

**FACTORES ASOCIADOS AL RENDIMIENTO EN LAS PRUEBAS SABER PRO
2010 DE LOS ESTUDIANTES DE INGENIERIA CIVIL**

**ADRIÁN DAVID MOJICA PERDOMO
WOLFANG ALEXANDER OSMA CASTELLANOS**



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2012**

**FACTORES ASOCIADOS AL RENDIMIENTO EN LAS PRUEBAS SABER-PRO
2010 DE LOS ESTUDIANTES DE INGENIERIA CIVIL**

**ADRIÁN DAVID MOJICA PERDOMO
WOLFANG ALEXANDER OSMA CASTELLANOS**

**Trabajo de Tesis para optar el grado académico de Especialista en
Estadística**

**Directora
Mg. TULIA ESTHER RIVERA FLÓREZ**



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2012**

AGRADECIMIENTOS

Queremos expresar nuestro agradecimiento a todos y cada una de las personas que contribuyeron de una u otra manera en la elaboración de esta tesis.

En primer lugar deseamos agradecer profundamente a nuestras familias, quienes nos brindaron el apoyo y la energía necesaria para poder llegar hoy a la culminación de este estudio.

A nuestra directora, Tulia Esther Rivera por sus orientaciones en la realización de esta tesis.

A cada uno de los profesores que nos brindaron sus experiencias y orientaciones a lo largo de nuestros estudios de especialización.

Nuestro mayor reconocimiento al Dr. Gabriel Yáñez Canal, el Profe Gabriel, por su constante apoyo y presión para avanzar en este estudio, por sus críticas, comentarios y amplios conocimientos que enriquecieron y direccionaron este trabajo. Profe, gracias por sus consejos que contribuyeron a nuestra formación profesional y a culminar de la mejor manera esta tesis.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	16
1. GENERALIDADES DE LA PRUEBA SABER PRO	20
1.1 EVALUACIÓN POR COMPETENCIAS Y PRUEBAS SABER PRO PARA ESTUDIANTES DE INGENIERÍA CIVIL	23
1.2 DESCRIPCIÓN DE LA BASE DE DATOS	24
2. ANTECEDENTES	29
3. MARCO TEÓRICO.....	36
3.1 TABLAS DE CONTINGENCIA (TC).....	36
3.2 DEPENDENCIA ENTRE VARIABLES	39
3.3 ANÁLISIS DE CORRESPONDENCIAS	42
3.3.1 Análisis de Correspondencia Simple (ACS)	43
3.3.2 Análisis de Correspondencias Múltiple (ACM)	49
4. RESULTADOS DEL ANÁLISIS DE CORRESPONDENCIAS	56
4.1 RESULTADOS DE LOS ESTUDIANTES	56
4.1.1 Niveles de las variables categóricas	57

4.1.2 Descripción de la variable <i>PUNTAJE</i>	60
4.1.3 Tablas de contingencia e Independencia de las variables.	62
4.1.4 Análisis de Correspondencia Simple (ACS).	73
4.1.5 Análisis de Correspondencias Múltiple (ACM)	82
4.2 RESULTADOS DE LAS INSTITUCIONES.....	88
4.2.1. Niveles de las variables categóricas	88
4.2.2 Descripción de la variable <i>PUNTAJE_PROM</i>	90
4.2.3 Análisis de Correspondencia Simple (ACS)	93
4.2.4 Análisis de Correspondencias Múltiple (ACM)	100
CONCLUSIONES.....	106
BIBLIOGRAFÍA.....	109

LISTA DE TABLAS

Pág.

<i>Tabla 1. Tabla de contingencia general f-filas * c-columnas.....</i>	<i>37</i>
<i>Tabla 2. Tabla de perfiles o frecuencias relativas condicionales.</i>	<i>39</i>
<i>Tabla 3. Tabla disyuntiva completa</i>	<i>50</i>
<i>Tabla 4. Tabla resumen de respuestas para un conjunto de n individuos.</i>	<i>51</i>
<i>Tabla 5. Tabla disyuntiva completa para los datos expuestos en la tabla 4.</i>	<i>52</i>
<i>Tabla 6. Variables categóricas y sus respectivos niveles.....</i>	<i>59</i>
<i>Tabla 7. Descriptivos, histogramas, diagramas de caja y pruebas de normalidad para la variable PUNTAJE.....</i>	<i>60</i>
<i>Tabla 8. Tabla de contingencia GÉNERO * PUNTAJE.....</i>	<i>62</i>
<i>Tabla 9. Tabla de contingencia TIPO * PUNTAJE.....</i>	<i>64</i>
<i>Tabla 10. Tabla de contingencia EDUCACIÓN PADRE * PUNTAJE</i>	<i>65</i>
<i>Tabla 11. Tabla de contingencia EDUCACIÓN MADRE * PUNTAJE.....</i>	<i>67</i>
<i>Tabla 12. Tabla de contingencia ESTRATO * PUNTAJE.....</i>	<i>68</i>
<i>Tabla 13. Tabla de contingencia TRABAJA * PUNTAJE</i>	<i>70</i>

<i>Tabla 14. Tabla de contingencia TITULO * PUNTAJE</i>	<i>71</i>
<i>Tabla 15. Pruebas Ji-cuadrado de Pearson.....</i>	<i>72</i>
<i>Tabla 16. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable género.</i>	<i>73</i>
<i>Tabla 17. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable tipo.</i>	<i>75</i>
<i>Tabla 18. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Educación Padre.....</i>	<i>76</i>
<i>Tabla 19. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Educación Madre.</i>	<i>78</i>
<i>Tabla 20. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Estrato.....</i>	<i>80</i>
<i>Tabla 21. Valores propios, coordenadas y contribuciones.</i>	<i>82</i>
<i>Tabla 22 . Variables categóricas y sus niveles para resultados de las instituciones.</i>	<i>90</i>
<i>Tabla 23. Descriptivos, histogramas, diagramas de caja y pruebas de normalidad para la variable PUNTAJE_ PROM.....</i>	<i>91</i>
<i>Tabla 24. Tabla unificada para la variable acreditado * puntaje_prom.</i>	<i>93</i>
<i>Tabla 25. Tabla unificada para la variable ciudad * puntaje_prom.</i>	<i>95</i>
<i>Tabla 26. Tabla unificada para la variable tipo* puntaje_prom.....</i>	<i>97</i>
<i>Tabla 27. Tabla unificada para la variable estudiantes * puntaje_prom</i>	<i>98</i>

<i>Tabla 28. Valores propios, coordenadas y contribuciones.</i>	<i>101</i>
--	------------

LISTA DE CUADROS

Pág.

<i>Cuadro 1. Modelo de evaluación del ICFES para estudiantes de Ingeniería Civil.</i>	<i>24</i>
<i>Cuadro 2. Número de estudiantes que han presentado la prueba saber Pro en el período 2004-2010.</i>	<i>25</i>
<i>Cuadro 3. Componentes evaluadas en las versiones de las Pruebas Saber Pro 2004-2010</i>	<i>27</i>
<i>Cuadro 4. Estructura de la prueba Saber Pro.</i>	<i>28</i>
<i>Cuadro 5. Evolución de los puntajes promedio del programa de Ingeniería Civil.</i>	<i>33</i>
<i>Cuadro 6. Puntajes promedio de hombres y mujeres evaluados en el 2009.</i>	<i>34</i>

LISTA DE FIGURAS

Pág.

<i>Figura 1. Plano factorial: eje 1 * eje 2.</i>	<i>84</i>
<i>Figura 2. Plano factorial: eje 1 * eje 3.</i>	<i>85</i>
<i>Figura 3. Plano factorial: eje 1 * eje 4.</i>	<i>86</i>
<i>Figura 4. Plano factorial: eje 2 * eje 3.</i>	<i>87</i>
<i>Figura 5. Plano factorial: eje 1 * eje 2.</i>	<i>103</i>
<i>Figura 6. Plano factorial: eje 1 * eje 3.</i>	<i>104</i>

1. TÍTULO:

FACTORES ASOCIADOS AL RENDIMIENTO EN LAS PRUEBAS SABER PRO 2010 DE LOS ESTUDIANTES DE INGENIERIA CIVIL*.

2. AUTORES:

Adrián David Mojica Perdomo
Wolfgang Alexander Osma Castellanos**

3. PALABRAS CLAVES:

Pruebas Saber Pro
Ingeniería Civil
Puntaje
Tablas de contingencia
Análisis de correspondencias

4. RESUMEN:

En este estudio se realizaron análisis estadísticos alrededor de los resultados obtenidos por los estudiantes de último año de Ingeniería Civil a nivel nacional en las pruebas académicas de vigilancia de la educación superior que tiene el estado y que hoy en día son de carácter obligatorio, Saber Pro 2010, con el propósito de observar qué factores se podían relacionar con los puntajes obtenidos tanto por los estudiantes, como por las Instituciones de Educación Superior que ofrecen este programa.

Para la elaboración del estudio, inicialmente se identificaron las variables cualitativas que podían relacionarse con los puntajes de los estudiantes y las instituciones, procediendo seguidamente a realizar un análisis de tablas de contingencia y pruebas de independencia. Con los resultados obtenidos en este último examen, se procedió a realizar el análisis de correspondencia simple y múltiple con la variable puntaje y las variables relacionadas con ella.

Los resultados de este estudio permitieron asociar los puntajes de los estudiantes con el género, el carácter público o privado de la institución, la educación académica de los padres y el estrato social al que pertenecen. Los puntajes de las instituciones se asociaron con la acreditación del programa y con la ubicación geográfica de la IES.

* Trabajo de grado.

** Facultad de Ciencias. Especialización en Estadística. Mg. Tulia Esther Rivera Flórez

1. TITLE:
FACTORS ASSOCIATED WITH PERFORMANCE TEST SABER PRO 2010 OF CIVIL
ENGINEERING STUDENTS *

2. AUTHORS:
Adrián David Mojica Perdomo
Wolfgang Alexander Osma Castellanos **

3. KEY WORDS:
Test Saber Pro
Civil Engineering
Score
Crosstabs
Correspondence analysis

4. DESCRIPTION:

In this study, performed statistical analyzes about the results obtained by students in their final year of Civil Engineering at the national level on academic tests to monitor the higher education, Saber Pro 2010, in order to observe which factors could be correlated with the scores obtained by the students and Higher Education Institutions offering this program.

In preparing the study, initially identified the qualitative variables that could relate to the scores of students and institutions, then proceeding to an analysis of contingency tables and tests of independence. With the results obtained in this last test, we proceeded to perform simple and multiple correspondence analysis with the variable score and the variables related with it.

The results of this study allowed to associate scores of students with gender, public or private institution, the academic education of parents and their social strata. The scores of institutions were associated with the accreditation of the program and with geographic location of the IES.

* Grade work.

** Facultad de Ciencias. Especialización en Estadística. Mg. Tulia Esther Rivera Flórez

INTRODUCCIÓN

Los exámenes de Calidad de la Educación Superior (ECAES), hoy conocidos como Prueba Saber Pro, se empezaron a aplicar en Colombia desde 2003 dando cumplimiento a las funciones de inspección y vigilancia de la educación que tiene el Estado y consignadas en la Constitución política de 1991 y Ley 30 de 1992. Estas pruebas son definidas como “pruebas académicas de carácter oficial y obligatorio que forman parte, con otros procesos y acciones, de un conjunto de instrumentos que el Gobierno Nacional dispone para evaluar la calidad del servicio educativo” (Artículo 1, decreto 1781).

Estas pruebas académicas son presentadas por los estudiantes de último año de los programas de pregrado de la Educación Superior y tienen por objeto comprobar los niveles mínimos de aptitudes y conocimientos de los estudiantes tal como se indica en el decreto 1781, en sus artículos 2 y 3 donde también se señala que será el ICFES la entidad encargada de diseñar, aplicar y administrar los resultados de las pruebas. En sus primeras versiones, es decir en el período del 2004-2008, la prueba fue de carácter opcional, a partir del 2009 es obligatoria y es requisito de grado.

En estos últimos años, los resultados de las pruebas han tenido algún impacto en el sector académico, en algunos casos se han utilizado como criterio de selección en algunas convocatorias de becas o selección de personal para trabajar en empresas nacionales y multinacionales

(Programa *Crédito mejores pruebas saber pro posgrado país*¹, Programa *Semillero de talentos regionales de Colfuturo*², Programa *Lo que pasa con los mejores ECAES*³).

Desafortunadamente, contrario a los objetivos planteados al implantar las pruebas, los resultados no se han convertido en una fuente de información importante para las Instituciones de Educación Superior, o al menos no al grado de convertirse en fuente primaria para autoevaluarse, formular políticas y reformas académicas acorde a los resultados observados⁴.

