8 29 24  3 24
      M  20 28 33  5 23
```

```
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect
```

```
Pearson's Chi-squared test
```

```
data: .Table
X-squared = 4.9201, df = 4, p-value = 0.2956
```

2. Cruzando Estado civil vs Desempeño en Inglés

```
Frequency table:
              MOD_INGLES_DESEM
Estado_civil_cod A- A1 A2 B+ B1
Casado           1  3  1  0  0
Separado         0  1  2  0  0
Soltero          27 52 54  7 45
Unión libre      0  1  0  1  2
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 13.41, df = 12, p-value = 0.34
```

3. Cruzando Hogar actual vs Desempeño en Inglés

```
Frequency table:
              MOD_INGLES_DESEM
Hogar_actual A- A1 A2 B+ B1
Permanente  18 39 36  3 26
temporal    10 18 21  5 22
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 4.3512, df = 4, p-value = 0.3606
```

4. Cruzando Cabeza de familia vs Desempeño en Inglés

```
Frequency table:
              MOD_INGLES_DESEM
Cabeza_familia_cod A- A1 A2 B+ B1
No                 27 51 54  8 48
Si                  1  6  3  0  0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 6.6759, df = 4, p-value = 0.154
```

5. Cruzando Estrato vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
Estrato_cod A- A1 A2 B+ B1
Estrato 1   1  5  5  0  3
Estrato 2  18 29 27  4 20
Estrato 3   6 22 21  4 23
Estrato 4   2  1  4  0  1
Estrato 5   1  0  0  0  0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 16.477, df = 16, p-value = 0.4202
```

6. Cruzando Valor anual de la matrícula vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
Valor_anual_matricula_cod A- A1 A2 B+ B1
Entre 1 millón y 3 millones 25 46 50  5 41
Entre 3 millones y 5 millones  1  1  2  2  1
Entre 500 mil y menos de de 1 millón 0  1  0  0  0
Menos de 500 mil           1  5  4  0  4
No pagó matrícula         1  3  1  1  2
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 18.124, df = 16, p-value = 0.3167
```

7. Cruzando Formas de pago matrícula vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
Forma_de_pagó_matricula_cod A- A1 A2 B+ B1
No cancelo matricula      3  4  5  0  2
Pagó por beca             20 26 26  5 21
Pagó por credito          1  5  4  0  1
Pagó por padres           1  4  1  1  2
Pagó propio               2 10  9  2 14
Varias formas de pagó     1  8 12  0  8
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 22.001, df = 20, p-value = 0.3404
```

8. Cruzando Ocupación del padre vs Desempeño en Inglés

```

Frequency table:
                                MOD_INGLES_DESEM
Ocupación_padre_cod            A-  A1  A2  B+  B1
Empleado como director o gerente      0  0  2  1  3
Empleado de nivel auxiliar o administrativo  2  0  0  0  2
empleado de nivel directivo             1  3  5  1  1
empleado de nivel técnico o profesional  3  7  6  1  4
Empleado obrero u operario             3 12  6  1  4
Empresario                             0  1  2  0  1
Hogar                                   0  2  3  0  1
Otra actividad u ocupación             3  6  8  1  3
Pensionado                             8  6  4  3  7
Pequeño empresario                     4  7  7  0  3
Profesional independiente              0  3  2  0  6
Trabajador por cuenta propia           4 10 12  0 13
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 48.603, df = 44, p-value = 0.2929

```

9. Cruzando Ocupación de la madre vs Desempeño en Inglés

```

Frequency table:
                                MOD_INGLES_DESEM
Ocupación_madre_cod            A-  A1  A2  B+  B1
Empleado como director o gerente      1  0  1  0  3
Empleado de nivel auxiliar o administrativo  4  4  2  0  2
empleado de nivel directivo            3  4  2  1  3
empleado de nivel técnico o profesional  2  5  3  1  5
Empleado obrero u operario            1  2  3  2  3
Hogar                                  5 17 19  1 17
Otra actividad u ocupación            3  5  6  1  1
Pensionado                            2  4  0  1  7
Pequeño empresario                    1  5  2  0  6
Profesional independiente              1  2  3  0  1
Trabajador por cuenta propia           5  9 16  1  0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 51.125, df = 40, p-value = 0.1117

```

10. Cruzando educación padre vs Desempeño en Inglés

```

Frequency table:
                                MOD_INGLES_DESEM
Educación_padre_cod            A-  A1  A2  B+  B1
Educación profesional completa         4  6 10  2  8
Educación profesional incompleta       0  0  5  0  2
Educación técnica o tecnologica completa  4  6  8  1  7
Educación técnica o tecnologica incompleta  2  3  2  0  1
Ninguno                                3  0  0  0  3
Potsgrado                              2  5  6  1  4
Primaria completa                      4  7  6  0  3
Primaria incompleta                    1 13  5  1  6
Secundaria completa                    5  6  6  2  9
Secundaria incompleta                  3 11  9  1  5
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

```

Pearson's Chi-squared test

```

data: .Table
X-squared = 36.342, df = 36, p-value = 0.4527

```

11. Cruzando educación madre vs Desempeño en Inglés

Frequency table:

```
MOD_INGLES_DESEM
Educación_madre_cod A- A1 A2 B+ B1
Educación profesional completa      5  7 11  2 11
Educación profesional incompleta    0  1  2  1  1
Educación técnica o tecnologica completa  1  6  5  0  5
Educación técnica o tecnologica incompleta 0  1  3  1  1
Ninguno                             1  0  0  0  0
No sabe                             0  1  0  0  0
Potsgrado                           3  4  4  1  8
Primaria completa                   4  3  4  0  1
Primaria incompleta                 1 12  6  0  5
Secundaria completa                 9 14 10  2 12
Secundaria incompleta               4  8 12  1  4
```

```
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect
```

Pearson's Chi-squared test

```
data: .Table
X-squared = 40.856, df = 40, p-value = 0.4327
```

12. Cruzando El hogar cuenta con celular vs Desempeño en Inglés

Frequency table:

```
MOD_INGLES_DESEM
El_hogar_cuenta_con_celular_cod A- A1 A2 B+ B1
No 7 12 12 3 8
Si 21 45 45 5 40
```

```
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect
```

Pearson's Chi-squared test

```
data: .Table
X-squared = 2.1054, df = 4, p-value = 0.7164
```

13. Cruzando El hogar cuenta con internet vs Desempeño en Inglés

Frequency table:

```
MOD_INGLES_DESEM
El_hogar_cuenta_con_internet_cod A- A1 A2 B+ B1
No 15 30 22 3 17
Si 13 27 35 5 31
```

```
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect
```

Pearson's Chi-squared test

```
data: .Table
X-squared = 5.0136, df = 4, p-value = 0.2859
```

14. Cruzando El hogar cuenta con televisión vs Desempeño en Inglés

```
Frequency table:
                                MOD_INGLES_DESEM
El_hogar_cuenta_con_televisión_cod A- A1 A2 B+ B1
                                No 10 20 16  1 15
                                Si 18 37 41  7 33
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 2.225, df = 4, p-value = 0.6944
```

15. Cruzando El hogar cuenta con lavadora vs Desempeño en Inglés

```
Frequency table:
                                MOD_INGLES_DESEM
El_hogar_cuenta_con_lavadora_cod A- A1 A2 B+ B1
                                No  9 19 17  0 11
                                Si 19 38 40  8 37
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 4.8256, df = 4, p-value = 0.3057
```

16. Cruzando El hogar cuenta con teléfono fijo vs Desempeño en Inglés

```
Frequency table:
                                MOD_INGLES_DESEM
El_hogar_cuenta_con_teléfono_cod A- A1 A2 B+ B1
                                No 15 31 32  3 21
                                Si 13 26 25  5 27
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 2.5116, df = 4, p-value = 0.6426
```

17. Cruzando Tomó algún curso de preparación al examen vs Desempeño en Inglés

```
Frequency table:
                                MOD_INGLES_DESEM
Tomó_curso A- A1 A2 B+ B1
            No 11 10 12  1 15
            Si 13 37 35  6 22
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 7.7547, df = 4, p-value = 0.101
```

18. Cruzando Bachillerato del que se graduó vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
ESTU_TITULO_BTO A- A1 A2 B+ B1
      A 15 27 37  4 31
      N  0  3  1  0  1
      T 13 27 19  4 16
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 7.3391, df = 8, p-value = 0.5005
```

19. Cruzando Trabaja actualmente vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
Estudiante_trabaja_cod A- A1 A2 B+ B1
      No                15 28 21  6 30
      Si, como ayudante sin remuneración  9 20 22  0 11
      Si, por experiencia o recursos en gastos generales  3  9 14  2  6
      Si, por ser práctica obligatoria del plan de estudios  1  0  0  0  1
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 16.365, df = 12, p-value = 0.1751
```

20. Cruzando Horas que trabaja vs Desempeño en Inglés

```
Frequency table:
      MOD_INGLES_DESEM
Horas_trabajo_cod A- A1 A2 B+ B1
      De 11 a 15 horas  3  7  9  0  4
      De 16 a 20 horas  8 10 16  0  6
      De 6 a 10 horas   2 12 11  2  8
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 8.0873, df = 8, p-value = 0.425
```

21. Cruzando Ingresos mensuales en salarios mínimos mensuales vs Desempeño en Inglés

```
Frequency table:
                MOD_INGLES_DESEM
Ingresos_cod   A-  A1  A2  B+  B1
  Entre 1 y menos de 2 SM  7 21 22  1 19
  Entre 2 y menos de 3 SM  8 20 17  3 11
  Entre 3 y menos de 5 SM  4  7  9  3 10
  Entre 5 y menos de 7 SM  3  0  2  1  4
  Entre 7 y menos de 10 SM 0  0  1  0  0
  Menos de 1 SM           6  9  6  0  4
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 20.695, df = 20, p-value = 0.4153
```

1. Cruzando Sexo vs Desempeño en Comunicación escrita

```
Frequency table:
      Desem_Comunica_Escrita
ESTU_GENERO Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      F      4      10      28      26      15      3
      M      4      17      36      39      10      3
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 3.7542, df = 5, p-value = 0.5853
```