En este sentido se ha observado que solo las instituciones educativas que logran ubicar estudiantes en los primeros puestos del examen a nivel nacional, hacen despliegue publicitario en sus páginas web y medios de comunicación. De hecho, son pocos los estudios donde se analice el desempeño de los estudiantes de un programa académico y de reformas académicas producto de éstos. Tal vez esta situación ha sido uno de los motivos por el cual el ICFES ha colocado a disposición de investigadores las bases de datos de las pruebas que administra incluyendo la de las pruebas Saber Pro.

Motivados en esta situación se realiza el presente trabajo con el objetivo de conocer la incidencia de los factores personales, socio-familiares, escolares o académicos, contextuales y demográficos en los resultados de las pruebas Saber-Pro, convencidos de que, el conocimiento de los factores que inciden en los resultados de las pruebas Saber Pro

¹ <http://www.icetex.gov.co/portal/Default.aspx?tabid=1064>

² <http://www.colfuturo.org/index.php?page=33&site=1&idFile=6871&id=6871&owner=0>

³ <http://www.dinero.com/administracion/articulo/lo-pasa-mejores/110598>

⁴ <http://www.universidad.edu.co/index.php?option>

permitirán a las Instituciones de Educación Superior (IES) construir planes y programas tendientes a mejorar los desempeños de sus estudiantes, y a adecuar sus procesos en procura de la excelencia académica.

Para la realización de este trabajo se analizaron los resultados en la prueba Saber Pro obtenidos por los estudiantes de Ingeniería Civil en el año 2010 a nivel nacional que fueron tomados de la página web del ICFES. Para realizar los análisis estadísticos se utilizaron los siguientes programas: Statistical Package for the Social Sciences® (SPSS 18.0), Process Analysis Design And Enactment® (SPAD 5.6), Microsoft® Excel 2010, Microsoft® Access 2010, Microsoft® Word 2010.

Se realizó el análisis mencionado desde dos puntos de vista diferentes y complementarios: 1. Los resultados individuales de los estudiantes y 2. Los resultados de las IES que ofrecen el programa de Ingeniería Civil. Para los estudiantes se asumió como variable básica los resultados obtenidos individualmente en la prueba, en tanto que para las IES se adoptó el promedio obtenido por sus estudiantes en la misma prueba y período de tiempo.

Los análisis realizados permitieron observar que algunas de las variables que se asocian con el puntaje obtenido por los estudiantes son: el género del estudiante, el carácter privado o público de la institución, la educación alcanzada por el padre y por la madre y el estrato social en el que vive. De otro lado, las variables que se asocian con los resultados de las instituciones son: sí el programa está acreditado y si la institución se encuentra en una de las siete ciudades principales del país.

El trabajo está organizado en cinco capítulos cuyos contenidos se describen brevemente así:

En el primer capítulo, *GENERALIDADES DE LA PRUEBA SABER PRO*, se hace una descripción de algunos aspectos relevantes de la prueba, los factores que evalúa, las variables que tiene en cuenta así como los datos disponibles en la página web del ICFES: www.icfes.gov.co

El segundo capítulo, *ANTECEDENTES*, contiene una revisión de algunos estudios a nivel nacional donde se analizan el rendimiento académico de estudiantes universitarios.

En el capítulo tres, *MARCO TEÓRICO*, se realiza una síntesis de los aspectos teóricos más importantes de las técnicas estadísticas utilizadas en el presente trabajo para el análisis de los datos obtenidos.

El capítulo cuatro, *ANALISIS DE CORRESPONDENCIAS*, se organiza en dos secciones: en la primera se muestran los resultados del estudio de las tablas de contingencia, análisis de correspondencia simple y múltiple realizados para los estudiantes, en la segunda sección, los mismos análisis realizados para las instituciones.

En el capítulo cinco, *CONCLUSIONES*, se precisan los resultados más importantes obtenidos en este estudio.

1. GENERALIDADES DE LA PRUEBA SABER PRO

La prueba Saber Pro es un examen a nivel nacional que pretende medir la calidad de la educación superior en Colombia. La prueba es un instrumento estandarizado administrado por el ICFES y forma parte, con otros procesos y acciones, de un conjunto de instrumentos que el Gobierno Nacional dispone para evaluar la calidad del servicio educativo y ejercer su inspección y vigilancia.

Como se mencionó en la introducción, según el ECAES, uno de los objetivos al implantar las pruebas Saber-Pro era que sirvieran de fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e Instituciones de Educación Superior.

Específicamente los objetivos de la prueba Saber-Pro son, según los *Lineamientos pruebas SABER PRO Noviembre 2011*⁵:

- Comprobar el desarrollo de competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las Instituciones de Educación Superior (IES).
- Producir indicadores de valor agregado de la educación superior en relación con el nivel de competencias de quienes ingresan a ella, proporcionar

⁵ http://www.icfes.gov.co/index.php?option=com_docman&task=doc_view&gid=4384&Itemid=59

información para la comparación entre programas, instituciones y metodologías, y para mostrar cambios 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 así como del servicio público educativo. Se espera que estos indicadores fomenten la cualificación de los procesos institucionales, la formulación de políticas y soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

La prueba es presentada por estudiantes que están próximos a graduarse, es decir, que hayan aprobado por lo menos el 75% de los créditos académicos del programa que cursan. La organización de la prueba es por módulos que evalúan las competencias consideradas fundamentales para los futuros egresados de los programas de pregrado ofrecidos en Colombia. Diferentes comunidades académicas, asociaciones de facultades y de programas han participado en el diseño y evaluación de los cuestionarios.

Las pruebas están organizadas por módulos, algunos evalúan competencias genéricas, es decir, aquellas competencias que deben desarrollar todos los estudiantes de educación superior independientemente del programa o nivel en el que se encuentren;

otros módulos evalúan competencias específicas comunes a grupos de programas de la misma o de distintas áreas de formación.

Las competencias cognitivas sobre las cuales se fundamenta el modelo de evaluación del ICFES, según la descripción dada por la Asociación Colombiana de Facultades de Ingeniería en 2005 en *Orientaciones para el examen de estado de calidad de la educación superior SABER PRO - ECAES*⁶, se definen así:

COMPETENCIA INTERPRETATIVA: Se define como aquella acción encaminada a encontrar el sentido de un texto, un problema, una gráfica, un plano de ingeniería, un diagrama de flujo, una ecuación, un circuito eléctrico, entre otras situaciones, donde se le proporciona un contexto al estudiante. La interpretación sigue unos criterios de veracidad, los cuales no implican sólo la comprensión de los contextos, sino que se debe dirigir a la situación concreta y reflexionar sobre sus implicaciones y los procesos de pensamiento involucrados son el recuerdo, la evocación, comprensión, análisis, medición, etc.

COMPETENCIA ARGUMENTATIVA: Es aquella acción dirigida a explicar, dar razones y desarrollar ideas de una forma coherente con el contexto de la disciplina evaluada. Los puntos relacionados con esta competencia exigen dar cuenta de un saber fundamentado en razones coherentes con los planteamientos que se encuentran en el texto. Se contextualiza la argumentación en acciones como la resolución

⁶ http://www.icfes.gov.co/index.php?option=com_docman&task=doc_view&gid=3456

de problemas, los fundamentos de un diseño de ingeniería, la organización de la información, la proyección de la información, la explicación de eventos, fenómenos, la formulación de soluciones a través de un gráfico, un plano, un diagrama, etc.

COMPETENCIA PROPÓSITIVA: Es aquella acción cuyo fin persigue que el estudiante proponga alternativas que puedan aplicarse en un contexto determinado; por lo tanto, se espera que la solución que escoja corresponda con las circunstancias que aparecen en la formulación de un problema. Así mismo, el estudiante deberá generar hipótesis y proponer alternativas de solución a los problemas de ingeniería que cubran aspectos como los ambientales, de manufacturabilidad, económicos, entre otros; y propondrá acciones de aplicación, evaluación o/y optimización de una solución en un contexto de ingeniería dado.

1.1 EVALUACIÓN POR COMPETENCIAS Y PRUEBAS SABER PRO PARA ESTUDIANTES DE INGENIERÍA CIVIL

Basados en el modelo de evaluación por competencias antes mencionadas, el ICFES estructuró los elementos de evaluación de la prueba Saber-Pro para cada programa académico. Desde la perspectiva de lo genérico y lo específico, se asumen unas características comunes para todas las ingenierías y unas específicas para cada programa a

evaluar, y que para efectos de la evaluación, se denominan Componentes, los cuales se pueden observar seguidamente en el Cuadro 1.

Cuadro 1. Modelo de evaluación del ICFES para estudiantes de Ingeniería Civil.

DEFINICIÓN DE ELEMENTOS DE LA ESTRUCTURA DE PRUEBA ECAES PARA INGENIERÍA CIVIL			
COMPONENTES DE LA PRUEBA	COMPETENCIA COGNITIVA INTERPRETACIÓN	COMPETENCIA COGNITIVA ARGUMENTACIÓN	COMPETENCIA COGNITIVA PROPOSICIÓN
Modelamiento de fenómenos y procesos	Identifica los aspectos y características relevantes de un fenómeno o proceso	Establece y analiza relaciones que representan fenómenos o procesos y modela fenómenos y procesos	Plantea hipótesis y genera alternativas de modelos que representan un fenómeno o proceso
Resolución de problemas, mediante la aplicación de las ciencias naturales y las matemáticas utilizando un lenguaje lógico y simbólico	Identifica y comprende las variables que definen un problema	Selecciona métodos apropiados y resuelve un problema	Plantea hipótesis y genera alternativas de solución de un problema
Comunicación efectiva y eficazmente en forma escrita, gráfica y simbólica	Lee, comprende e interpreta textos científicos, gráficas, datos e información experimental, planos e imágenes de sistemas mecánicos	Argumenta ideas técnicas a través de textos, gráficas, reportes de datos experimentales, planos e imágenes.	Propone ideas técnicas a través de textos, gráficas, reportes de datos experimentales, planos e imágenes.
Diseño de sistemas, componentes o procesos que cumplan con especificaciones deseadas	Definir especificaciones apropiadas para diseños de Ingeniería	Analiza, establece y relaciona elementos de las especificaciones	Plantea posibles nuevos diseños a partir de la aplicación de principios científicos y conocimientos tecnológicos
Planeación, diseño, evaluación del impacto (social, económico, tecnológico y ambiental) y gestión proyectos de ingeniería civil	Identifica aspectos relevantes de un proyecto y traslada su definición a términos de ingeniería.	Analiza y establece las mejores prácticas aplicables en un proyecto y dimensiona sus consecuencias de tipo social y ambiental y formula proyectos	Propone nuevas formas de gestionar proyectos de ingeniería
Planear, diseñar, evaluar y definir metodologías para la adquisición de datos y el diseño experimentos.	Identifica la información necesaria para el diseño o la modelación de un problema en ingeniería	Interpreta y analiza eficientemente información experimental o de campo dentro el contexto del problema que está resolviendo.	Propone metodologías para la adquisición de datos y diseña experimentos como parte de la solución de problemas.

1.2 DESCRIPCIÓN DE LA BASE DE DATOS

Las bases de datos están condensadas en archivos Microsoft Office Access en donde se puede encontrar información para todas las pruebas Saber Pro presentadas desde el año 2004 hasta el año 2010. La prueba

ha sufrido cambios por cuanto no se tienen las mismas variables medidas en todo el período de tiempo en algunos casos, se tiene acceso a puntajes de los exámenes genéricos y a los específicos por competencias; pero en otros, sólo se tiene el puntaje total tanto para las pruebas genéricas como específicas.

El Cuadro 2 muestra la cantidad de estudiantes de Ingeniería Civil que han presentado la prueba en este período de tiempo a nivel nacional. Cabe advertir que a partir de la expedición de la Ley 1324 de 2009 la prueba es de carácter obligatorio para los estudiantes.

Cuadro 2. Número de estudiantes que han presentado la prueba saber Pro en el período 2004-2010.

AÑO	TOTAL ESTUDIANTES
2004	2658
2005	2172
2006	2003
2007	1695
2008	1615
2009	2953
2010	2268
Total	15364

La base de datos cuenta con cerca de 63 variables clasificadas en los siguientes grupos:

- *Información de la prueba:* Se identifican aspectos de la prueba que presenta el estudiante. Por ejemplo nombre de la prueba y área del conocimiento que se evalúa.

- *Información general del estudiante:* Diferentes datos familiares, información socioeconómica, datos personales y/o de identificación del estudiante. Por ejemplo, código de identificación de la prueba, si sabe leer o hablar otro idioma, etc.
- *Información de discapacidad del estudiante:* Variables acerca de la discapacidad del estudiante (discapacidad motriz, auditiva o invidencia).
- *Información sobre la institución educativa:* Información sobre la institución educativa del estudiante. Por ejemplo ciudad sede de la Institución, nombre de la institución, carácter académico (universidad, técnica profesional, normalista, etc.), si es de carácter privado u oficial, etc.
- *Resultados de desempeño académico:* Para el caso de exámenes específicos se presenta el puntaje individual del evaluado, así como el puntaje obtenido en cada una de las competencias evaluadas.