2. Cruzando Estado civil vs Desempeño en Comunicación escrita

```
Frequency table:
      Desem_Comunica_Escrita
Estado_civil_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
  Casado           1      3      1      0      0      0
  Separado          0      1      2      0      0      0
  Soltero           6     23     60     63     25     6
  Unión libre       1      0      1      2      0      0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 23.519, df = 15, p-value = 0.07372
```

3. Cruzando Hogar actual vs Desempeño en Comunicación escrita

```
Frequency table:
      Desem_Comunica_Escrita
Hogar_actual Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
  Permanente      5      16      46      37      15      2
  temporal        3      11      18      28      10      5
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 6.7856, df = 5, p-value = 0.2371
```

4. Cruzando Cabeza de familia vs Desempeño en Comunicación escrita

```
Frequency table:
      Desem_Comunica_Escrita
Cabeza_familia_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      No          7      23      62      62      25      7
      Si           1       4       2       3       0      0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 8.434, df = 5, p-value = 0.1339
```

5. Cruzando Estrato vs Desempeño en Comunicación escrita

```
Frequency table:
      Desem_Comunica_Escrita
Estrato_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
  Estrato 1      0      4      1      5      2      2
  Estrato 2      4     18     28     36     11      0
  Estrato 3      4      5     31     20     11      4
  Estrato 4      0      0      4      3      1      0
  Estrato 5      0      0      0      1      0      0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 28.535, df = 20, p-value = 0.09732
```

6. Cruzando Valor anual de la matrícula vs Desempeño en Comunicación escrita

Frequency table:

Valor_anual_matricula_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Entre 1 millón y 3 millones	7	20	55	56	20	7
Entre 3 millones y 5 millones	0	2	1	3	1	0
Entre 500 mil y menos de de 1 millón	0	0	1	0	0	0
Menos de 500 mil	1	4	5	2	2	0
No pagó matrícula	0	0	2	4	2	0

Warning in chisq.test(.Table, correct = FALSE) :
Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table
X-squared = 13.482, df = 20, p-value = 0.8558

7. Cruzando Formas de pago matrícula vs Desempeño en Comunicación escrita

Frequency table:

Forma_de_pagó_matricula_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
No cancelo matrícula	0	1	5	5	3	0
Pagó por beca	3	12	35	34	10	3
Pagó por credito	1	2	4	1	1	2
Pagó por padres	0	1	2	4	2	0
Pagó propio	3	5	10	11	5	2
Varias formas de pagó	1	6	8	10	4	0

Warning in chisq.test(.Table, correct = FALSE) :
Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table
X-squared = 19.959, df = 25, p-value = 0.749

8. Cruzando Ocupación del padre vs Desempeño en Comunicación escrita

Frequency table:

Ocupación_padre_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Empleado como director o gerente	0	0	2	1	2	1
Empleado de nivel auxiliar o administrativo	0	0	1	2	1	0
empleado de nivel directivo	0	1	2	5	1	2
empleado de nivel técnico o profesional	1	2	9	6	1	1
Empleado obrero u operario	1	3	10	9	3	0
Empresario	0	0	1	1	2	0
Hogar	0	1	4	1	0	0
Otra actividad u ocupación	0	6	3	8	2	2
Pensionado	2	4	11	8	2	1
Pequeño empresario	0	4	3	10	4	0
Profesional independiente	0	1	5	1	3	0
Trabajador por cuenta propia	4	5	13	13	4	0

Warning in chisq.test(.Table, correct = FALSE) :
Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table
X-squared = 57.422, df = 55, p-value = 0.3855

9. Cruzando Ocupación de la madre vs Desempeño en Comunicación escrita

Frequency table:

Ocupación_madre_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Empleado como director o gerente	0	0	3	1	0	1
Empleado de nivel auxiliar o administrativo	1	1	4	5	1	0
empleado de nivel directivo	0	1	4	4	3	1
empleado de nivel técnico o profesional	0	2	7	5	1	1
Empleado obrero u operario	0	2	5	1	2	1
Hogar	3	6	18	21	9	1
Otra actividad u ocupación	1	3	5	5	1	0
Pensionado	1	1	3	6	2	1
Pequeño empresario	0	4	6	3	1	0
Profesional independiente	0	0	2	4	1	0
Trabajador por cuenta propia	2	7	7	10	4	1

Warning in chisq.test(.Table, correct = FALSE) :

Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table

X-squared = 33.649, df = 50, p-value = 0.9632

10. Cruzando educación padre vs Desempeño en Comunicación escrita

Frequency table:

Educación_padre_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Educación profesional completa	1	3	13	8	3	2
Educación profesional incompleta	0	0	1	2	4	0
Educación técnica o tecnologica completa	1	1	9	8	5	2
Educación técnica o tecnologica incompleta	0	1	3	1	2	0
Ninguno	0	3	1	1	0	1
Potsgrado	0	1	6	7	3	1
Primaria completa	1	5	3	8	3	0
Primaria incompleta	1	8	10	6	0	1
Secundaria completa	0	1	10	13	3	0
Secundaria incompleta	4	4	8	11	2	0

Warning in chisq.test(.Table, correct = FALSE) :

Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table

X-squared = 65.282, df = 45, p-value = 0.02562

11. Cruzando educación madre vs Desempeño en Comunicación escrita

Frequency table:

Educación_madre_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Educación profesional completa	0	2	11	12	8	3
Educación profesional incompleta	0	0	1	0	4	0
Educación técnica o tecnologica completa	0	3	8	4	2	0
Educación técnica o tecnologica incompleta	0	0	2	3	1	0
Ninguno	0	1	0	0	0	0
No sabe	0	0	0	0	0	0
Potsgrado	1	3	5	7	2	2
Primaria completa	1	3	3	4	1	0
Primaria incompleta	1	8	8	6	1	0
Secundaria completa	4	5	18	16	2	2
Secundaria incompleta	1	2	8	13	4	0

Warning in chisq.test(.Table, correct = FALSE) :

Chi-squared approximation may be incorrect

Pearson's Chi-squared test

data: .Table

X-squared = NaN, df = 50, p-value = NA

12. Cruzando El hogar cuenta con celular vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
El_hogar_cuenta_con_celular_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      1      9      12      9      10      1
                Si      7     18     52     56     15      6
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 10.477, df = 5, p-value = 0.06279
.
```

13. Cruzando El hogar cuenta con internet vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
El_hogar_cuenta_con_internet_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      3     15     28     32      8      1
                Si      5     12     36     33     17      6
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 6.2698, df = 5, p-value = 0.2808
```

14. Cruzando El hogar cuenta con televisión vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
El_hogar_cuenta_con_televisión_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      4      8     18     21      8      3
                Si      4     19     46     44     17      4
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 2.0852, df = 5, p-value = 0.8372
```

15. Cruzando El hogar cuenta con lavadora vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
El_hogar_cuenta_con_lavadora_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      2      11      19      19      5      0
                Si      6      16      45      46      20      7
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 5.7622, df = 5, p-value = 0.3301
```

16. Cruzando El hogar cuenta con teléfono fijo vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
El_hogar_cuenta_con_teléfono_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      2      15      34      38      10      3
                Si      6      12      30      27      15      4
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 5.27, df = 5, p-value = 0.3838
```

17. Cruzando Tomó algún curso de preparación al examen vs Desempeño en Comunicación escrita

```
Frequency table:
                Desem_Comunica_Escrita
Tomó_curso Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
                No      3      8      19      14      4      1
                Si      5      14      30      44      13      5
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data: .Table
X-squared = 4.153, df = 5, p-value = 0.5276
```

18. Cruzando Bachillerato del que se graduó vs Desempeño en Comunicación escrita

```

Frequency table:
      Desem_Comunica_Escrita
ESTU_TITULO_BTO Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      A          6      15      39      35      14      5
      N          0       1       2       1       1       0
      T          2      11      23      29      10      2
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data:  .Table
X-squared = 3.3098, df = 10, p-value = 0.9732

```

19. Cruzando Trabaja actualmente vs Desempeño en Comunicación escrita

```

Frequency table:
      Desem_Comunica_Escrita
Estudiante_trabaja_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      No                4      10      30      33      16      5
      Si, como ayudante sin remuneración
      Si, por experiencia o recursos en gastos generales
      Si, por ser práctica obligatoria del plan de estudios
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data:  .Table
X-squared = 12.273, df = 15, p-value = 0.6583

```

20. Cruzando Horas que trabaja vs Desempeño en Comunicación escrita

```

Frequency table:
      Desem_Comunica_Escrita
Horas_trabajo_cod Nivel 2 Nivel 3 Nivel 4 Nivel 5 Nivel 6 Nivel 7
      De 11 a 15 horas  0       4       7       9       3       0
      De 16 a 20 horas  2       8      14      11       3       2
      De 6 a 10 horas   2       5      13      12       3       0
Warning in chisq.test(.Table, correct = FALSE) :
  Chi-squared approximation may be incorrect

      Pearson's Chi-squared test

data:  .Table
X-squared = 5.8435, df = 10, p-value = 0.8282

```

21. Cruzando Ingresos mensuales en salarios mínimos vs Desempeño en Comunicación escrita

Frequency table:

Ingresos_cod	Desem_Comunica_Escrita					
	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
Entre 1 y menos de 2 SM	4	11	18	25	9	3
Entre 2 y menos de 3 SM	2	8	27	11	7	3
Entre 3 y menos de 5 SM	2	2	11	12	4	1
Entre 5 y menos de 7 SM	0	0	4	4	2	0
Entre 7 y menos de 10 SM	0	0	0	1	0	0
Menos de 1 SM	0	6	4	12	3	0

```
Warning in chisq.test(.Table, correct = FALSE) :  
Chi-squared approximation may be incorrect
```

Pearson's Chi-squared test

```
data: .Table  
X-squared = 23.764, df = 25, p-value = 0.533
```

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 en razonamiento cuantitativo, lectura crítica y comunicación escrita tienden a distribuirse normalmente. En cuanto al puntaje en inglés, este no asume una distribución normal.

PRUEBA DE NORMALIDAD PUNTAJE RAZONAMIENTO CUANTITATIVO

```
> with(Dataset, shapiro.test(MOD_RAZONA_CUANTITATIVO_PUNT))
```

Shapiro-Wilk normality test

```
data: MOD_RAZONA_CUANTITATIVO_PUNT  
W = 0.99092, p-value = 0.2493
```

PRUEBA DE NORMALIDAD PUNTAJE LECTURA CRÍTICA

```
> with(Dataset, shapiro.test(MOD_LECTURA_CRITICA))
```

Shapiro-Wilk normality test

```
data: MOD_LECTURA_CRITICA  
W = 0.99416, p-value = 0.6324
```

PRUEBA DE NORMALIDAD PUNTAJE COMUNICACIÓN ESCRITA

```
> with(Dataset, shapiro.test(MOD_COMUNICA_ESCRITA_PUNT))

Shapiro-Wilk normality test

data:  MOD_COMUNICA_ESCRITA_PUNT
W = 0.9928, p-value = 0.4501
```

PRUEBA DE NORMALIDAD PUNTAJE INGLÉS

```
> with(Dataset, shapiro.test(MOD_INGLES_PUNT))

Shapiro-Wilk normality test

data:  MOD_INGLES_PUNT
W = 0.67957, p-value < 2.2e-16
```

Luego, para aquellas variables normales se realizó el test de varianzas de Leven's, con el fin de determinar varianzas iguales, determinando diferencia de medias entre algunas variables con los puntajes de razonamiento cuantitativo, lectura crítica y comunicación escrita respectivamente.

1. Test de varianzas sexo vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_GENERO, var, na.rm=TRUE))
      F      M
0.6287722 0.7046109
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_GENERO, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.004 0.9497
      195
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~ESTU_GENERO, alternative='two.sided', conf.level=.95,
+        var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by ESTU_GENERO
t = -3.0305, df = 195, p-value = 0.002773
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5871751 -0.1242094
sample estimates:
mean in group F mean in group M
 10.02045      10.37615
```

2. Test de varianzas Estado civil vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Estado_civil_cod, var, na.rm=TRUE))
      Casado      Separado      Soltero Unión libre
0.0670000  2.0133333  0.7006187  0.3800000
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estado_civil_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  1.8629 0.1372
      193
>
> AnovaModel.2 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estado_civil_cod, data=Dataset)
> summary(AnovaModel.2)
              Df Sum Sq Mean Sq F value Pr(>F)
Estado_civil_cod  3   2.61  0.8709   1.251 0.292
Residuals        193 134.35  0.6961
3 observations deleted due to missingness
> |
```

3. Test de varianzas situación de su hogar actual vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Hogar_actual, var, na.rm=TRUE))
Permanente temporal
0.7159599  0.6467982
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Hogar_actual, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.3473 0.5563
      196
>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~Hogar_actual, alternative='two.sided',
+        conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_RAZONA_CUANTITATIVO_PUNT by Hogar_actual
t = -1.7252, df = 196, p-value = 0.08607
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.44864367  0.02996377
sample estimates:
mean in group Permanente mean in group temporal
      10.13934             10.34868
```

4. Test de varianzas Cabeza de familia vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Cabeza_familia_cod, var,
+ na.rm=TRUE))
      No      Si
0.72508676 0.07288889
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Cabeza_familia_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  6.659 0.0106 *
      196
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~Cabeza_familia_cod, alternative='two.sided',
+        conf.level=.95, var.equal=FALSE, data=Dataset)

Welch Two Sample t-test

data: MOD_RAZONA_CUANTITATIVO_PUNT by Cabeza_familia_cod
t = 2.9898, df = 20.765, p-value = 0.007041
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.09593485 0.53534175
sample estimates:
mean in group No mean in group Si
      10.23564             9.92000
```

5. Test de varianzas Estrato vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Estrato_cod, var, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5
1.0400000 0.6246003 0.7570579 0.7942857      NA
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estrato_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  0.7177 0.5808
      192
>
> AnovaModel.6 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estrato_cod, data=Dataset)
> summary(AnovaModel.6)
      Df Sum Sq Mean Sq F value Pr(>F)
Estrato_cod  4  0.52  0.1289   0.181  0.948
Residuals  192 136.45  0.7107
3 observations deleted due to missingness
> |
```

6. Test de varianzas Valor anual de la matrícula vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Valor_anual_matricula_cod, var,
+ na.rm=TRUE))
      Entre 1 millón y 3 millones      Entre 3 millones y 5 millones Entre 500 mil y menos de de 1 millón      Menos de 500 mil
      0.6260126                    2.9295238                    NA                    0.7380220
      No pagó matrícula
      0.3455357
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Valor_anual_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  4.0946 0.003302 **
      192
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> AnovaModel.7 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Valor_anual_matricula_cod, data=Dataset)
> summary(AnovaModel.7)
      Df Sum Sq Mean Sq F value Pr(>F)
Valor_anual_matricula_cod  4  3.41  0.8534  1.227  0.301
Residuals                192 133.51  0.6954
3 observations deleted due to missingness
> |
```

7. Test de varianzas Forma de pagó matrícula vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Forma_de_pagó_matricula_cod, var,
+ na.rm=TRUE))
      No cancelo matrícula      Pagó por beca      Pagó por credito      Pagó por padres      Pagó propio Varias formas de pagó
      0.8210989                    0.7274774                    0.7096364                    0.3150000                    0.7892342                    0.4614778
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Forma_de_pagó_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  0.7069 0.6189
      192
>
> AnovaModel.8 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Forma_de_pagó_matricula_cod,
+ data=Dataset)
> summary(AnovaModel.8)
      Df Sum Sq Mean Sq F value Pr(>F)
Forma_de_pagó_matricula_cod  5  5.0  1.0007  1.453  0.207
Residuals                192 132.2  0.6885
2 observations deleted due to missingness
> |
```

8. Test de varianzas Ocupación del padre vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNTI, Ocupación_padre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      2.4306667 1.7891667 0.3949091
      empleado de nivel técnico o profesional Empleado obrero u operario Empresario
      0.9565714 0.7314615 0.6225000
      Hogar Otra actividad u ocupación Pensionado
      1.6546667 0.8034762 0.5397222
      Pequeño empresario Profesional independiente Trabajador por cuenta propia
      0.4714762 0.2960000 0.4987989

> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNTI ~ Ocupación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group 11 1.1918 0.2952
      186

>
> AnovaModel.9 <- aov(MOD_RAZONA_CUANTITATIVO_PUNTI ~ Ocupación_padre_cod, data=Dataset)
> summary(AnovaModel.9)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_padre_cod 11 6.18 0.5617 0.797 0.643
Residuals          186 131.01 0.7044
2 observations deleted due to missingness
> |
```

9. Test de varianzas Ocupación de la madre vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNTI, Ocupación_madre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      0.3030000 0.7069697 0.8443590
      empleado de nivel técnico o profesional Empleado obrero u operario Hogar
      0.7336250 0.4601818 0.8465400
      Otra actividad u ocupación Pensionado Pequeño empresario
      0.5345000 0.3934615 0.3191758
      Profesional independiente Trabajador por cuenta propia
      0.4900000 0.8171183

> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNTI ~ Ocupación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group 10 0.6911 0.732
      187

>
> AnovaModel.10 <- aov(MOD_RAZONA_CUANTITATIVO_PUNTI ~ Ocupación_madre_cod, data=Dataset)
> summary(AnovaModel.10)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_madre_cod 10 8.63 0.8631 1.255 0.259
Residuals          187 128.56 0.6875
2 observations deleted due to missingness
> |
```

10. Test de varianzas El hogar cuenta con internet vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNTI, El_hogar_cuenta_con_internet_cod,
+ var, na.rm=TRUE))
      No Si
0.5772654 0.7492826

> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNTI ~ El_hogar_cuenta_con_internet_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group 1 0.3366 0.5625
      196

>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNTI~El_hogar_cuenta_con_internet_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_RAZONA_CUANTITATIVO_PUNTI by El_hogar_cuenta_con_internet_cod
t = -2.7585, df = 196, p-value = 0.006356
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.55603293 -0.09242932
sample estimates:
mean in group No mean in group Si
10.03793 10.36216
```

11. Test de varianzas El hogar cuenta con celular vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, El_hogar_cuenta_con_celular_cod, var,
+   na.rm=TRUE))
      No      Si
0.8131998 0.6699983
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ El_hogar_cuenta_con_celular_cod, data=Dataset,
+   center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.3187  0.573
      196
>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~El_hogar_cuenta_con_celular_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by El_hogar_cuenta_con_celular_cod
t = 0.056668, df = 196, p-value = 0.9549
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2785841  0.2950676
sample estimates:
mean in group No mean in group Si
      10.22619      10.21795
```

12. Test de varianzas El hogar cuenta con televisión vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, El_hogar_cuenta_con_televisión_cod,
+   var, na.rm=TRUE))
      No      Si
0.5022528 0.7888448
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ El_hogar_cuenta_con_televisión_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.6611  0.199
      196
>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~El_hogar_cuenta_con_televisión_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by El_hogar_cuenta_con_televisión_cod
t = -0.29701, df = 196, p-value = 0.7668
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2908511  0.2147126
sample estimates:
mean in group No mean in group Si
      10.19355      10.23162
```

13. Test de varianzas El hogar cuenta con lavadora vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, El_hogar_cuenta_con_lavadora_cod,
+   var, na.rm=TRUE))
      No      Si
0.6019091 0.7371077
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ El_hogar_cuenta_con_lavadora_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.6948  0.4056
      196
>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~El_hogar_cuenta_con_lavadora_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_RAZONA_CUANTITATIVO_PUNT by El_hogar_cuenta_con_lavadora_cod
t = -0.47236, df = 196, p-value = 0.6372
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3225337  0.1978858
sample estimates:
mean in group No mean in group Si
      10.17500      10.23732
```

14. Test de varianzas El hogar cuenta con teléfono vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, El_hogar_cuenta_con_teléfono_cod,
+ var, na.rm=TRUE))
      No      Si
0.6888692 0.7048246
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ El_hogar_cuenta_con_teléfono_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0552 0.8144
      196