A continuación en el Cuadro 3 se presentan las modificaciones de las componentes que han sido evaluadas en las versiones de esta prueba.

Cuadro 3. Componentes evaluadas en las versiones de las Pruebas Saber Pro 2004-2010

COMPETENCIAS	2004-2008	2009	2010
Genéricas	Comprensión lectora común	Pensamiento crítico	Pensamiento crítico
		Entendimiento interpersonal	Entendimiento interpersonal
			Comprensión comunicativa escrita
		Comprensión, escritura, argumentación	Comprensión lectora puntaje
			Comprensión lectora desempeño
	Inglés común	Inglés puntaje	Inglés puntaje
		Inglés desempeño	Inglés desempeño
Específicas	Resolución de Problemas en ingeniería	Solución de problemas	Solución de problemas
	Diseño de Sistemas Componentes procesos		
	Modelamiento de fenómenos		
	Planeación, diseño, evaluación definición		
	Planeación, diseño, evaluación del impacto		

Para el año 2010 las pruebas de carácter genérico se han separado en aspectos más específicos y por el contrario, las pruebas de carácter específico se han ido agrupando en varios otros tópicos.

Las pruebas de competencias genéricas han sido estructuradas con un número de preguntas y tiempo disponible para su respuesta como se muestra en el Cuadro 4.

Cuadro 4. Estructura de la prueba Saber Pro.

PRUEBAS GENÉRICAS	NÚMERO DE PREGUNTAS	TIEMPO ASIGNADO
Comunicación escrita	Un escrito	40 minutos
Solución de problemas	93 preguntas	3 horas 50 minutos
Pensamiento Crítico		
Entendimiento Interpersonal		
Comprensión lectora	15 preguntas	
Inglés	40 preguntas	
Tiempo total de la prueba		4 horas 30 minutos

Fuente: ICFES. Orientaciones para el examen de estado de calidad de la educación superior Saber Pro (ECAES). Guías.

La forma de evaluar la prueba de comunicación escrita es asignando al estudiante un tema para que elabore un escrito argumentativo. Para valorar las siguientes tres pruebas de la tabla: solución de problemas, pensamiento crítico y entendimiento interpersonal se proponen 93 preguntas de selección múltiple.

La prueba de comprensión lectora es abordada con dos textos de interés general, uno de corte periodístico y el otro de carácter expositivo. Sobre los cuales se plantean 15 preguntas de selección múltiple con única respuesta.

La prueba de inglés está organizada en 6 partes y demanda la respuesta de 40 preguntas basadas en la observación de imágenes, listas, conversaciones cortas y textos.

2. ANTECEDENTES

En la literatura relacionada con el tema, pruebas SABER PRO, se encuentran pocas referencias de trabajos sobre el desempeño académico de los jóvenes universitarios colombianos. Sin embargo, en la página web del ICFES se pueden consultar los resultados de los estudios estadísticos a nivel nacional que realiza este organismo con el ánimo de “brindar una fuente de información que permita a las instituciones de Educación Superior fomentar la cualificación de los procesos institucionales facilitando el proceso de toma de decisiones en el sistema educativo”.

En estudios sobre el desempeño de los estudiantes universitarios se muestra que existen factores o categorías de índole personal, social, económica e institucional que inciden en su rendimiento académico.

A continuación se construye un resumen con los principales hallazgos de dos estudios relacionados directamente con las pruebas Saber Pro.

□ *LEIDY JOHANA CÁRDENAS ACEVEDO, YERLY KATHERINE RINCÓN DELGADO*

Cárdenas & Rincón, (2011) en su tesis de pregrado “DETERMINANTES DE LA CALIDAD DE LA EDUCACIÓN PÚBLICA UNIVERSITARIA EN

COLOMBIA” pretenden participar en el proceso de búsqueda de los determinantes de la calidad de la educación superior de carácter público en el país estimando un modelo econométrico con los datos suministrados por el Sistema Nacional de Información de Educación Superior (SNIES) para el año 2009.

Para tal fin, las autoras utilizan un método analítico que comprende dos niveles: El primer nivel realiza un análisis general de las características del sistema educativo colombiano y el segundo nivel establece mediante un *modelo de regresión probit*⁷ la relación entre la cualificación de los docentes y los programas acreditados de Alta calidad con el promedio del puntaje del ECAES de las 32 universidades públicas que componen el Sistema Universitario estatal (SUE).

Como variables para la estimación del modelo tuvieron en cuenta:

- *Variable dependiente:*

PROMEDIO DE ECAES: Que categorizaron en dos niveles: un primer nivel las universidades cuyo promedio se encuentra por encima del puntaje promedio de ECAES de las 32 universidades públicas y un segundo nivel con las que se encuentran por debajo de esa media.

- *Variables explicativas:*

⁷ Modelo de regresión no lineal pensado para realizar regresiones con variables categóricas binarias (que solo pueden adquirir dos posibles valores)

Consideraron las siguientes variables: infraestructura física por estudiante, docentes TCE⁸ con estudios de especialización, docentes TCE con estudios de maestría, docentes TCE con estudios de Doctorado, aportes de la nación por estudiante matriculado, relación de programas acreditados de alta calidad sobre el total de programas de pregrado por universidad, grupos de investigación en las categorías A1 y A reconocidos por Colciencias, número de revistas indexadas de la institución, número de artículos publicados en revistas indexadas según índice en Colciencias, participación de estudiantes vinculados a la extensión del total de estudiantes matriculados.

Las autoras utilizaron un modelo de regresión probit al haber reducido la variable dependiente a sólo dos alternativas de elección mutuamente excluyentes.

Los resultados que obtuvieron indican que el promedio del puntaje en la prueba ECAES en las universidades públicas es explicado significativamente por tres de las variables consideradas: la participación de docentes con estudios de maestría, los docentes con estudios de doctorado y la relación entre los programas académicos acreditados y los no acreditados, presentando niveles de significación inferiores al 10%.

La bondad de ajuste del modelo con un valor de $R^2=0,6053$ lo consideraron adecuado ya que explicaba el 60,53% de la variabilidad de la variable dependiente.

⁸ Tiempo Completo Equivalente = (docentes tiempo completo)+(docentes medio tiempo/2)+ (suma horas cátedra docente/18*40)

□ *EXÁMENES DE ESTADO DE CALIDAD DE LA EDUCACIÓN
SUPERIOR SABER PRO. RESULTADOS DEL PERÍODO 2004-2009*

En el informe de resultados del periodo 2005-2009 que el ICFES entrega sobre los exámenes de estado de calidad de la educación superior Saber Pro, se resalta que la prueba 2009 fue la primera llevada a cabo después de entrar en vigencia la nueva reglamentación sobre la obligatoriedad como requisito de grado y señalan que las dos mayores implicaciones que trajo la nueva ley están en el aumento en el número de estudiantes evaluados y en la aplicación, por primera vez, de pruebas comunes de competencias genéricas: solución de problemas, pensamiento crítico, entendimiento interpersonal y comunicación escrita.

Con la lectura de este informe se busca compactar la información correspondiente al programa de Ingeniería Civil a nivel nacional que permita más adelante hacer comparaciones con los resultados obtenidos en el presente trabajo.

En el 2009 el crecimiento del número de estudiantes de pregrado evaluados en comparación con el 2008 fue del 93% (77,2 mil) a través de las pruebas de competencias genéricas. Para el programa de Ingeniería Civil ese aumento estuvo por debajo de la media general con un 83% representado por 1.337 estudiantes.

El programa de Ingeniería Civil está en el 14° puesto entre los programas que han sido evaluados con más de 10.000 estudiantes en

estas pruebas ⁹, siendo el programa de administración el que cuenta con el mayor número de estudiantes (81.486).

Del total de estudiantes evaluados, los de instituciones oficiales representaron algo menos de la mitad (46%) y por el contrario en Ingeniería civil representaron algo más de la mitad (56,7%). El 43% de los estudiantes valorados son del género masculino, en su programa, los Ingenieros civiles tienen una participación del 70,9 posicionándose en el 9º lugar entre los programas y 4º entre las Ingenierías con mayor participación de este género.

En los Cuadros 5 y 6 se extractan los resultados del programa de Ingeniería Civil del informe generado por el ICFES.

Cuadro 5. Evolución de los puntajes promedio del programa de Ingeniería Civil.

Área	Programa evaluado		2004	2005	2006	2007	2008	2009
<i>Ingeniería, Arquitectura, Urbanismo y afines</i>	<i>Ingeniería Civil</i>	<i>Nivel nacional</i>	-----	103,4	102,6	100,0	100,6	99,99
		<i>IES Oficiales</i>	-----	105,2	103,9	101,5	101,6	100,3
		<i>IES Privadas</i>	-----	95,7	98,2	97,4	96,6	93,0

Fuente: ICFES. Bases de datos de Saber Pro. Cálculos: Dirección de evaluación del ICFES.

“En Ingeniería Civil, los puntajes oficiales superaron ampliamente los de los privados en 6 puntos en el período; sin

⁹ Para el período 2004-2009 el total de estudiantes evaluados en el programa de Ingeniería Civil es de 13.095

embargo, ambos grupos mostraron una clara tendencia negativa en sus resultados y bajaron cerca de 5 y de 4 puntos en el período los oficiales y los privados, respectivamente” (p.52)

Se presentan seguidamente los resultados en los puntajes tanto para hombres, como para mujeres y su respectiva diferencia para el año 2009.

Cuadro 6. Puntajes promedio de hombres y mujeres evaluados en el 2009.

Exámenes	Mujeres	Hombres	Diferencia Hombres-Mujeres
<i>Ingeniería Civil</i>	<i>95,9</i>	<i>101,3</i>	<i>5,4</i>

Fuente: ICFES. Bases de datos de Saber Pro. Cálculos: Dirección de evaluación del ICFES.

“La diferencia estadísticamente significativa favorece a los hombres” (p.60).

En la síntesis final, el informe concluye con respecto a los puntajes en el programa de Ingeniería Civil:

- ✓ La evolución de los puntajes promedio en el período 2004-2009 se encuentra entre los catalogados “con tendencia negativa moderada”.
- ✓ Respecto a la evolución de los puntajes por sector los estudiantes

oficiales superaron a los privados.

- ✓ Por género, los hombres superaron en los puntajes a las mujeres.
- ✓ Se encontró una correlación positiva (0,091) y estadísticamente significativa ($\text{valor_p} = 0,000$) entre el nivel socioeconómico y puntajes.

3. MARCO TEÓRICO

En este capítulo se expone el marco teórico concerniente a las técnicas estadísticas utilizadas en el desarrollo de este trabajo. En el primer apartado se describen algunos aspectos de la teoría básica del Análisis de Tablas de Contingencia y de las pruebas de independencia entre las variables cualitativas, posteriormente, se relacionan algunos conceptos básicos del Análisis de Correspondencias.

El planteamiento teórico en mención se realiza tomando como guía los textos de Peña (2002), Díaz (2007) y Greenacre (2008) (Greenacre, 2008).

3.1 TABLAS DE CONTINGENCIA (TC)

Una tabla de contingencia es una tabla en la cual se relaciona información de dos variables categóricas que presentan diferentes modalidades o niveles, es decir, la tabla tiene doble entrada, en cada una de las casillas aparece el número de casos o individuos que poseen un nivel de una de las variables o características analizadas y otro nivel, de la otra variable con la que se quiere observar la relación.

En general, una tabla de contingencia es un arreglo bidimensional de f filas por c -columnas las cuales conforman $f \times c$ celdas informativas.

A continuación en la Tabla 1 se expone una representación de este tipo de tablas.

Tabla 1. Tabla de contingencia general f -filas * c -columnas

Variable Fila	Variable Columna						Total fila ($n_{i.}$)
	1	2	...	j	...	c	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
I	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
F	n_{f1}	n_{f2}	...	n_{fj}	...	n_{fc}	$n_{f.}$
Total columna ($n_{.j}$)	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.c}$	$n_{..} = N$

Dónde:

n_{ij} = Es la frecuencia de la i -ésima modalidad de la variable fila y j -ésima modalidad de la variable columna.

$$n_{i.} = \sum_{j=1}^c n_{ij} \quad n_{.j} = \sum_{i=1}^f n_{ij}$$

Son los totales (marginales) de cada uno de los niveles de las variables fila y columna respectivamente.