>
> t.test(MOD_RAZONA_CUANTITATIVO_PUNT~El_hogar_cuenta_con_teléfono_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_RAZONA_CUANTITATIVO_PUNT by El_hogar_cuenta_con_teléfono_cod
t = -0.97268, df = 196, p-value = 0.3319
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3495013  0.1186190
sample estimates:
mean in group No mean in group Si
      10.16373      10.27917
```

15. Test de varianzas Tomó algún curso de preparación al examen vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZO_CUANTITATIVO_PUNT, Tomó_curso, var, na.rm=TRUE))
      No      Si
0.8506803 0.6216546
> leveneTest(MOD_RAZO_CUANTITATIVO_PUNT ~ Tomó_curso, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.1466 0.2859
      160

> t.test(MOD_RAZO_CUANTITATIVO_PUNT~Tomó_curso, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_RAZO_CUANTITATIVO_PUNT by Tomó_curso
t = 1.8006, df = 160, p-value = 0.07366
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02477862  0.53657201
sample estimates:
mean in group No mean in group Si
      10.38776      10.13186
```

16. Test de varianzas Tipo de bachillerato del que se graduó vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, ESTU_TITULO_BTO, var, na.rm=TRUE))
      A      N      T
0.8413825 0.3730000 0.5040409
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_TITULO_BTO, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  2.729 0.06778 .
      195

---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> AnovaModel.12 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ ESTU_TITULO_BTO, data=Dataset)
> summary(AnovaModel.12)

      Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TITULO_BTO  2  1.31  0.6549  0.94  0.392
Residuals      195 135.88  0.6968
2 observations deleted due to missingness
> |
```

17. Test de varianzas Educación del padre vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Educación_padre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      0.8133333      Educación profesional incompleta      0.5528571      Educación técnica o tecnologica completa      0.6985385
Educación técnica o tecnologica incompleta      0.2898214      Ninguno      0.9026667      Potgrado      0.7269281
      Primaria completa      0.8901053      Primaria incompleta      0.4663538      Secundaria completa      0.8571958
      Secundaria incompleta      0.5228079
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Educación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group   9  0.3295 0.9644
      188
>
> AnovaModel.13 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Educación_padre_cod, data=Dataset)
> summary(AnovaModel.13)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_padre_cod  9  7.57  0.8414  1.22 0.285
Residuals          188 129.62  0.6895
2 observations deleted due to missingness
.
```

18. Test de varianzas Educación de la madre vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Educación_madre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      0.7859286      Educación profesional incompleta      1.7670000      Educación técnica o tecnologica completa      0.4486029
Educación técnica o tecnologica incompleta      0.4960000      Ninguno      NA      No sabe      NA
      Potsgrado      0.4120000      Primaria completa      0.9353788      Primaria incompleta      0.4921558
      Secundaria completa      0.8601018      Secundaria incompleta      0.4660099
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Educación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  10  0.9669 0.4738
      187
>
> AnovaModel.14 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Educación_madre_cod, data=Dataset)
> summary(AnovaModel.14)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_madre_cod  10 10.91  1.0910  1.616 0.105
Residuals          187 126.28  0.6753
2 observations deleted due to missingness
.
```

19. Test de varianzas Trabaja actualmente vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Estudiante_trabaja_cod, var,
+ na.rm=TRUE))
      No      Si, como ayudante sin remuneración      Si, por experiencia o recursos en gastos generales
Si, por ser práctica obligatoria del plan de estudios      0.7653929      0.5475198      0.7565152
      0.2450000
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estudiante_trabaja_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group   3  0.9878 0.3996
      194
>
> AnovaModel.15 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Estudiante_trabaja_cod,
+ data=Dataset)
> summary(AnovaModel.15)
      Df Sum Sq Mean Sq F value Pr(>F)
Estudiante_trabaja_cod  3  2.81  0.9369  1.352 0.259
Residuals          194 134.38  0.6927
2 observations deleted due to missingness
.>
```

20. Test de varianzas Horas de trabajo vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Horas_trabajo_cod, var, na.rm=TRUE))
De 11 a 15 horas De 16 a 20 horas De 6 a 10 horas
 0.5839130      0.6915321      0.5454958
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Horas_trabajo_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  0.2867 0.7514
      95
>
> AnovaModel.16 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Horas_trabajo_cod, data=Dataset)
> summary(AnovaModel.16)
              Df Sum Sq Mean Sq F value Pr(>F)
Horas_trabajo_cod  2   1.47  0.7336   1.194  0.308
Residuals          95  58.36  0.6143
102 observations deleted due to missingness
> |
```

21. Test de varianzas Ingresos mensuales en salarios mínimos vs Razonamiento cuantitativo

```
> with(Dataset, tapply(MOD_RAZONA_CUANTITATIVO_PUNT, Ingresos_cod, var, 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 Entre 7 y menos de 10 SM Menos de 1 SM
 0.7663023      0.5725365      0.6405871      1.3298889      NA      0.5645667
> leveneTest(MOD_RAZONA_CUANTITATIVO_PUNT ~ Ingresos_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  1.4893 0.195
      192
>
> AnovaModel.17 <- aov(MOD_RAZONA_CUANTITATIVO_PUNT ~ Ingresos_cod, data=Dataset)
> summary(AnovaModel.17)
              Df Sum Sq Mean Sq F value Pr(>F)
Ingresos_cod  5   5.09  1.019  1.481  0.198
Residuals    192 132.10  0.688
2 observations deleted due to missingness
> |
```

1. Test de varianzas sexo vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, ESTU_GENERO, var, na.rm=TRUE))
      F      M
0.5705277 0.4740741
> leveneTest(MOD_LECTURA_CRITICA ~ ESTU_GENERO, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.3275 0.5678
      195
>
> t.test(MOD_LECTURA_CRITICA~ESTU_GENERO, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_LECTURA_CRITICA by ESTU_GENERO
t = -0.41902, df = 195, p-value = 0.6757
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2464272  0.1600636
sample estimates:
mean in group F mean in group M
 10.35682      10.40000
```

2. Test de varianzas Estado civil vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Estado_civil_cod, var, na.rm=TRUE))
      Casado      Separado      Soltero Unión libre
0.4670000 0.5733333 0.5280306 0.1225000
> leveneTest(MOD_LECTURA_CRITICA ~ Estado_civil_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  0.8108 0.4893
      193
>
> AnovaModel.2 <- aov(MOD_LECTURA_CRITICA ~ Estado_civil_cod, data=Dataset)
> summary(AnovaModel.2)
              Df Sum Sq Mean Sq F value Pr(>F)
Estado_civil_cod  3  0.39  0.1290  0.248  0.863
Residuals       193 100.54  0.5209
3 observations deleted due to missingness
> |
```

3. Test de varianzas situación de su hogar actual vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Hogar_actual, var, na.rm=TRUE))
Permanente  temporal
0.4907113  0.5417035
> leveneTest(MOD_LECTURA_CRITICA ~ Hogar_actual, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.4006 0.5275
      196
> t.test(MOD_LECTURA_CRITICA~Hogar_actual, alternative='two.sided',
+        conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_LECTURA_CRITICA by Hogar_actual
t = -1.3496, df = 196, p-value = 0.1787
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.34673206  0.06498055
sample estimates:
mean in group Permanente    mean in group temporal
           10.32623                10.46711
```

4. Test de varianzas Cabeza de familia vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Cabeza_familia_cod, var,
+ na.rm=TRUE))
      No      Si
0.5200054 0.3426667
> leveneTest(MOD_LECTURA_CRITICA ~ Cabeza_familia_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.5819 0.4465
      196
> t.test(MOD_LECTURA_CRITICA~Cabeza_familia_cod, alternative='two.sided',
+        conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_LECTURA_CRITICA by Cabeza_familia_cod
t = 1.09, df = 196, p-value = 0.277
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2048124  0.7109826
sample estimates:
mean in group No mean in group Si
           10.39309                10.14000
```

5. Test de varianzas Estrato vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Estrato_cod, var, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5
0.2837912 0.4793814 0.6043140 0.3198214      NA
> leveneTest(MOD_LECTURA_CRITICA ~ Estrato_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  1.371 0.2455
      192
> AnovaModel.6 <- aov(MOD_LECTURA_CRITICA ~ Estrato_cod, data=Dataset)
> summary(AnovaModel.6)
      Df Sum Sq Mean Sq F value Pr(>F)
Estrato_cod  4   3.18  0.7938   1.559 0.187
Residuals  192  97.75  0.5091
3 observations deleted due to missingness

Attaching package: 'TH.data'

The following object is masked from 'package:MASS':

  geyser
```

6. Test de varianzas Valor anual de la matrícula vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Valor_anual_matricula_cod, var,
+ na.rm=TRUE))
      Entre 1 millón y 3 millones Entre 3 millones y 5 millones Entre 500 mil y menos de de 1 millón Menos de 500 mil
      0.5269411                0.8195238                NA                0.2440110
      No pagó matrícula
      0.5569643
> leveneTest(MOD_LECTURA_CRITICA ~ Valor_anual_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  1.3714 0.2454
      192
>
> AnovaModel.7 <- aov(MOD_LECTURA_CRITICA ~ Valor_anual_matricula_cod, data=Dataset)
> summary(AnovaModel.7)
      Df Sum Sq Mean Sq F value Pr(>F)
Valor_anual_matricula_cod  4  1.24  0.3103  0.599 0.664
Residuals                192  99.46  0.5180
3 observations deleted due to missingness
> |
```

7. Test de varianzas Forma de pagó matrícula vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Forma_de_pagó_matricula_cod, var,
+ na.rm=TRUE))
      No cancelo matricula      Pagó por beca      Pagó por credito      Pagó por padres      Pagó propio Varias formas de pagó
      0.6129670                0.5266621                0.5165455                0.5019444                0.6788589                0.2514286
> leveneTest(MOD_LECTURA_CRITICA ~ Forma_de_pagó_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  1.5313 0.1818
      192
>
> AnovaModel.8 <- aov(MOD_LECTURA_CRITICA ~ Forma_de_pagó_matricula_cod,
+ data=Dataset)
> summary(AnovaModel.8)
      Df Sum Sq Mean Sq F value Pr(>F)
Forma_de_pagó_matricula_cod  5  1.22  0.2437  0.469 0.799
Residuals                192  99.71  0.5193
2 observations deleted due to missingness
> |
```