Las frecuencias pueden ser transformadas en proporciones o porcentajes. La frecuencia relativa absoluta se obtiene al dividir cada frecuencia n_{ij} por el número total de observaciones N . Este porcentaje se escribe como f_{ij}

$$f_{ij} = \frac{n_{ij}}{N}$$

Luego la cantidad f_{ij} corresponde a la proporción o porcentaje de individuos que tienen las dos características determinadas por la fila y la columna que definen esa celda.

De igual manera la frecuencia relativa marginal se obtiene de la siguiente manera tanto para filas como para columnas

$$f_{i.} = \frac{n_{i.}}{N} \quad f_{.j} = \frac{n_{.j}}{N}$$

El i -ésimo perfil fila se obtiene al dividir cada frecuencia n_{ij} de la fila i por la respectiva frecuencia marginal fila $n_{i.}$, esto es:

$$f_{j|i} = \frac{n_{ij}}{n_{i.}}$$

El perfil columna se obtiene al dividir cada frecuencia n_{ij} por la respectiva frecuencia marginal columna $n_{.j}$

$$f_{i|j} = \frac{n_{ij}}{n_{.j}}$$

La Tabla 2 presenta la tabla de perfiles con sus respectivos perfiles marginales.

Tabla 2. Tabla de perfiles o frecuencias relativas condicionales.

Variable Fila	Variable Columna						Marginal fila ($f_{i.}$)
	1	2	...	j	...	c	
1	f_{11}	f_{12}	...	f_{1j}	...	f_{1c}	$f_{1.}$
2	f_{21}	f_{22}	...	f_{2j}	...	f_{2c}	$f_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	f_{i1}	f_{i2}	...	f_{ij}	...	f_{ic}	$f_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
F	f_{f1}	f_{f2}	...	f_{fj}	...	f_{fc}	$f_{f.}$
Marginal columna ($f_{.j}$)	$f_{.1}$	$f_{.2}$...	$f_{.j}$...	$f_{.c}$	$f_{..} = N$

3.2 DEPENDENCIA ENTRE VARIABLES

A partir de una tabla de contingencia se puede analizar si existe alguna relación de dependencia o independencia entre las modalidades de las

variables categóricas que son objeto de estudio. Es de anotar que el hecho de que dos variables sean dependientes significa que los valores de una de ellas están influidos por el nivel que adopte la otra variable.

Para identificar estas relaciones de dependencia entre par de variables se utiliza la *Prueba ji-cuadrado* cuyo cálculo permite afirmar con un nivel de confianza determinado si las modalidades de una variable influyen en las de la otra variable.

Dos variables son independientes si se cumple:

- a) Que la frecuencias relativas condicionadas sean iguales a las frecuencias relativas marginales

$$f_{j|i} = \frac{n_{ij}}{n_{i.}} = f_{i.} = \frac{n_{i.}}{N}$$

$$f_{i|j} = \frac{n_{ij}}{n_{.j}} = f_{.j} = \frac{n_{.j}}{N}$$

- b) O bien que la frecuencia relativa conjunta sea igual al producto de las frecuencias relativas marginales

$$f_j \cap f_i = \frac{n_{ij}}{N} = \frac{n_{i.}}{N} \times \frac{n_{.j}}{N}$$

Así, al comparar las frecuencias teóricas esperadas en caso de independencia entre las variables con las frecuencias observadas en los datos, se puede concluir qué tipo de relación existe. Por otro lado, utilizando el concepto de probabilidad se puede estimar la probabilidad

de un evento determinado a partir de sus frecuencias relativas de la siguiente manera:

$$P_{ij} = \frac{n_{ij}}{N} \quad P_{i.} = \frac{n_{i.}}{N} \quad P_{.j} = \frac{n_{.j}}{N}$$

Si las variables son independientes se tendría la expresión:

$$P_{ij} = \frac{E_{ij}}{N} = \frac{n_{i.}}{N} \times \frac{n_{.j}}{N}$$

Donde E_{ij} sería la frecuencia absoluta esperada o teórica si las variables son independientes, pudiéndose calcular con la expresión:

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{N}$$

De esta forma se obtienen tantos valores E_{ij} esperados y n_{ij} observados como celdas tiene la tabla confirmando que si hay poca diferencia entre estos valores se estará en condiciones de independencia y por el contrario, si las diferencias son muy grandes, se estará en condiciones de dependencia.

Como la variable fila tiene f filas, y la variable columna c columnas, la tabla es de orden $f \times c$. La prueba ji-cuadrado utiliza el estadístico χ^2 para analizar la independencia entre variables cualitativas y está definido por:

$$\hat{\chi}^2 = \frac{\sum_{i=1}^c \sum_{j=1}^f (n_{ij} - E_{ij})^2}{E_{ij}}$$

Siendo las hipótesis a contrastar con la prueba:

H_o : Las variables son independientes

H_a : Las variables no son independientes

El valor de $\hat{\chi}^2$ calculado se compara con el valor tabulado de una χ^2 para el nivel de confianza determinado (α) con $(f-1)(c-1)$ grados de libertad y se observa el valor_p (significancia) para concluir según la regla de decisión así:

$valor\ p < \alpha \rightarrow$ Se rechaza $H_o \rightarrow$ SON DEPENDIENTES

$valor\ p > \alpha \rightarrow$ Se acepta $H_o \rightarrow$ SON INDEPENDIENTES

3.3 ANÁLISIS DE CORRESPONDENCIAS

El análisis de correspondencia es una técnica estadística que describe las relaciones existentes entre las categorías de dos variables nominales organizadas en una tabla de correspondencias o contingencia y permite representar gráficamente la relación entre ellas. Con algunas reservas se puede afirmar que para cada variable en un gráfico, las distancias

entre los puntos de las categorías de una y otra, reflejan las relaciones entre ellas.

3.3.1 Análisis de Correspondencia Simple (ACS)

(Díaz Monroy & Morales Rivera, Análisis estadístico de datos categóricos, 2009)

Se dice que el ACS es una metodología estadística que sirve para analizar la homogeneidad entre las modalidades de cada una de las dos variables cualitativas dispuestas en una tabla de contingencia donde el interés se centra en representar de manera conjunta en un plano factorial las distintas modalidades de la tabla de contingencia, de forma que la distancia entre los puntos está relacionada con el nivel de asociación entre dichas modalidades.

Idealmente se desea explicar la mayor inercia posible con el menor número de factores, generalmente 2 o 3 dimensiones son retenidas.

Sea X la matriz de datos de $n \times p$ que relaciona las mediciones de 2 variables categóricas cada una de ellas con n y p modalidades o niveles para una población de N individuos.

$$X = \begin{pmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,p} \end{pmatrix}$$

Donde cada $a_{i,j}$ $i=1,.., n$ y $j=1,..., p$ representa el conteo de individuos que poseen la característica i de una de las variables y la característica j de la otra variable.

En la tabla de contingencia X se pueden contemplar dos espacios, el primero de ellos el espacio fila \mathbb{R}^p y el espacio de las columnas \mathbb{R}^n .

Con respecto a los datos y la tabla de contingencia se tiene que:

- El número total de individuos por fila se nota por:

$$n_{i.} = \sum_{j=1}^p n_{ij}, \text{ para } i = 1, \dots, n. \quad (3.1)$$

- El número total de individuos por columna se nota por:

$$n_{.j} = \sum_{i=1}^n n_{ij}, \text{ para } j = 1, \dots, p. \quad (3.2)$$

- El número total de individuos de la tabla está dado por:

$$N = \sum_{i=1}^n \sum_{j=1}^p n_{i,j} \quad (3.3)$$

- Las frecuencias relativas absolutas:

$$f_{i,j} = \frac{n_{i,j}}{N} \quad (3.4)$$

- Las frecuencias marginales de filas y columnas:

$$f_{i.} = \sum_{j=1}^p f_{i,j} = \frac{n_{i.}}{N} \quad (3.5.a) \quad f_{.j} = \sum_{i=1}^n f_{i,j} = \frac{n_{.j}}{N} \quad (3.5.b)$$

El perfil fila (columna) de una categoría de la variable fila (columna) es el vector de las frecuencias absolutas de la variable divididas por la frecuencia marginal de la fila (columnas) que tiene:

$$\square \text{ Perfil fila: } \left(\frac{n_{i1}}{n_{i.}}, \dots, \frac{n_{ip}}{n_{i.}} \right) \quad i = 1, 2, \dots, n \quad (3.6)$$

$$\square \text{ Perfil columna: } \left(\frac{n_{1j}}{n_{.j}}, \dots, \frac{n_{nj}}{n_{.j}} \right) \quad j = 1, 2, \dots, p \quad (3.7)$$

El centroide de los perfiles fila está dado por las frecuencias marginales es decir $\mathcal{G}_f = (f_{1.}, \dots, f_{p.})$; y el centroide de los perfiles columnas se representa por $\mathcal{G}_c = (f_{1.}, \dots, f_{n.})$. Y es la media ponderada de los perfiles fila con pesos para cada fila iguales a su masa (son las frecuencias marginales de la fila). Igualmente el baricentro o centroide de los perfiles columna son las frecuencias marginales de las columnas y se obtiene como la media ponderada de los perfiles columna con pesos iguales a las frecuencias marginales de la columna (masa de los perfiles columna).

La distancia χ^2 es utilizada para la medir la distancia (similitud) entre los perfiles fila o perfiles columnas es decir la distancia, "dos a dos" entre filas o columnas pertenecientes a la tabla de contingencia. Para dos perfiles fila i y i' dicha distancia está definida de la siguiente forma:

$$d^2(i, i') = \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \quad (3.8)$$

De igual forma para dos perfiles columna j y j' se tiene:

$$d^2(j, j') = \sum_{i=1}^n \frac{1}{f_{i.}} \left(\frac{f_{i,j}}{f_{.j}} - \frac{f_{i,j'}}{f_{.j'}} \right)^2 \quad (3.9)$$

Con respecto a las distancia χ^2 para perfiles fila y columna se tiene que esta distancia da mayor prioridad a las modalidades “raras”, dado a que por no estar presentes en la mayoría de los individuos son las más diferenciadoras.

La inercia (total) de una tabla de contingencia es igual al estadístico χ^2 dividido por el total N de la tabla. “Geométricamente, la inercia mide lo lejos que se hallan los perfiles fila (o los perfiles columna) de su perfil medio (baricentro). Se puede considerar que el perfil medio simboliza la hipótesis de homogeneidad (es decir, la igualdad de los perfiles)” (Greenacre, 2008, p.53). La inercia es una medida de la variabilidad o de la falta de homogeneidad existente entre los perfiles fila o columna.

La función del Análisis de Correspondencias es explicar la asociación entre filas y columnas a través de un menor número de dimensiones en la cual la primera dimensión (eje o factor) explique la mayor parte de asociación, y que la segunda dimensión explique la mayor parte de asociación no explicada por el primer eje, y así sucesivamente.

El máximo de número de dimensiones existentes es igual al menor número de modalidades o categorías presente en cualquiera de las variables menos uno. En la práctica es común tomar dos o tres dimensiones pues con esta cantidad de dimensiones se pueden obtener una representación aceptable de la asociación entre filas y columnas.

La cantidad $\frac{\chi^2}{N}$ se denomina inercia total, la cual se descompone en k valores característicos, donde cada uno de ellos representa la inercia principal de cada dimensión. Cada una de las dimensiones caracteriza las categorías o modalidades de las filas y columnas de la tabla de contingencia bidimensional situándolas como coordenadas en el espacio geométrico, siendo las puntuaciones fila las coordenadas para cada modalidad de la variable fila en las dimensiones de la tabla de contingencia y las puntuaciones columna son las coordenadas para cada modalidad de la variable columna en esas mismas dimensiones.

Díaz, Morales (2009, p. 110) con respecto a las coordenadas de las modalidades de las variables agregan:

Para calcular las coordenadas de las modalidades de las variables se procede de la siguiente forma¹⁰:

- La primera dimensión o eje principal produce un valor singular (valor propio) λ_1 que explica la mayor parte de la inercia (Asociación) total y constituye el primer conjunto de coordenadas: uno para la variable fila y otro para la variable columna.
- La segunda dimensión produce un nuevo valor singular λ_2 que explica la mayor parte de la inercia residual y constituye el segundo conjunto de coordenadas: una para la variable fila y otra para la variable de columna.

¹⁰ Una explicación más detallada puede ser consultada en (Díaz, 2007, cap 11).

- El resto de dimensiones se obtiene con el mismo procedimiento.

Por otro lado el procedimiento para el análisis de correspondencias simple se puede resumir de la siguiente forma:

- a) Se parte de los datos originales; las filas y columnas cumplen papeles simétricos. Estas son las modalidades de las dos variables, respectivamente. La suma de todos los términos de la tabla es N , el cual es el número total de individuos o efectivos.
- b) Se construye una tabla de las frecuencias relativas las cuales conforman las probabilidades. Donde $(f_{i.}: i = 1, \dots, n)$ y $((f_{.j}: j = 1, \dots, p)$ son las probabilidades marginales o perfiles fila o columna, respectivamente.
- c) Para estudiar las filas de la tabla, se transforman en perfiles fila. Análogamente se procede con las columnas. Se dispone entonces de dos tablas; una para el perfil fila y otra para el perfil columna. Un perfil se interpreta como una probabilidad condicional.
- d) Un perfil fila es un arreglo de p -números y está representado por un punto de \mathbb{R}^p . La nube de puntos \mathcal{H}_c . de los perfiles fila, está en un hiperplano \mathcal{H}_f de vectores tales que la suma de sus componentes es igual a 1. Cada perfil fila i es afectado por los puntos $f_{i.}$; de manera que la nube \mathcal{H}_f está “equilibrada” en los

perfiles medios o baricentro C_i . En la nube \mathcal{H}_f se busca la semejanza entre los perfiles, semejanza medida a través de una distancia χ^2 .

- e) La representación de los perfiles columnas de \mathbb{R}^n se hace de forma análoga a la representación de los perfiles fila en \mathbb{R}^p .
- f) El análisis factorial de la nube consiste en poner en evidencia una sucesión de direcciones ortogonales, tales que la inercia, con relación al origen O de la proyección de la nube de puntos sobre tales direcciones, sea máxima.
- g) Simétricamente se desarrolla un procedimiento análogo para las columnas.
- h) Los planos factoriales, determinados por dos factores sobre las filas o sobre las columnas, proporcionan imágenes aproximadas de las nubes \mathcal{H}_c y \mathcal{H}_f ; sobre este plano, la distancia entre los puntos se interpreta como la semejanza entre los perfiles de esos puntos. El origen de los ejes se considera como el perfil promedio.

3.3.2. Análisis de Correspondencias Múltiple (ACM)

El Análisis de Correspondencia Múltiple (ACM) es un análisis de correspondencia simple aplicado no solo a una tabla de contingencia, sino a una tabla disyuntiva completa (Ver Tabla 3).

Partiendo de una matriz de datos en la que cada individuo tiene asociada más de dos variables de tipo categórico (nominal u ordinal), donde las respuestas son mutuamente excluyentes y una categoría de cada respuesta es elegida de manera obligatoria, se construye una tabla de doble entrada mediante el uso de variables indicadoras por cada una de las categorías dentro de cada una de las variables o preguntas (Tabla 3), en la cual:

m_k : Modalidades de una pregunta o Variable.

V_j : Preguntas o variables.

Q : número de variables.

n : Número de individuos u observaciones.

Tabla 3. Tabla disyuntiva completa

<i>Individuos</i>	V_1 1 ... m_k	V_2 1 ... m_k	V_k 1 ... m_k	V_j 1 ... m_k	<i>Suma</i>
<i>1</i>						Q
<i>2</i>	Q
<i>i</i>				X_{i1}	X_{ik}	Q
...	
<i>N</i>						Q
<i>Suma</i>						$n \times Q$

Esta tabla es la que se conoce como tabla disyuntiva completa. Para facilitar la comprensión de este tipo de tabla a continuación se presenta un ejemplo expuesto en Díaz (2007 p.406).

Sea un conjunto de datos compuesto por 5 variables medidas para un conjunto de n individuos con las siguientes descripciones:

- *Grupo de Edad*: Modalidades → joven (1), adulto (2), anciano (3).
- *Género*: Modalidades → masculino (1), femenino (2).
- *Nivel de estudios*: Modalidades → Primaria (1), secundaria (2), Universitaria (3), Otra (4).
- *Categoría Socioeconómica*: Modalidades → Bajo (1), medio (2), alto (3), otra (4).
- Posesión de vivienda: Modalidades → Propietario (1), no propietario (2).

Entonces una tabla que resume una de las posibles repuestas del conjunto de individuos se presenta en la Tabla 4.

Tabla 4. Tabla resumen de respuestas para un conjunto de n individuos.

Individuo	G. de Edad	Género	N. Estudios	C. Socioeconómica	P. Vivienda
1	2	1	2	2	1
2	1	2	3	3	2
⋮	⋮	⋮	⋮	⋮	⋮
n	3	1	4	3	1

La cual tiene asociada la siguiente tabla dicotómica o disyuntiva y describe las k-respuestas para un grupo de n-individuos a través de un código binario.

Tabla 5. Tabla disyuntiva completa para los datos expuestos en la tabla 4.

Individuo	G. de Edad	Género	N. Estudios	C. Socioeconómica	P. Vivienda
1	0 1 0	1 0	0 1 0 0	0 1 0	1 0
2	1 0 0	0 1	0 0 1 0	0 0 1	0 1
⋮	⋮	⋮	⋮	⋮	⋮
n	0 0 1	1 0	0 0 0 1	0 0 1	1 0

□ *Distancia entre Filas y Columnas de una Matriz de Datos*

Para medir la distancia entre filas y columnas se recurre a la denominada distancia Chi cuadrado χ^2 , la cual es expresada para medir la distancia entre las filas i, i' como:

$$d^2(i, i') = \frac{1}{k} \sum_{j=1}^n \frac{n}{x_{.j}} (x_{ij} - x_{i'j})^2, \text{ individuos } i \text{ e } i' \text{ en } \mathbb{R}^p \quad (3.10)$$

Si los individuos estudiados tienen la característica de interés, la expresión entre paréntesis tomará el valor de 0, en caso contrario será 1. Esta distancia es ponderada por un peso igual a $\frac{n}{x_{.j}}$, lo que hace que una modalidad con una frecuencia baja tenga un peso mayor y una con

alta frecuencia un peso menor, de esta forma las modalidades menos comunes “raras” presentes en un individuo lo alejen de los demás.

En el caso de las columnas, la distancia entre las modalidades j y j' viene expresada por:

$$d^2(j, j') = \sum_{i=1}^n n \left(\frac{x_{ij}}{x_{.j}} - \frac{x_{ij'}}{x_{.j'}} \right)^2, \text{ modalidades } j \text{ e } j' \text{ en } \mathbb{R}^n \quad (3.11)$$

Con respecto a la distancia entre columnas o modalidades (Díaz, 2007, p. 412) comenta: “Dos modalidades que son seleccionadas por los mismos individuos coinciden pues $x_{ij} = x_{i'j}$ ” y agrega: “Las modalidades de frecuencia baja (las raras) están alejadas de las otras. En forma semejante, dos individuos están próximos si ellos han seleccionado las mismas modalidades; los individuos están alejados si no han respondido de la misma manera.”

Díaz (2007, p. 414) indica que algunas propiedades del ACM son las siguientes:

1. Es una representación gráfica de la asociación entre variables categóricas dos a dos; en consecuencia al análisis de correspondencia simple es un caso especial para un par de variables en particular.
2. A diferencia del análisis de componentes principales, los primeros ejes, aunque son los de mayor inercia explicada, justifican una pequeña parte de la variabilidad total.

3. La distancia de una modalidad al origen en el ACM es inversamente proporcional a su participación n_j . Es decir, las modalidades con participación baja aparecen más alejadas del origen que las modalidades de mayor frecuencia.
4. Las modalidades o categorías de una variable están centradas: es decir el centro de las modalidades de una misma variable es el origen del “nuevo” sistema de coordenadas. Así, las modalidades de una variable dicotómica se ubicarán en forma opuesta al origen.
5. El ACM es una descomposición de la nube de puntos de la varianza o inercia total del espacio de individuos (filas) o del espacio de las modalidades (columnas), en ciertas direcciones ortogonales, de tal forma que en cada dirección se maximice la inercia explicada respecto a la inercia no explicada por las direcciones previas.
6. La proyección de un individuo es el centro de gravedad de las modalidades que este ha escogido. Simétricamente, la proyección de una modalidad es el centro de gravedad de los individuos que la han escogido.

□ *Puntos Suplementarios.*

Son puntos correspondientes a una (o unas) variable de particular interés que aun cuando no interviene en la formación de los ejes, sus modalidades son proyectadas en el espacio factorial con la finalidad de ver su relación con las variables que están incluidas en el Análisis de Correspondencias. A estas variables, sin embargo, se les puede calcular la calidad de representación de fila o columna, así como también su contribución relativa a cada uno de los ejes construidos.

4. RESULTADOS DEL ANÁLISIS DE CORRESPONDENCIAS

Con el propósito de observar qué variables podrían influir en los resultados de las pruebas SABER PRO 2010 que presentan los estudiantes del programa de Ingeniería Civil en el país, se analizan en este capítulo los resultados obtenidos al aplicar el Análisis de Correspondencias descrito previamente en el Capítulo 3.

La primera parte de este apartado se enfoca en los resultados de los estudiantes, se expone la forma en que se categorizaron las variables continuas utilizadas, el estudio de las tablas de contingencia y los resultados del análisis de independencia entre las variables categóricas utilizando la prueba χ^2 . Finalmente, se realiza el Análisis de Correspondencia Simple (ACS) Y el Análisis de Correspondencias Múltiple (ACM).

En la segunda parte, se sigue el mismo procedimiento anterior pero enfocado al análisis de los resultados observados a nivel global o de las Instituciones de Educación Superior (IES).

4.1 RESULTADOS DE LOS ESTUDIANTES

Como para el estudio se cuenta con variables cualitativas y cuantitativas, se categorizan estas últimas considerando no solo su

distribución sino también, la representatividad de cada una de las modalidades propuestas.

4.1.1 Niveles de las variables categóricas

Las variables que se tienen en cuenta en esta primera parte del análisis son las que refieren a algunos aspectos de los estudiantes tales como: el género, el tipo de Institución de Educación Superior en la que estudia, el nivel de estudios alcanzado por los padres, la modalidad de bachillerato, el estrato social al que pertenece, si labora o no y el puntaje del estudiante.

- *GENERO*: sexo del estudiante: masculino (1), femenino (2).
- *TIPO*: corresponde al tipo de Institución de Educación Superior según el origen de los recursos económicos que obtiene: privada (1), pública (2).
- *ED_PADRE*: Corresponde al nivel de estudios alcanzado por el padre del estudiante: (1) ninguno, primaria incompleta, primaria completa; (2) bachillerato incompleto, bachillerato completo, tecnólogo incompleto o tecnólogo completo, (3) profesional incompleto, profesional completo y posgrado.
- *ED_MADRE*: Corresponde al nivel de estudios alcanzado por la madre del estudiante: (1) ninguno, primaria incompleta, primaria

completa; (2) bachillerato incompleto, bachillerato completo, tecnólogo incompleto o tecnólogo completo, (3) profesional incompleto, profesional completo y posgrado.

- ☐ *TRABAJA*: Refiere si el estudiante labora simultáneamente con sus estudios: (1) No trabaja, (2) Ayuda al sostenimiento del hogar, trabaja en prácticas de la carrera o para auto sostenerse y adquirir experiencia.
- ☐ *ESTRATO*: Relaciona el estrato socioeconómico al que pertenece el estudiante: (1) estratos uno (bajo-bajo), dos (bajo) y rural; (2) estrato tres (medio); (3) estrato cuatro (medio alto) y (4) estratos cinco (alto) y seis (alto – alto).
- ☐ *TITULO*: Corresponde al título de bachiller obtenido por el estudiante: (1) Bachiller Académico, (2) Otro (Bachiller Normalista o Bachiller Técnico).
- ☐ *PUNTAJE*: Puntuación obtenida por el estudiante en la evaluación Saber Pro: (1) puntajes inferiores a una desviación estándar por debajo de la media¹¹ (90 puntos), (2) puntajes entre una desviación estándar por debajo de la media y la media (entre 90 y 100 puntos), (3) puntajes entre la media y una desviación estándar por encima de la media (entre 100 y 110 puntos) y (4) los puntajes superiores a la media más una desviación estándar (110).

¹¹ Valores redondeados

La Tabla 6 presenta a continuación el resumen de estas variables con su correspondiente categorización.