8. Test de varianzas Ocupación del padre vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Ocupación_padre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      1.4120000                        1.4625000                        0.7356364
      empleado de nivel técnico o profesional Empleado obrero u operario      Empresario
      0.4291429                        0.4312154                        0.2300000
      Hogar      Otra actividad u ocupación      Pensionado
      0.3350000                        0.2985714                        0.5862963
      Pequeño empresario      Profesional independiente      Trabajador por cuenta propia
      0.4735714                        0.6467273                        0.4530904

> leveneTest(MOD_LECTURA_CRITICA ~ Ocupación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  11  0.6045 0.8236
      186

>
> AnovaModel.9 <- aov(MOD_LECTURA_CRITICA ~ Ocupación_padre_cod, data=Dataset)
> summary(AnovaModel.9)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_padre_cod  11  5.44  0.4949  0.964  0.481
Residuals           186 95.49  0.5134
2 observations deleted due to missingness
> |
```

9. Test de varianzas Ocupación de la madre vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Ocupación_madre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      0.2530000                        0.7935606                        0.3625641
      empleado de nivel técnico o profesional Empleado obrero u operario      Hogar
      0.5589583                        0.3307273                        0.4916657
      Otra actividad u ocupación      Pensionado      Pequeño empresario
      0.3251667                        0.5090110                        0.6474725
      Profesional independiente      Trabajador por cuenta propia
      0.4390476                        0.5749032

> leveneTest(MOD_LECTURA_CRITICA~ Ocupación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  10  0.5514 0.8516
      187

>
> AnovaModel.10 <- aov(MOD_LECTURA_CRITICA ~ Ocupación_madre_cod, data=Dataset)
> summary(AnovaModel.10)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_madre_cod  10  6.84  0.6840  1.359  0.202
Residuals           187 94.09  0.5032
2 observations deleted due to missingness
> |
```

10. Test de varianzas El hogar cuenta con internet vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, El_hogar_cuenta_con_internet_cod,
+ var, na.rm=TRUE))
      No      Si
0.4743892 0.5408714
> leveneTest(MOD_LECTURA_CRITICA ~ El_hogar_cuenta_con_internet_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0586 0.809
      196

> t.test(MOD_LECTURA_CRITICA~El_hogar_cuenta_con_internet_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by El_hogar_cuenta_con_internet_cod
t = -1.1182, df = 196, p-value = 0.2648
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.31654152  0.08746416
sample estimates:
mean in group No mean in group Si
      10.31609      10.43063
```

11. Test de varianzas El hogar cuenta con celular vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, El_hogar_cuenta_con_celular_cod, var,
+   na.rm=TRUE))
      No      Si
0.4987747 0.5184103
> leveneTest(MOD_LECTURA_CRITICA ~ El_hogar_cuenta_con_celular_cod, data=Dataset,
+   center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.223 0.6373
      196
> t.test(MOD_LECTURA_CRITICA~El_hogar_cuenta_con_celular_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by El_hogar_cuenta_con_celular_cod
t = -0.50243, df = 196, p-value = 0.6159
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3085003  0.1832256
sample estimates:
mean in group No mean in group Si
      10.33095      10.39359
```

12. Test de varianzas El hogar cuenta con televisión vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, El_hogar_cuenta_con_televisión_cod,
+   var, na.rm=TRUE))
      No      Si
0.5701454 0.4898976
> leveneTest(MOD_LECTURA_CRITICA ~ El_hogar_cuenta_con_televisión_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.6035 0.4382
      196
> t.test(MOD_LECTURA_CRITICA~El_hogar_cuenta_con_televisión_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by El_hogar_cuenta_con_televisión_cod
t = -0.18767, df = 196, p-value = 0.8513
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2374836  0.1962122
sample estimates:
mean in group No mean in group Si
      10.36613      10.38676
```

13. Test de varianzas El hogar cuenta con lavadora vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, El_hogar_cuenta_con_lavadora_cod,
+   var, na.rm=TRUE))
      No      Si
0.4454253 0.5407092
> leveneTest(MOD_LECTURA_CRITICA~ El_hogar_cuenta_con_lavadora_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.8573 0.1745
      196
> t.test(MOD_LECTURA_CRITICA~El_hogar_cuenta_con_lavadora_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by El_hogar_cuenta_con_lavadora_cod
t = -0.61562, df = 196, p-value = 0.5389
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2927441  0.1534583
sample estimates:
mean in group No mean in group Si
      10.33036      10.40000
```

14. Test de varianzas El hogar cuenta con teléfono vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, El_hogar_cuenta_con_teléfono_cod,
+ var, na.rm=TRUE))
      No      Si
0.5416502 0.4861831
> leveneTest(MOD_LECTURA_CRITICA ~ El_hogar_cuenta_con_teléfono_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1      0 0.9992
      196
> t.test(MOD_LECTURA_CRITICA~El_hogar_cuenta_con_teléfono_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by El_hogar_cuenta_con_teléfono_cod
t = -0.27567, df = 196, p-value = 0.7831
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2293304  0.1730804
sample estimates:
mean in group No mean in group Si
 10.36667          10.39479
```

15. Test de varianzas Tomó algún curso de preparación al examen vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Tomó_curso, var, na.rm=TRUE))
      No      Si
0.5636224 0.4929883
> leveneTest(MOD_LECTURA_CRITICA ~ Tomó_curso, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.5028 0.4793
      160
> t.test(MOD_LECTURA_CRITICA~Tomó_curso, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_LECTURA_CRITICA by Tomó_curso
t = -0.11912, df = 160, p-value = 0.9053
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2568379  0.2276163
sample estimates:
mean in group No mean in group Si
 10.36327          10.37788
```

16. Test de varianzas Tipo de bachillerato del que se graduó vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, ESTU_TITULO_BTO, var, na.rm=TRUE))
      A      N      T
0.5794636 0.3880000 0.4285751
> leveneTest(MOD_LECTURA_CRITICA ~ ESTU_TITULO_BTO, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  1.0259 0.3604
      195
> AnovaModel.12 <- aov(MOD_LECTURA_CRITICA ~ ESTU_TITULO_BTO, data=Dataset)
> summary(AnovaModel.12)
              Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TITULO_BTO  2  0.47  0.2365  0.459  0.633
Residuals      195 100.46  0.5152
2 observations deleted due to missingness
> |
```

17. Test de varianzas Educación del padre vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Educación_padre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      0.7593678      0.7214286      0.4175538
Educación técnica o tecnologica incompleta      0.2126786      Ninguno      Potgrado
      Primaria completa      Primaria incompleta      Secundaria completa
Educación profesional completa      0.2237895      0.5158154      0.7543386
      Secundaria incompleta
Educación profesional completa      0.2538177

> leveneTest(MOD_LECTURA_CRITICA ~ Educación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  9  2.2344 0.02158 *
      188

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

>
> AnovaModel.13 <- aov(MOD_LECTURA_CRITICA ~ Educación_padre_cod, data=Dataset)
> summary(AnovaModel.13)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_padre_cod  9   4.40   0.4894   0.953  0.48
Residuals           188  96.53   0.5135
2 observations deleted due to missingness
> |
```

18. Test de varianzas Educación de la madre vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Educación_madre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      0.5389444      1.2780000      0.2747059
Educación técnica o tecnologica incompleta      0.2336667      Ninguno      No sabe
      Potgrado      Primaria completa      Primaria incompleta
Educación profesional completa      0.5272368      0.4935606      0.2560688
      Secundaria completa      Secundaria incompleta
Educación profesional completa      0.7129880      0.3599261

> leveneTest(MOD_LECTURA_CRITICA ~ Educación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  10  1.4482 0.1622
      187

>
> AnovaModel.14 <- aov(MOD_LECTURA_CRITICA ~ Educación_madre_cod, data=Dataset)
> summary(AnovaModel.14)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_madre_cod  10   7.18   0.7183   1.433  0.169
Residuals           187  93.75   0.5013
2 observations deleted due to missingness
< |
```

19. Test de varianzas Trabaja actualmente vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Estudiante_trabaja_cod, var,
+ na.rm=TRUE))
      No      Si, como ayudante sin remuneración      Si, por experiencia o recursos en gastos generales
Educación profesional completa      0.5359303      0.3860418      0.6586542
Si, por ser práctica obligatoria del plan de estudios
Educación profesional completa      0.5000000

> leveneTest(MOD_LECTURA_CRITICA ~ Estudiante_trabaja_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  1.337 0.2636
      194

>
> AnovaModel.15 <- aov(MOD_LECTURA_CRITICA ~ Estudiante_trabaja_cod,
+ data=Dataset)
> summary(AnovaModel.15)
      Df Sum Sq Mean Sq F value Pr(>F)
Estudiante_trabaja_cod  3   2.09   0.6973   1.369  0.254
Residuals           194  98.84   0.5095
2 observations deleted due to missingness
> |
```

20. Test de varianzas Horas de trabajo vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Horas_trabajo_cod, var, na.rm=TRUE))
De 11 a 15 horas De 16 a 20 horas De 6 a 10 horas
0.5358498      0.5243526      0.4149244
> leveneTest(MOD_LECTURA_CRITICA ~ Horas_trabajo_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  0.1398 0.8697
      95
>
> AnovaModel.16 <- aov(MOD_LECTURA_CRITICA ~ Horas_trabajo_cod, data=Dataset)
> summary(AnovaModel.16)
              Df Sum Sq Mean Sq F value Pr(>F)
Horas_trabajo_cod  2    1.52   0.7597   1.557  0.216
Residuals          95   46.35   0.4879
102 observations deleted due to missingness
> |
```

21. Test de varianzas Ingresos mensuales en salarios mínimos vs Lectura crítica

```
> with(Dataset, tapply(MOD_LECTURA_CRITICA, Ingresos_cod, var, 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 Entre 7 y menos de 10 SM Menos de 1 SM
0.5124327      0.4273933      0.4267235      1.0711111      NA      0.6779000
> leveneTest(MOD_LECTURA_CRITICA ~ Ingresos_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  1.5387 0.1796
      192
>
> AnovaModel.17 <- aov(MOD_LECTURA_CRITICA ~ Ingresos_cod, data=Dataset)
> summary(AnovaModel.17)
              Df Sum Sq Mean Sq F value Pr(>F)
Ingresos_cod  5    1.22   0.2444   0.471  0.798
Residuals    192   99.71   0.5193
2 observations deleted due to missingness
> |
```

1. Test de varianzas sexo vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, ESTU_GENERO, var, na.rm=TRUE))
      F      M
1.1204378 0.9868417
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_GENERO, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.4568 0.2289
      193
>
> t.test(MOD_COMUNICA_ESCRITA_PUNT~ESTU_GENERO, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_COMUNICA_ESCRITA_PUNT by ESTU_GENERO
t = 1.1772, df = 193, p-value = 0.2405
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1172643  0.4645227
sample estimates:
mean in group F mean in group M
 10.18372      10.01009
```

2. Test de varianzas Estado civil vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Estado_civil_cod, var, na.rm=TRUE))
  Casado Separado Soltero Unión libre
  0.363000 0.520000 1.011608 1.789167
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Estado_civil_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  0.939 0.4229
      191
> AnovaModel.2 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Estado_civil_cod, data=Dataset)
> summary(AnovaModel.2)
              Df Sum Sq Mean Sq F value Pr(>F)
Estado_civil_cod  3  11.29   3.764   3.745 0.012 *
Residuals       191 191.97   1.005
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
5 observations deleted due to missingness
> |
```

3. Test de varianzas situación de su hogar actual vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Hogar_actual, var, na.rm=TRUE))
Permanente temporal
  0.9611226 1.2130378
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Hogar_actual, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.8157 0.3676
      194
> t.test(MOD_COMUNICA_ESCRITA_PUNT~Hogar_actual, alternative='two.sided',
+ conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_COMUNICA_ESCRITA_PUNT by Hogar_actual
t = -1.571, df = 194, p-value = 0.1178
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.53541295  0.06063609
sample estimates:
mean in group Permanente mean in group temporal
           10.00661           10.24400
```

4. Test de varianzas Cabeza de familia vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Cabeza_familia_cod, var,
+ na.rm=TRUE))
  No Si
  1.0562116 0.9067778
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Cabeza_familia_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0048 0.9449
      194
> t.test(MOD_COMUNICA_ESCRITA_PUNT~Cabeza_familia_cod, alternative='two.sided',
+ conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_COMUNICA_ESCRITA_PUNT by Cabeza_familia_cod
t = 1.9884, df = 194, p-value = 0.04817
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.005365457 1.317000134
sample estimates:
mean in group No mean in group Si
           10.13118           9.47000
```

5. Test de varianzas Estrato vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Estrato_cod, var, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5
1.9264286 0.8977341 1.1440541 0.4400000      NA
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Estrato_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  1.8689 0.1175
      190
> AnovaModel.6 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Estrato_cod, data=Dataset)
> summary(AnovaModel.6)
      Df Sum Sq Mean Sq F value Pr(>F)
Estrato_cod  4    4.3   1.075   1.026  0.395
Residuals  190  199.0   1.047
5 observations deleted due to missingness
> |
```

6. Test de varianzas Valor anual de la matrícula vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Valor_anual_matricula_cod, var,
+ na.rm=TRUE))
      Entre 1 millón y 3 millones      Entre 3 millones y 5 millones Entre 500 mil y menos de de 1 millón      Menos de 500 mil
      1.0668647                    0.9433333                    NA                    1.2956593
      No pagó matrícula
      0.7371429
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT~ Valor_anual_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  4  0.8505 0.4949
      190
>
> AnovaModel.7 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Valor_anual_matricula_cod, data=Dataset)
> summary(AnovaModel.7)
      Df Sum Sq Mean Sq F value Pr(>F)
Valor_anual_matricula_cod  4  2.51  0.6286  0.589  0.671
Residuals                  190 202.63  1.0665
5 observations deleted due to missingness
> |
```

7. Test de varianzas Forma de pagó matrícula vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Forma_de_pagó_matricula_cod, var,
+ na.rm=TRUE))
      No cancelo matrícula      Pagó por beca      Pagó por credito      Pagó por padres      Pagó propio Varias formas de pagó
      0.7033516                0.8728458                2.1740000                1.0894444                1.6380000                0.9175123
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Forma_de_pagó_matricula_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  1.6657 0.1448
      190
>
> AnovaModel.8 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Forma_de_pagó_matricula_cod,
+ data=Dataset)
> summary(AnovaModel.8)
      Df Sum Sq Mean Sq F value Pr(>F)
Forma_de_pagó_matricula_cod  5  1.3  0.2592  0.239  0.945
Residuals                    190 206.4  1.0864
4 observations deleted due to missingness
> |
```

8. Test de varianzas Ocupación del padre vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNTI, Ocupación_padre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      0.7776667                        0.4425000                        1.0187273
      empleado de nivel técnico o profesional Empleado obrero u operario Empresario
      0.8064211                        0.8737385                        0.5366667
      Hogar Otra actividad u ocupación Pensionado
      0.4520000                        1.2654762                        1.1980291
      Pequeño empresario Profesional independiente Trabajador por cuenta propia
      0.9584762                        1.1761111                        1.2298920

> leveneTest(MOD_COMUNICA_ESCRITA_PUNTI ~ Ocupación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  11  0.5513 0.8658
      184

>
> AnovaModel.9 <- aov(MOD_COMUNICA_ESCRITA_PUNTI ~ Ocupación_padre_cod, data=Dataset)
> summary(AnovaModel.9)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_padre_cod  11  17.12  1.557  1.503 0.133
Residuals          184 190.59  1.036
4 observations deleted due to missingness
> |
```

9. Test de varianzas Ocupación de la madre vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNTI, Ocupación_madre_cod, var,
+ na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo empleado de nivel directivo
      1.0120000                        0.8451515                        1.1925641
      empleado de nivel técnico o profesional Empleado obrero u operario Hogar
      0.8853333                        1.6880000                        1.0513037
      Otra actividad u ocupación Pensionado Pequeño empresario
      0.8909524                        1.1602747                        0.7809341
      Profesional independiente Trabajador por cuenta propia
      0.3057143                        1.4098495

> leveneTest(MOD_COMUNICA_ESCRITA_PUNTI ~ Ocupación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  10  0.6359 0.7818
      185

>
> AnovaModel.10 <- aov(MOD_COMUNICA_ESCRITA_PUNTI ~ Ocupación_madre_cod, data=Dataset)
> summary(AnovaModel.10)
      Df Sum Sq Mean Sq F value Pr(>F)
Ocupación_madre_cod  10  8.13  0.813  0.754 0.673
Residuals          185 199.58  1.079
4 observations deleted due to missingness
\ |
```

10. Test de varianzas El hogar cuenta con internet vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNTI, El_hogar_cuenta_con_internet_cod,
+ var, na.rm=TRUE))
      No Si
0.9670837 1.1173394
> leveneTest(MOD_COMUNICA_ESCRITA_PUNTI ~ El_hogar_cuenta_con_internet_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.2211 0.6387
      194

>
> t.test(MOD_COMUNICA_ESCRITA_PUNTI ~ El_hogar_cuenta_con_internet_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_COMUNICA_ESCRITA_PUNTI by El_hogar_cuenta_con_internet_cod
t = -1.9184, df = 194, p-value = 0.05653
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.573353536 0.007941747
sample estimates:
mean in group No mean in group Si
 9.94023          10.22294
```

11. Test de varianzas El hogar cuenta con celular vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, El_hogar_cuenta_con_celular_cod, var,
+   na.rm=TRUE))
      No      Si
1.3536353 0.9936037
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ El_hogar_cuenta_con_celular_cod, data=Dataset,
+   center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  3.1532 0.07734 .
      194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> t.test(MOD_COMUNICA_ESCRITA_PUNT~El_hogar_cuenta_con_celular_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_COMUNICA_ESCRITA_PUNT by El_hogar_cuenta_con_celular_cod
t = -0.41958, df = 194, p-value = 0.6753
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4306311  0.2795488
sample estimates:
mean in group No mean in group Si
      10.03810      10.11364
```

12. Test de varianzas El hogar cuenta con televisión vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, El_hogar_cuenta_con_televisión_cod,
+   var, na.rm=TRUE))
      No      Si
1.3346298 0.9481472
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT~ El_hogar_cuenta_con_televisión_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  2.2759 0.133
      194
>
> t.test(MOD_COMUNICA_ESCRITA_PUNT~El_hogar_cuenta_con_televisión_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_COMUNICA_ESCRITA_PUNT by El_hogar_cuenta_con_televisión_cod
t = 0.42447, df = 194, p-value = 0.6717
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2458783  0.3807363
sample estimates:
mean in group No mean in group Si
      10.14355      10.07612
```

13. Test de varianzas El hogar cuenta con lavadora vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, El_hogar_cuenta_con_lavadora_cod,
+   var, na.rm=TRUE))
      No      Si
0.9276234 1.1044702
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT~ El_hogar_cuenta_con_lavadora_cod,
+   data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.1396 0.7091
      194
>
> t.test(MOD_COMUNICA_ESCRITA_PUNT~El_hogar_cuenta_con_lavadora_cod,
+   alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data:  MOD_COMUNICA_ESCRITA_PUNT by El_hogar_cuenta_con_lavadora_cod
t = -1.7334, df = 194, p-value = 0.08461
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.60163111  0.03877397
sample estimates:
mean in group No mean in group Si
      9.896429      10.177857
```