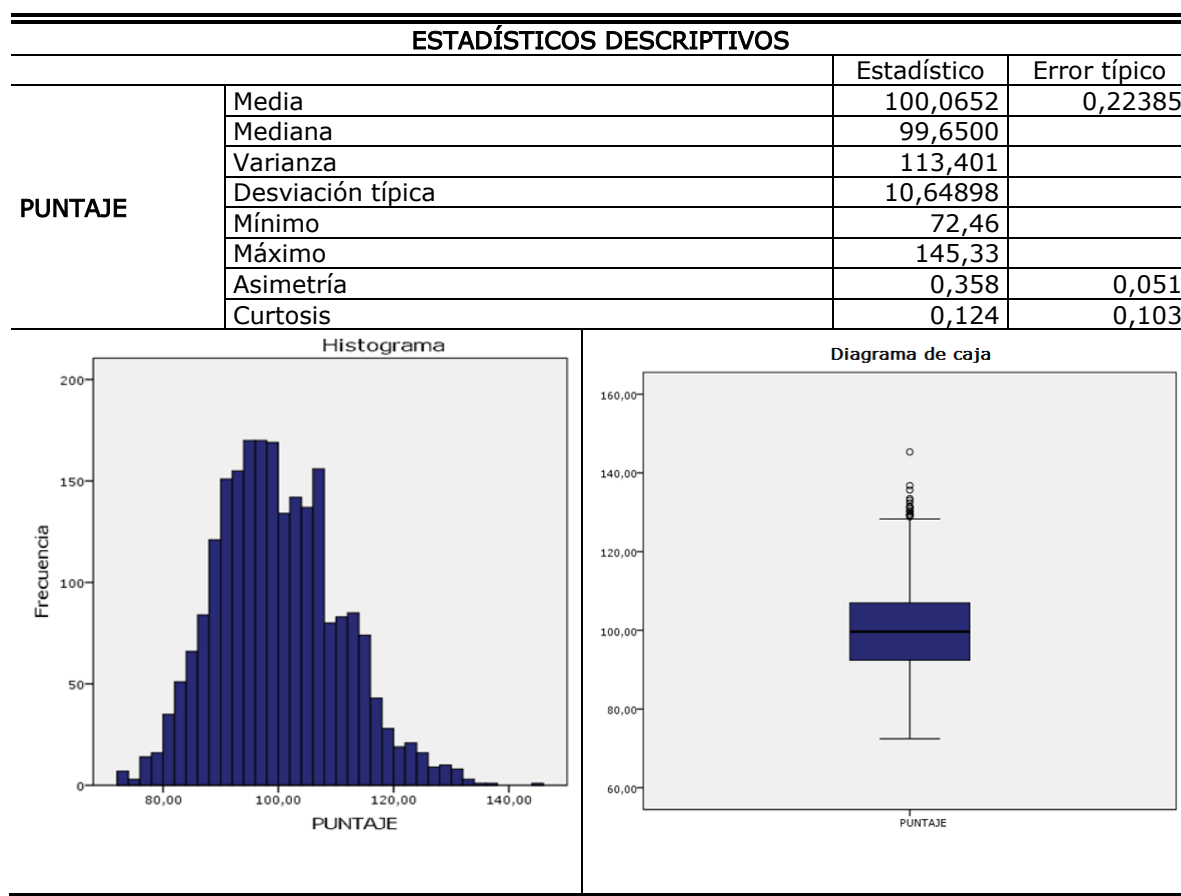
Tabla 6. Variables categóricas y sus respectivos niveles.

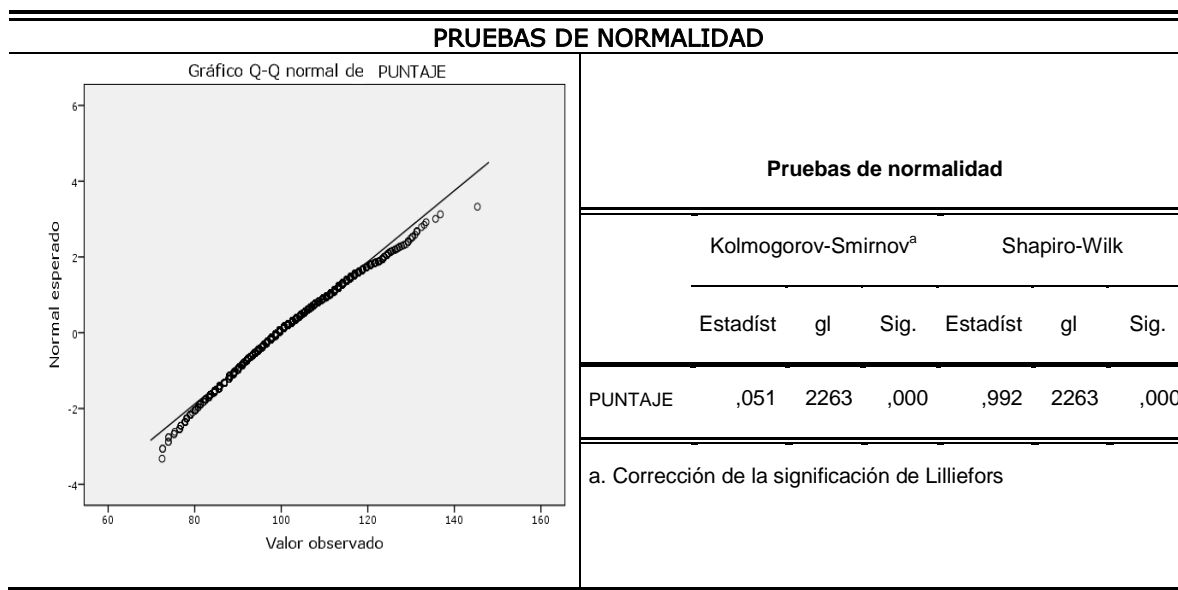
VARIABLE	No.	VALOR	
GÉNERO	1	MASCULINO	
	2	FEMENINO	
TIPO	1	PRIVADA	
	2	PÚBLICA	
EDUCACIÓN_PADRE	1	PRIMARIA	
	2	SECUNDARIA Y/O TECNOLOGÍA	
	3	UNIVERSITARIA Y/O POSGRADOS	
EDUCACIÓN_MADRE	1	PRIMARIA	
	2	SECUNDARIA Y/O TECNOLOGÍA	
	3	UNIVERSITARIA Y/O POSGRADOS	
ESTRATO	1	ESTRATOS BAJOS	
	2	ESTRATO MEDIO	
	3	ESTRATO MEDIO ALTO	
	4	ESTRATOS ALTOS	
TRABAJA	1	NO	
	2	SI	
TITULO	1	BACHILLER ACADÉMICO	
	2	OTRO	
PUNTAJE	1	[< 90]	BAJO
	2	[90 - 100)	MEDIO
	3	[100 - 110)	ALTOS
	4	[> = 110]	MUY ALTOS

4.1.2 Descripción de la variable *PUNTAJE*

Se pueden ver en la Tabla 7 los principales estadísticos de tendencia central y variabilidad para la variable Puntuación_prom, así como también el histograma de frecuencias, diagramas de caja y pruebas de normalidad, estas últimas, para este estudio en particular no son necesarias, pero sí para estudios posteriores.

Tabla 7. Descriptivos, histogramas, diagramas de caja y pruebas de normalidad para la variable *PUNTAJE*.





La información que se condensa en la tabla indica que la variable presenta una distribución con media igual a 100,0652, desviación 10,64898 y un valor en la mediana (99,65) muy cerca al valor de la media, el 50% de los estudiantes fueron evaluados con un puntaje superior a este valor, siendo el puntaje máximo alcanzado de 145,33 y el mínimo de 72,46.

El histograma de frecuencias presenta una curva de asimetría positiva, y en el diagrama de cajas deja ver una distribución con algunos datos extremos. Las pruebas de normalidad con un valor de significancia nulo rechazan la hipótesis de que los datos se distribuyen como una normal y la nube de puntos en el diagrama Q-Q presenta hacia los extremos un alejamiento de los puntos de la diagonal.

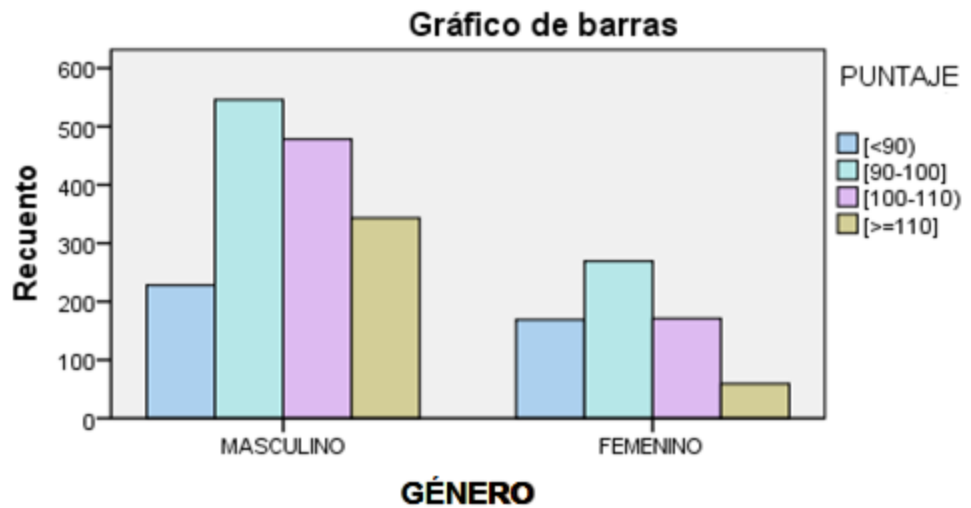
Para indagar sobre la asociación de algunas características de los estudiantes con relación a los puntajes obtenidos en la evaluación Saber Pro se realiza el análisis de las tablas de contingencia de las variables con relación al puntaje.

4.1.3 Tablas de contingencia e Independencia de las variables.

La Tabla 8 presenta la tabla de contingencia para la variable *Género* según el *Puntaje* obtenido por los estudiantes en la prueba, así como también los correspondientes valores de los perfiles fila para cada nivel de la variable género respecto del puntaje y en la parte inferior de la misma, el gráfico de barras.

Tabla 8. Tabla de contingencia GÉNERO * PUNTAJE.

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
GÉNERO	MASCULINO	Recuento	228	546	478	343	1595
		Perfil fila	14,3%	34,2%	30,0%	21,5%	100%
	FEMENINO	Recuento	169	269	171	59	668
		Perfil fila	25,3%	40,3%	25,6%	8,8%	100%
Total		Recuento	397	815	649	402	2263
		Perfil medio	17,5%	36,0%	28,7%	17,8%	100%



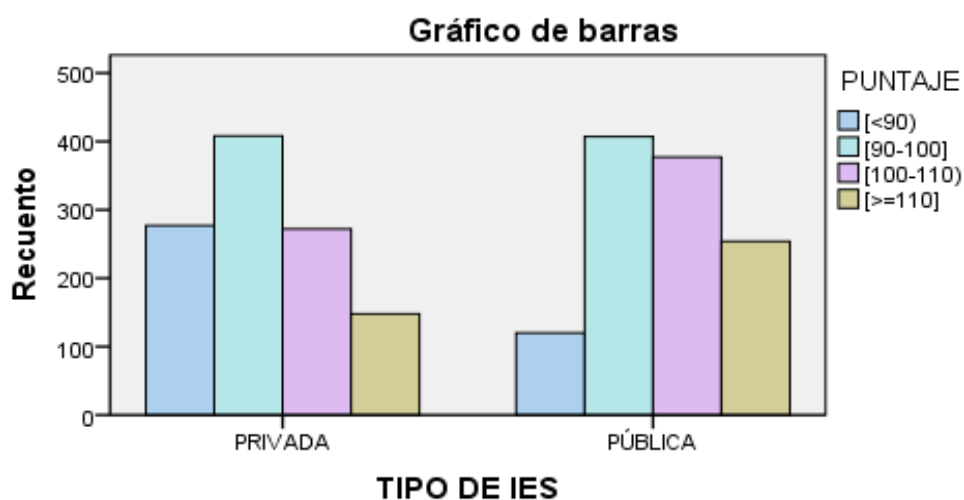
Como era de esperarse se observa que es mayor el número de estudiantes del género masculino (70,5%) que del género femenino (29,5%). Entre las mujeres el 40,3% obtuvieron puntajes entre una desviación estándar debajo de la media y la media, y en proporciones prácticamente iguales (25,3% y 25,6%) están por debajo y por encima de éste y sólo el 8,8% alcanzan los puntajes más altos superiores a 110. Entre los hombres, el 64,2% lograron puntajes a ± 1 desviación estándar de la media, presentando el menor número (14,3%), de estudiantes con puntajes bajos o inferiores a 90.

El gráfico de barras permite comparar los perfiles fila de un género con el otro. En el género masculino el perfil de los puntajes superiores a 110 es mayor que el de los puntajes menores de 90, opuesto a lo que se observa en el género femenino donde el perfil para los puntajes muy altos es menor que el perfil de los puntajes bajos.

La Tabla 9 presenta la tabla de contingencia para las variables *Tipo* por *Puntaje*, los perfiles fila y su correspondiente gráfico de barras en la parte inferior de la misma.

Tabla 9. Tabla de contingencia TIPO * PUNTAJE.