14. Test de varianzas El hogar cuenta con teléfono vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, El_hogar_cuenta_con_teléfono_cod,
+ var, na.rm=TRUE))
      No      Si
0.9413328 1.2100938
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ El_hogar_cuenta_con_teléfono_cod,
+ data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.9278 0.1666
      194
>
> t.test(MOD_COMUNICA_ESCRITA_PUNT~El_hogar_cuenta_con_teléfono_cod,
+ alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_COMUNICA_ESCRITA_PUNT by El_hogar_cuenta_con_teléfono_cod
t = 0.29856, df = 194, p-value = 0.7656
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2475542  0.3358729
sample estimates:
mean in group No mean in group Si
      10.11863      10.07447
```

15. Test de varianzas Tomó algún curso de preparación al examen vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Tomó_curso, var, na.rm=TRUE))
      No      Sí
0.9630017 1.1347043
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Tomó_curso, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.4808 0.4891
      158
> t.test(MOD_COMUNICA_ESCRITA_PUNT~Tomó_curso, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Dataset)

Two Sample t-test

data: MOD_COMUNICA_ESCRITA_PUNT by Tomó_curso
t = -1.2269, df = 158, p-value = 0.2217
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5713975  0.1335229
sample estimates:
mean in group No mean in group Sí
      9.930612      10.149550
```

16. Test de varianzas Tipo de bachillerato del que se graduó vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, ESTU_TITULO_BTO, var, na.rm=TRUE))
      A      N      T
1.110973 0.997000 1.028151
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_TITULO_BTO, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  0.0753 0.9275
      193
> AnovaModel.12 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ ESTU_TITULO_BTO, data=Dataset)
> summary(AnovaModel.12)
      Df Sum Sq Mean Sq F value Pr(>F)
ESTU_TITULO_BTO  2  0.04  0.0206  0.019  0.981
Residuals      193 207.67  1.0760
4 observations deleted due to missingness
> |
```

17. Test de varianzas Educación del padre vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Educación_padre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      1.0430920      0.5880952      1.0841846
Educación técnica o tecnologica incompleta      0.9747619      Ninguno      Potsgrado
      Primaria completa      Primaria incompleta      Secundaria completa
      1.2285263      0.8745385      0.7312418
      Secundaria incompleta      0.5956410
      1.3666749
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Educación_padre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  9  0.7997  0.617
      186
>
> AnovaModel.13 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Educación_padre_cod, data=Dataset)
> summary(AnovaModel.13)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_padre_cod  9  21.15  2.350  2.343 0.0159 *
Residuals          186 186.55  1.003
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 observations deleted due to missingness
> |
```

18. Test de varianzas Educación de la madre vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Educación_madre_cod, var,
+ na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
Educación profesional completa      0.9206587      0.6630000      0.5647059
Educación técnica o tecnologica incompleta      0.4670000      Ninguno      No sabe
      Potsgrado      NA      NA
      1.4325000      Primaria completa      Primaria incompleta
      1.4415152      0.7604167
      Secundaria completa      Secundaria incompleta
      1.1191767      0.7980952
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Educación_madre_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  9  0.9993  0.4422
      186
>
> AnovaModel.14 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Educación_madre_cod, data=Dataset)
> summary(AnovaModel.14)
      Df Sum Sq Mean Sq F value Pr(>F)
Educación_madre_cod  9  27.87  3.0965  3.203 0.00125 **
Residuals          186 179.84  0.9669
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 observations deleted due to missingness
> |
```

19. Test de varianzas Trabaja actualmente vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Estudiante_trabaja_cod, var,
+ na.rm=TRUE))
      No      Si, como ayudante sin remuneración      Si, por experiencia o recursos en gastos generales
Si, por ser práctica obligatoria del plan de estudios      1.1133484      1.1729905      0.6589037
      0.0800000
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT~ Estudiante_trabaja_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  3  1.9712  0.1197
      192
>
> AnovaModel.15 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Estudiante_trabaja_cod,
+ data=Dataset)
> summary(AnovaModel.15)
      Df Sum Sq Mean Sq F value Pr(>F)
Estudiante_trabaja_cod  3  6.34  2.113  2.014 0.113
Residuals          192 201.37  1.049
4 observations deleted due to missingness
> |
```

20. Test de varianzas Horas de trabajo vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Horas_trabajo_cod, var, na.rm=TRUE))
De 11 a 15 horas De 16 a 20 horas De 6 a 10 horas
 0.6590909      1.3115641      0.8649916
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Horas_trabajo_cod, data=Dataset,
+ center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  2  1.1935 0.3077
      95
>
> AnovaModel.16 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Horas_trabajo_cod, data=Dataset)
> summary(AnovaModel.16)
              Df Sum Sq Mean Sq F value Pr(>F)
Horas_trabajo_cod  2    0.68   0.3409   0.341  0.712
Residuals          95   95.06   1.0006
102 observations deleted due to missingness
> |
```

21. Test de varianzas Ingresos mensuales en salarios mínimos vs Comunicación escrita

```
> with(Dataset, tapply(MOD_COMUNICA_ESCRITA_PUNT, Ingresos_cod, var, 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 Entre 7 y menos de 10 SM Menos de 1 SM
1.2650207      1.0991077      0.9536190      0.4405556      NA      0.9447333
> leveneTest(MOD_COMUNICA_ESCRITA_PUNT ~ Ingresos_cod, data=Dataset, center="median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  5  1.0721 0.3772
      190
>
> AnovaModel.17 <- aov(MOD_COMUNICA_ESCRITA_PUNT ~ Ingresos_cod, data=Dataset)
> summary(AnovaModel.17)
              Df Sum Sq Mean Sq F value Pr(>F)
Ingresos_cod  5    1.57   0.3145   0.29  0.918
Residuals    190  206.14   1.0849
4 observations deleted due to missingness
> |
```

Como el puntaje en inglés no asume una distribución normal se realizó el test no paramétrico de Kruskal Wallis, donde se determinó asociación entre algunas variables y el puntaje en inglés.

1. Test no paramétrico Kruskal-Wallis sexo vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, ESTU_GENERO, median, na.rm=TRUE))
  F  M
10.1 10.1
> kruskal.test(MOD_INGLES_PUNT ~ ESTU_GENERO, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by ESTU_GENERO
Kruskal-Wallis chi-squared = 1.189, df = 1, p-value = 0.2755
```

2. Test no paramétrico Kruskal-Wallis Estado civil vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Estado_civil_cod, median, na.rm=TRUE))
  Casado  Separado  Soltero Unión libre
    10.0    10.2    10.1    11.1
> kruskal.test(MOD_INGLES_PUNT ~ Estado_civil_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Estado_civil_cod
Kruskal-Wallis chi-squared = 3.8327, df = 3, p-value = 0.2801
```

3. Test no paramétrico Kruskal-Wallis situación de su hogar actual vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Hogar_actual, median, na.rm=TRUE))
Permanente  temporal
    10.08    10.30
> kruskal.test(MOD_INGLES_PUNT ~ Hogar_actual, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Hogar_actual
Kruskal-Wallis chi-squared = 2.0459, df = 1, p-value = 0.1526
```

4. Test no paramétrico Kuskal-Wallis Cabeza de familia vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Cabeza_familia_cod, median, na.rm=TRUE))
  No  Si
10.10 9.95
> kruskal.test(MOD_INGLES_PUNT ~ Cabeza_familia_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Cabeza_familia_cod
Kruskal-Wallis chi-squared = 0.90651, df = 1, p-value = 0.341
```

5. Test no paramétrico Kruskal-Wallis Estrato vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Estrato_cod, median, na.rm=TRUE))
Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5
    10.06    10.06    10.30    10.08    9.00
> kruskal.test(MOD_INGLES_PUNT ~ Estrato_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Estrato_cod
Kruskal-Wallis chi-squared = 8.8728, df = 4, p-value = 0.06436
```

6. Test no paramétrico Kruskal-Wallis Valor anual de la matrícula vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNTI, Valor_anual_matricula_cod, median, na.rm=TRUE))
      Entre 1 millón y 3 millones      Entre 3 millones y 5 millones Entre 500 mil y menos de de 1 millón      Menos de 500 mil
                10.10                10.50                9.90                10.18
      No pagó matrícula
                10.11
> kruskal.test(MOD_INGLES_PUNTI ~ Valor_anual_matricula_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNTI by Valor_anual_matricula_cod
Kruskal-Wallis chi-squared = 2.3475, df = 4, p-value = 0.6721
```

7. Test no paramétrico Kruskal-Wallis Forma de pagó matrícula vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNTI, Forma_de_pagó_matricula_cod, median, na.rm=TRUE))
      No cancelo matrícula      Pagó por beca      Pagó por credito      Pagó por padres      Pagó propio Varias formas de pagó
                10.045                10.100                10.000                9.860                10.300                10.300
> kruskal.test(MOD_INGLES_PUNTI ~ Forma_de_pagó_matricula_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNTI by Forma_de_pagó_matricula_cod
Kruskal-Wallis chi-squared = 5.9072, df = 5, p-value = 0.3154
```

8. Test no paramétrico Kruskal-Wallis Ocupación del padre vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNTI, Ocupación_padre_cod, median, na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo      empleado de nivel directivo
                11.100                10.050                10.100
      empleado de nivel técnico o profesional      Empleado obrero u operario      Empresario
                10.100                9.945                10.400
      Hogar      Otra actividad u ocupación      Pensionado
                10.130                10.100                10.150
      Pequeño empresario      Profesional independiente      Trabajador por cuenta propia
                10.000                10.600                10.300
> kruskal.test(MOD_INGLES_PUNTI ~ Ocupación_padre_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNTI by Ocupación_padre_cod
Kruskal-Wallis chi-squared = 12.996, df = 11, p-value = 0.2936
```

9. Test no paramétrico Kruskal-Wallis Ocupación de la madre vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNTI, Ocupación_madre_cod, median, na.rm=TRUE))
      Empleado como director o gerente Empleado de nivel auxiliar o administrativo      empleado de nivel directivo
                10.900                9.845                10.000
      empleado de nivel técnico o profesional      Empleado obrero u operario      Hogar
                10.400                10.400                10.300
      Otra actividad u ocupación      Pensionado      Pequeño empresario
                10.000                10.700                10.450
      Profesional independiente      Trabajador por cuenta propia
                10.300                10.060
> kruskal.test(MOD_INGLES_PUNTI ~ Ocupación_madre_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNTI by Ocupación_madre_cod
Kruskal-Wallis chi-squared = 13.099, df = 10, p-value = 0.2182
```