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
TIPO	PRIVADA	Recuento	277	408	272	148	1105
		Perfil fila	25,1%	36,9%	24,6%	13,4%	100,0%
	PÚBLICA	Recuento	120	407	377	254	1158
		Perfil fila	10,4%	35,1%	32,6%	21,9%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil medio	17,5%	36,0%	28,7%	17,8%	100,0%



El 51,2% de los estudiantes que presentaron la prueba pertenecen a instituciones de carácter público y el restante 48,8% a instituciones privadas. Tanto para las unas como para las otras, un poco más del 60% de los estudiantes obtuvieron puntajes entre [90 – 110), siendo

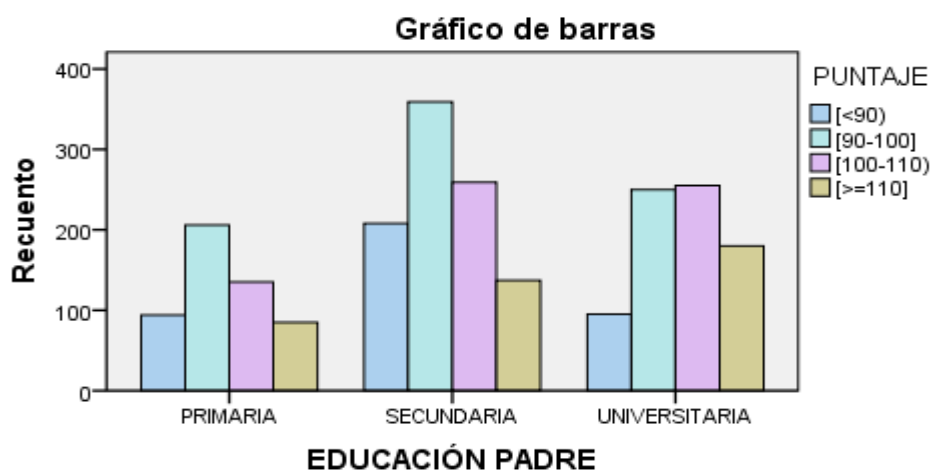
opuestos entre sí los porcentajes para los otros dos niveles de la variable puntuación.

Al comparar los perfiles fila entre los tipos de Instituciones de Educación Superior (IES), se observa un comportamiento diferente para cada uno de ellos, con diferencias regulares entre los perfiles de las privadas y diferencias muy irregulares entre los perfiles de las instituciones públicas.

La Tabla 10 muestra la tabla de contingencia para la variable *Educación Padre* por *Puntajes*, los perfiles filas y en la parte inferior el gráfico de barras respectivo.

Tabla 10. Tabla de contingencia EDUCACIÓN PADRE * PUNTAJE

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
EDUCACIÓN PADRE	PRIMARIA	Recuento	94	206	135	85	520
		Perfil fila	18,1%	39,6%	26,0%	16,3%	100,0%
	SECUNDARIA	Recuento	208	359	259	137	963
		Perfil fila	21,6%	37,3%	26,9%	14,2%	100,0%
	UNIVERSITARIA	Recuento	95	250	255	180	780
		Perfil fila	12,2%	32,1%	32,7%	23,1%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil medio	17,5%	36,0%	28,7%	17,8%	100,0%



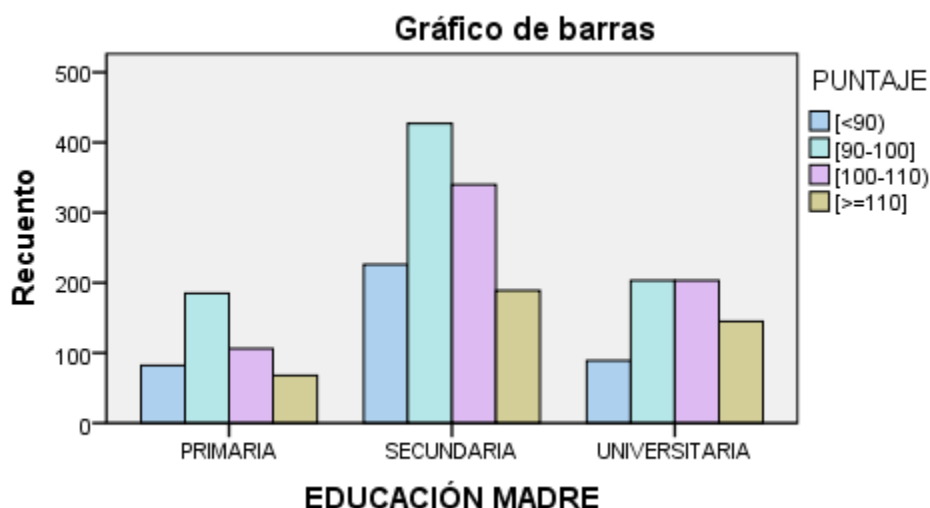
El 23% de los padres de los estudiantes de Ing. Civil que presentaron las pruebas Saber Pro en el año 2010 habían alcanzado a hacer máximo la primaria completa, el 42,6% la secundaria completa o una tecnología y el 34,5% estudios universitarios o posgrados. De los 402 estudiantes que obtuvieron puntajes más altos, el 44,8% tienen padres con estudios universitarios, siendo éste el nivel de mayor frecuencia. De los 397 que obtuvieron puntajes bajos, más de la mitad (52,4%) tienen padres con estudios de secundaria.

Para el nivel *universitario* de la variable *Educación padre* se observa un comportamiento muy diferente de los perfiles fila con relación a los otros: mayor el perfil del tercer nivel que el segundo y mayor el perfil del cuarto nivel de puntuación que el primero.

La Tabla 11 muestra la tabla de contingencia para la variable *Educación Madre* por *Puntaje*, los valores correspondientes a los perfiles fila y en la parte inferior el gráfico de barras.

Tabla 11. Tabla de contingencia EDUCACIÓN MADRE * PUNTAJE

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
EDUCACIÓN MADRE	PRIMARIA	Recuento	82	185	106	68	441
		Perfil fila	18,6%	42,0%	24,0%	15,4%	100,0%
	SECUNDARIA	Recuento	226	427	340	189	1182
		Perfil fila	19,1%	36,1%	28,8%	16,0%	100,0%
	UNIVERSITARIA	Recuento	89	203	203	145	640
		Perfil fila	13,9%	31,7%	31,7%	22,7%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil fila	17,5%	36,0%	28,7%	17,8%	100,0%



Del total de 2263 estudiantes que presentaron la prueba en el programa de Ing. Civil las madres del 19,5% de ellos habían alcanzado como máximo nivel educativo la educación primaria, el 52,2% la secundaria y el 28,3 la universitaria. Un poco más de la mitad de las madres para todos los niveles de puntuación lograron estudios secundarios y el

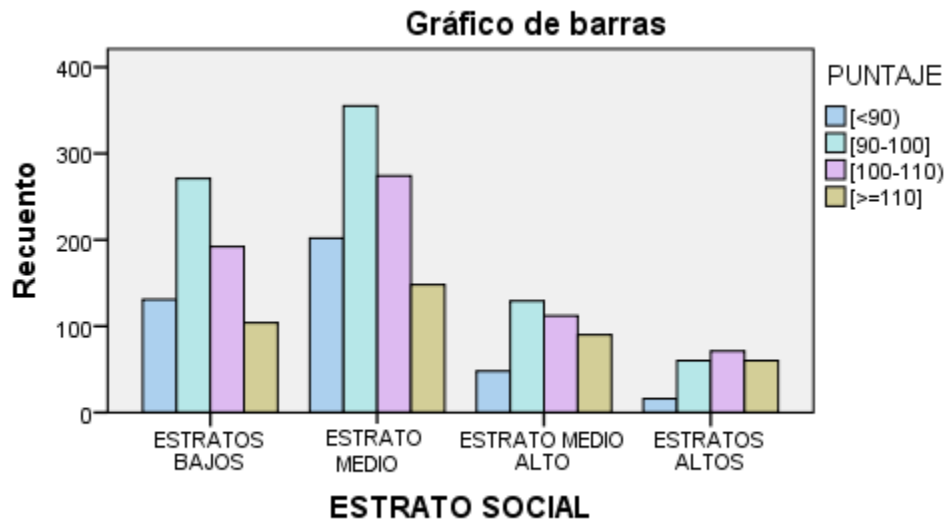
31,3% y el 36,1% de las madres de los estudiantes que lograron los puntajes altos y muy altos respectivamente habían realizado estudios universitarios.

La distribución de los perfiles fila para el nivel *universitario* de la variable *Educación madre* presenta un comportamiento diferente al de los niveles *primaria y secundaria*.

En la Tabla 12 se presenta la tabla de contingencia para la variable *Trabaja por Puntaje*, así como también los valores correspondientes a los perfiles fila; en la parte final el gráfico de barras correspondiente.

Tabla 12. Tabla de contingencia ESTRATO * PUNTAJE.

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
ESTRATO	ESTRATOS BAJOS	Recuento	131	271	192	104	698
		Perfil fila	18,8%	38,8%	27,5%	14,9%	100,0%
	ESTRATO MEDIO	Recuento	202	355	274	148	979
		Perfil fila	20,6%	36,3%	28,0%	15,1%	100,0%
	ESTRATO MEDIO ALTO	Recuento	48	129	112	90	379
		Perfil fila	12,7%	34,0%	29,6%	23,7%	100,0%
	ESTRATOS ALTOS	Recuento	16	60	71	60	207
		Perfil fila	7,7%	29,0%	34,3%	29,0%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil fila	17,5%	36,0%	28,7%	17,8%	100,0%



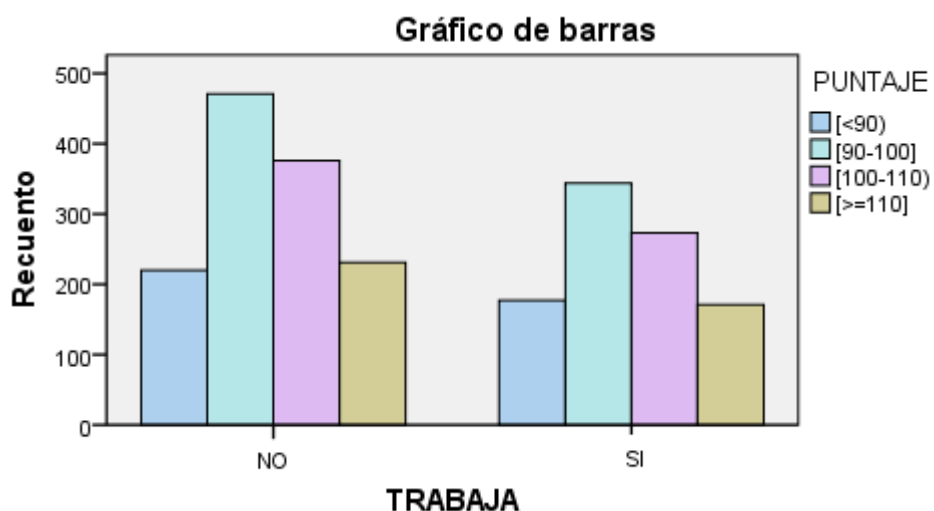
El 74,1% de los estudiantes pertenecen a los estratos bajos (30,8%) y medios (43,3%). Entre los estudiantes que obtuvieron menos de 90 puntos la mitad (50,9%) están en el estrato medio y el 4% en los estratos altos. De forma similar, entre los estudiantes con puntajes superiores a 110 la mayor frecuencia es de estudiantes en estrato medio (36,8%) y la menor en estratos altos (14,9%).

Al observar el gráfico de los perfiles fila para todos los niveles de la variable *Estrato* se aprecia que para los *estratos bajo y medio* se presentan similitudes entre sí pero difieren de las distribuciones de los estratos medio alto y altos que también presentan diferencias entre ellos.

A continuación la Tabla 13 presenta la respectiva tabla de contingencia para las variables *Trabaja por Puntaje*, así como también los valores de los perfiles fila con su diagrama de barras.

Tabla 13. Tabla de contingencia TRABAJA * PUNTAJE

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
TRABAJA	NO	Recuento	220	471	376	231	1298
		Perfil fila	16,9%	36,3%	29,0%	17,8%	100,0%
	SI	Recuento	177	344	273	171	965
		Perfil fila	18,3%	35,6%	28,3%	17,7%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil fila	17,5%	36,0%	28,7%	17,8%	100,0%



El 42,6% de los estudiantes ayudan al sostenimiento del hogar, trabajan en prácticas de la carrera o para auto sostenerse y el 42,6% no realizan ninguna labor remunerada económicamente, porcentaje que se mantiene muy parecido para todos los niveles de la variable puntuación.