10. Test no paramétrico Kruskal-Wallis El hogar cuenta con internet vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, El_hogar_cuenta_con_internet_cod, median, na.rm=TRUE))
  No  Si
10.0 10.3
> kruskal.test(MOD_INGLES_PUNT ~ El_hogar_cuenta_con_internet_cod, data=Dataset)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by El_hogar_cuenta_con_internet_cod
Kruskal-Wallis chi-squared = 5.9867, df = 1, p-value = 0.01441
```

11. Test no paramétrico Kruskal-Wallis El hogar cuenta con celular vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, El_hogar_cuenta_con_celular_cod, median, na.rm=TRUE))
  No  Si
10.2 10.1
> kruskal.test(MOD_INGLES_PUNT ~ El_hogar_cuenta_con_celular_cod, data=Dataset)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by El_hogar_cuenta_con_celular_cod
Kruskal-Wallis chi-squared = 0.070559, df = 1, p-value = 0.7905
```

12. Test no paramétrico Kruskal-Wallis El hogar cuenta con televisión vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, El_hogar_cuenta_con_televisión_cod, median,
+ na.rm=TRUE))
  No  Si
10.06 10.30
> kruskal.test(MOD_INGLES_PUNT ~ El_hogar_cuenta_con_televisión_cod, data=Dataset)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by El_hogar_cuenta_con_televisión_cod
Kruskal-Wallis chi-squared = 2.8422, df = 1, p-value = 0.09182
```

13. Test no paramétrico Kruskal-Wallis El hogar cuenta con lavadora vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, El_hogar_cuenta_con_lavadora_cod, median, na.rm=TRUE))
  No  Si
9.98 10.25
> kruskal.test(MOD_INGLES_PUNT ~ El_hogar_cuenta_con_lavadora_cod, data=Dataset)

Kruskal-Wallis rank sum test

data:  MOD_INGLES_PUNT by El_hogar_cuenta_con_lavadora_cod
Kruskal-Wallis chi-squared = 5.754, df = 1, p-value = 0.01645
```

14. Test no paramétrico Kruskal-Wallis El hogar cuenta con teléfono vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, El_hogar_cuenta_con_teléfono_cod, median, na.rm=TRUE))
  No  Si
10.08 10.30
> kruskal.test(MOD_INGLES_PUNT ~ El_hogar_cuenta_con_teléfono_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by El_hogar_cuenta_con_teléfono_cod
Kruskal-Wallis chi-squared = 2.0816, df = 1, p-value = 0.1491
```

15. Test no paramétrico Kruskal-Wallis Tomó algún curso de preparación vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Tomó_curso, median, na.rm=TRUE))
  No  Sí
10.3 10.1
> kruskal.test(MOD_INGLES_PUNT ~ Tomó_curso, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Tomó_curso
Kruskal-Wallis chi-squared = 0.018226, df = 1, p-value = 0.8926
```

16. Test no paramétrico Kruskal-Wallis Tipo de bachillerato del que se graduó vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, ESTU_TITULO_BTO, median, na.rm=TRUE))
  A  N  T
10.25 9.80 10.00
> kruskal.test(MOD_INGLES_PUNT ~ ESTU_TITULO_BTO, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by ESTU_TITULO_BTO
Kruskal-Wallis chi-squared = 2.4475, df = 2, p-value = 0.2941
```

17. Test no paramétrico Kruskal-Wallis Educación del padre vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Educación_padre_cod, median, na.rm=TRUE))
Educación profesional completa Educación profesional incompleta Educación técnica o tecnologica completa
10.500 10.300 10.350
Educación técnica o tecnologica incompleta Ninguno Postgrado
9.615 10.100 10.150
Primaria completa Primaria incompleta Secundaria completa
9.900 10.000 10.300
Secundaria incompleta
10.060
> kruskal.test(MOD_INGLES_PUNT ~ Educación_padre_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Educación_padre_cod
Kruskal-Wallis chi-squared = 8.7761, df = 9, p-value = 0.4582
```

18. Test no paramétrico Kruskal-Wallis Educación de la madre vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Educación_madre_cod, median, na.rm=TRUE))
      Educación profesional completa      Educación profesional incompleta      Educación técnica o tecnologica completa
      10.400                          10.300                          10.300
Educación técnica o tecnologica incompleta      Ninguno      No sabe
      10.460                          8.200                          9.730
      Potsgrado      Primaria completa      Primaria incompleta
      10.470                          9.795                          10.000
      Secundaria completa      Secundaria incompleta
      10.060                          10.060
> kruskal.test(MOD_INGLES_PUNT ~ Educación_madre_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Educación_madre_cod
Kruskal-Wallis chi-squared = 14.327, df = 10, p-value = 0.1586
```

19. Test no paramétrico Kruskal-Wallis Trabaja actualmente vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Estudiante_trabaja_cod, median, na.rm=TRUE))
      No      Si, como ayudante sin remuneración      Si, por experiencia o recursos en gastos generales
      10.25      10.06      10.25
Si, por ser práctica obligatoria del plan de estudios
      10.50
> kruskal.test(MOD_INGLES_PUNT ~ Estudiante_trabaja_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Estudiante_trabaja_cod
Kruskal-Wallis chi-squared = 3.1246, df = 3, p-value = 0.3728
```

20. Test no paramétrico Kruskal-Wallis Horas de trabajo vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Horas_trabajo_cod, median, na.rm=TRUE))
De 11 a 15 horas De 16 a 20 horas De 6 a 10 horas
      10.10      10.08      10.10
> kruskal.test(MOD_INGLES_PUNT ~ Horas_trabajo_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Horas_trabajo_cod
Kruskal-Wallis chi-squared = 2.1999, df = 2, p-value = 0.3329
```

21. Test no paramétrico Kruskal-Wallis Ingresos mensuales en salarios mínimos vs Inglés

```
> with(Dataset, tapply(MOD_INGLES_PUNT, Ingresos_cod, 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 Entre 7 y menos de 10 SM      Menos de 1 SM
      10.10      10.10      10.40      10.65      10.10      9.86
> kruskal.test(MOD_INGLES_PUNT ~ Ingresos_cod, data=Dataset)

Kruskal-Wallis rank sum test

data: MOD_INGLES_PUNT by Ingresos_cod
Kruskal-Wallis chi-squared = 8.4365, df = 5, p-value = 0.1338
```

REFERENCIAS BIBLIOGRÁFICAS

- AGRESTI, A. (2002), *AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS*, University of Florida.
- Arias, I. (2014), ‘INFLUENCIA DE LOS PADRES EN EL RENDIMIENTO ACADÉMICO DE SUS HIJOS: UNA APROXIMACIÓN ECONÓMICA EN EL CONTEXTO DE LA EDUCACIÓN MEDIA COLOMBIANA. EDUCACIÓN Y DESARROLLO SOCIAL’, *REVISTA EDUCACIÓN Y DESARROLLO SOCIAL* **8**(2), 184–199.
- Bogoya, D. (2013), ‘EDUCACIÓN SUPERIOR EN COLOMBIA.’, Universidad Nacional.
- Camarero, L. (2013), *ESTADÍSTICA PARA LA INVESTIGACIÓN SOCIAL*, Garceta.
- Canadas, L. (2013), *REGRESIÓN LOGÍSTICA TRATAMIENTO COMPUTACIONAL EN R*, Universidad de Granada.
- CNA (2013), *LINEAMIENTOS PARA LA ACREDITACIÓN DE PROGRAMAS DE PREGRADO*, Consejo Nacional de Acreditación (CNA).
- Díaz, L. (2012), *ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS*, Universidad Nacional de Colombia.
- Dobson, A. (1990), *AN INTRODUCTION TO GENERALIZED LINEAR MODELS*, University of Newcastle.
- Fabián G Alejandra S Luz Rondón L Martín Alonso Gómez C, . (2013), ‘IMPACTO DE LAS FACULTADES DE MEDICINA Y DE LOS ESTUDIANTES SOBRE LOS RESULTADOS EN LA PRUEBA NACIONAL DE CALIDAD DE LA EDUCACIÓN SUPERIOR (SABER PRO)’ , *Revista Colombiana de Anestesiología* **41**(3), 203–204.
- Icfes, . (2012), ‘LINEAMIENTOS PARA EL EXAMEN SABER PRO 2012-1’ , Icfes, 2012.
- Icfes, . (2017), ‘GUÍA DE INTERPRETACIÓN Y USO DE RESULTADOS DEL EXAMEN SABER PRO’ , Icfes.

- Iglesias, T. (2013), *MÉTODOS DE BONDAD DE AJUSTE EN REGRESIÓN LOGÍSTICA*, Universidad de la Granada.
- Mesa, H. (2009), *DE LOS ECAES LA AUTONOMÍA UNIVERSITARIA Y EL DERECHO*, Revista electrónica.
- Mora Borbón, W. A. (2014), 'EDICIÓN DE TEXTOS CIENTÍFICOS LATEX', Revista digital.
- Mundial, . (2012), 'LA EDUCACIÓN SUPERIOR EN COLOMBIA 2012', OCDE,2012.
- Pérez, M. (2016), *MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALA Y FORMA (GAMLSS)*, Universidad de Vigo.
- Ramírez, C. (2014a), *FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO SEGÚN NIVEL DE FORMACIÓN PREGRADO Y GÉNERO DE LOS ESTUDIANTES DE EDUCACIÓN SUPERIORE EN COLOMBIA*, Revista Colombiana de Educacin. 220-221.
- Ramírez, C. (2014b), 'FACTORES SOCIECONÓMICOS Y EDUCATIVOS ASOCADOS CON EL DESEMPEÑO ACADÉMICO, SEGÚN NIVEL DE FORMACIÓN Y GÉNERO DE LOS ESTUDIANTES QUE PRESENTARON LA PRUEBA SABER PRO 2009', Icfes.
- Rigby, B. (2015), *FLEXIBLE REGRESSION AND SMOOTHING THE GAMLSS PACKAGES IN R*, Mikis Stasinopoulos, Bob Rigby, Vlasios Voudouris, Gillian Heller and Fernanda De Bastiani.
- Sánchez, A. (2013), *BILINGUISMO EN COLOMBIA*, Banco de la República.
- Sánchez, I. d. J. (2010), 'GUÍA DE INTERPRETACIÓN DE RESULTADOS DE LAS PRUEBAS SABER PRO', Icfes.