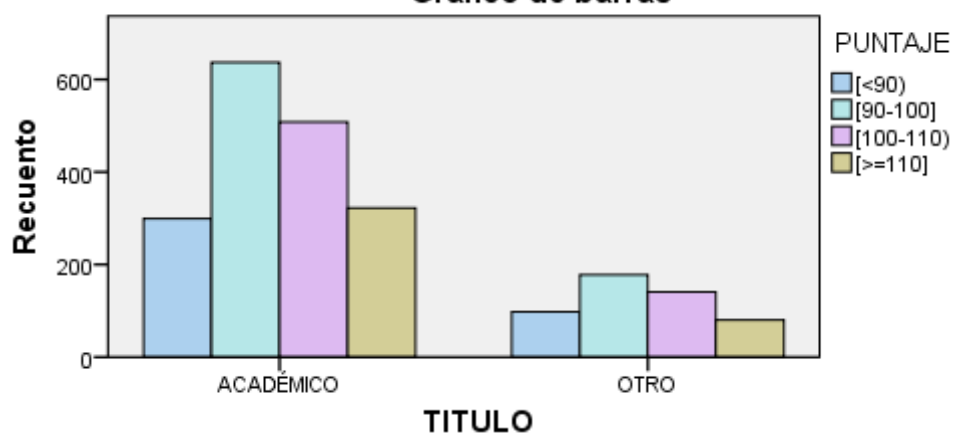
El gráfico de barras muestra una distribución de frecuencias para cada nivel de la variable *Trabaja* muy parecida, con un gran aumento entre el primer y segundo nivel y descensos regulares hacia el cuarto nivel.

En la Tabla 14 se puede observar la tabla de contingencia para la variable *Título* por *Puntaje*, los valores de los perfiles fila y el gráfico de barras correspondiente.

Tabla 14. Tabla de contingencia TITULO * PUNTAJE

			PUNTAJE				Total
			[<90)	[90-100)	[100-110)	[>=110]	
TITULO	ACADÉMICO	Recuento	299	637	508	322	1766
		Perfil fila	16,9%	36,1%	28,8%	18,2%	100,0%
	OTRO	Recuento	98	178	141	80	497
		Perfil fila	19,7%	35,8%	28,4%	16,1%	100,0%
Total		Recuento	397	815	649	402	2263
		Perfil fila	17,5%	36,0%	28,7%	17,8%	100,0%

Gráfico de barras



El 78% de los estudiantes evaluados tienen el título de Bachiller Académico y el restante 22% el de Bachiller Normalista o Bachiller Técnico, porcentajes que se mantienen muy similares en todos los niveles de la variable *puntaje*. De igual manera, al observar el diagrama de barras se advierte que para cada nivel de la variable *Título* la distribución de los perfiles fila es muy similar.

Previo al análisis de correspondencias se relacionan en la Tabla 15 los resultados del estudio de independencia entre las variables categóricas utilizando la prueba χ^2 para tablas de contingencia.

Tabla 15. Pruebas Ji-cuadrado de Pearson.

	χ^2	gl	Valor_p	Resultado
GÉNERO * PUNTAJE	82,964 ^a	3	,000	Dependientes
TIPO * PUNTAJE	105,844 ^a	3	,000	Dependientes
EDUCACIÓN PADRE * PUNTAJE	54,163 ^a	6	,000	Dependientes
EDUCACIÓN MADRE * PUNTAJE	31,838 ^a	6	,000	Dependientes
ESTRATO * PUNTAJE	59,502 ^a	9	,000	Dependientes
TRABAJA * PUNTAJE	,765 ^a	3	,858	Independientes
TÍTULO * PUNTAJE	2,743 ^a	3	,433	Independientes

El valor_p (0,000) menor que el nivel de significancia (0,05) indica que se rechaza el supuesto de homogeneidad para las tablas de esas variables, concluyendo que es probable que existan diferencias entre los niveles de éstas en relación con los perfiles de los niveles de puntaje.

La prueba ji-cuadrado respalda la independencia de las variables *trabaja* y *título* en asociación con la variable puntaje, señalando que ni el título de bachiller obtenido por el estudiante, ni si el individuo trabaja o no, tienen relación con los resultados que obtiene en la pruebas. Por el contrario para las variables *género*, *tipo*, *educación padre*, *educación madre* y *estrato* se concluye que sí juegan algún papel en los resultados.

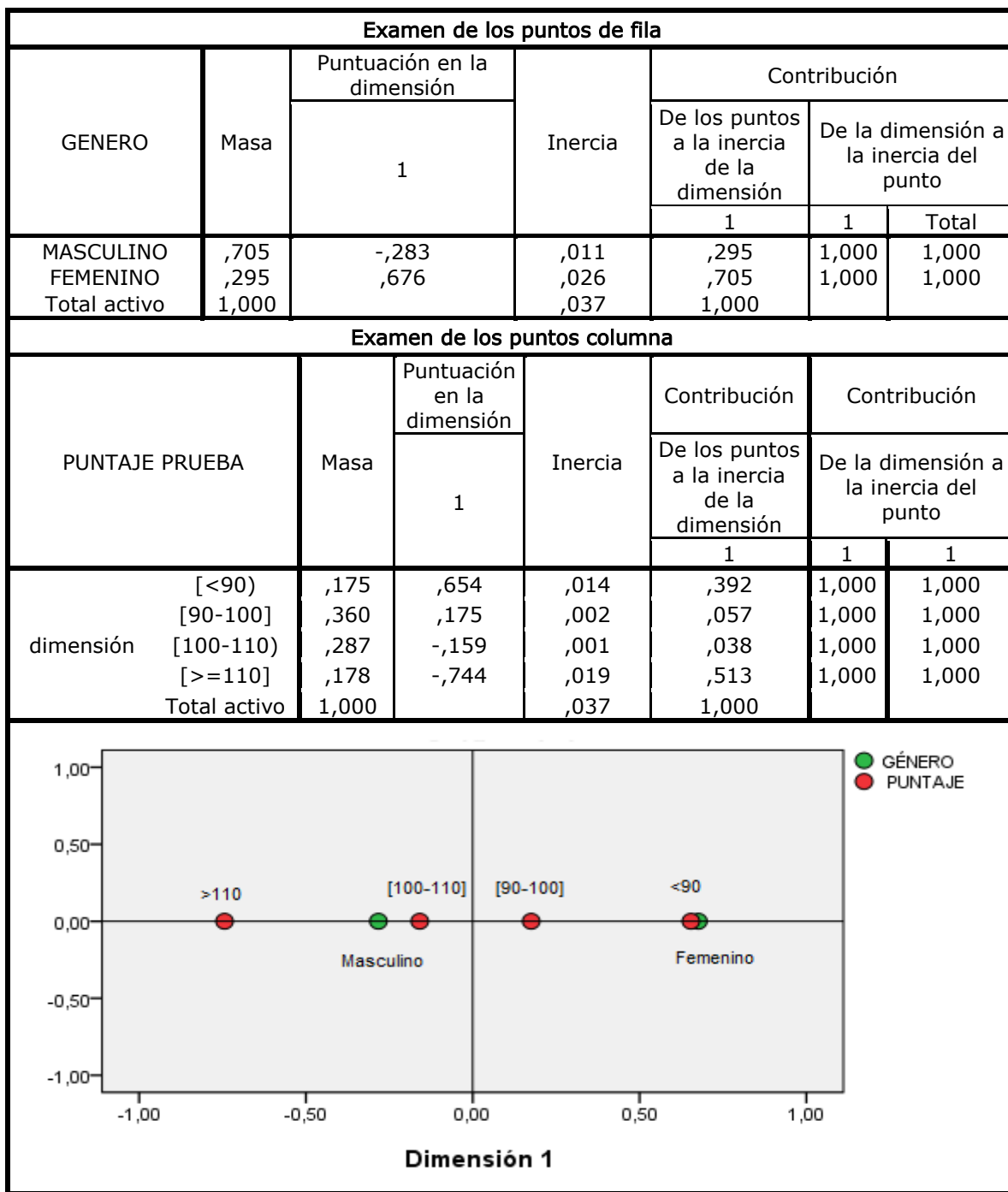
4.1.4 Análisis de Correspondencia Simple (ACS).

Se conoce que la inercia es el promedio de las distancias ponderadas por la masa de los distintos puntos a su centro de gravedad y teniendo en cuenta que cada punto tiene un peso o ponderación igual a su masa ahora el interés se extiende a medir la variabilidad de la tabla de correspondencias mediante la inercia.

□ GÉNERO * PUNTAJE

Tabla 16. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable género.

Resumen					
Dimensión	Valor propio	Inercia	Proporción de inercia	Proporción de inercia	Confianza para el Valor propio
			Explicada	Acumulada	Desviación típica
1 dimensión	,191	,037	1,000	1,000	,019
Total		,037	1,000	1,000	



En el resumen del modelo se observa que una sola dimensión contribuye 100% en la inercia total. En el examen de los puntos fila se muestran

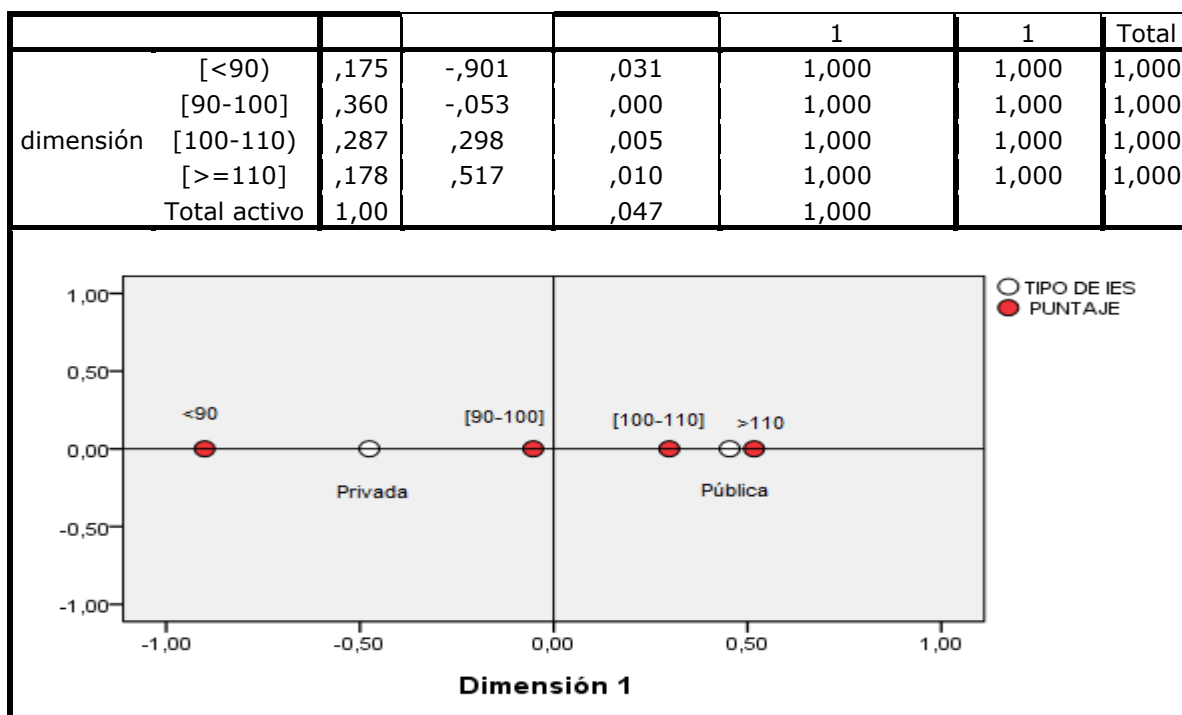
las contribuciones totales de cada perfil fila a la inercia de la dimensión, siendo el punto fila de mayor contribución (0,705) el género femenino.

Las modalidades de la variable puntuación quedan ubicadas en un orden descendente hacia la derecha: el género femenino se asocia con los puntajes bajos y el masculino con los puntajes altos.

□ TIPO * PUNTAJE

Tabla 17. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable tipo.

Resumen						
Dimensión	Valor propio	Proporción de inercia		Confianza para el Valor propio		
		Explicada	Acumulada	Desviación típica		
dimensión 1	,216	1,000	1,000	,020		
Total		1,000	1,000			
Examen de los puntos de fila						
TIPO DE IES	Masa	Puntuación en la dimensión 1	Inercia	Contribución		
				De los puntos a la inercia de la dimensión	De la dimensión a la inercia del punto	
		1		1	1	Total
PRIVADA	,488	-,476	,024	,512	1,00	1,000
PÚBLICA	,512	,454	,023	,488	1,00	1,000
Total activo	1,00		,047	1,00		
Examen de los puntos columna						
PUNTAJE PRUEBA	Masa	Puntuación en la dimensión	Inercia	Contribución		
		1		De los puntos a la inercia de la dimensión	De la dimensión a la inercia del punto	

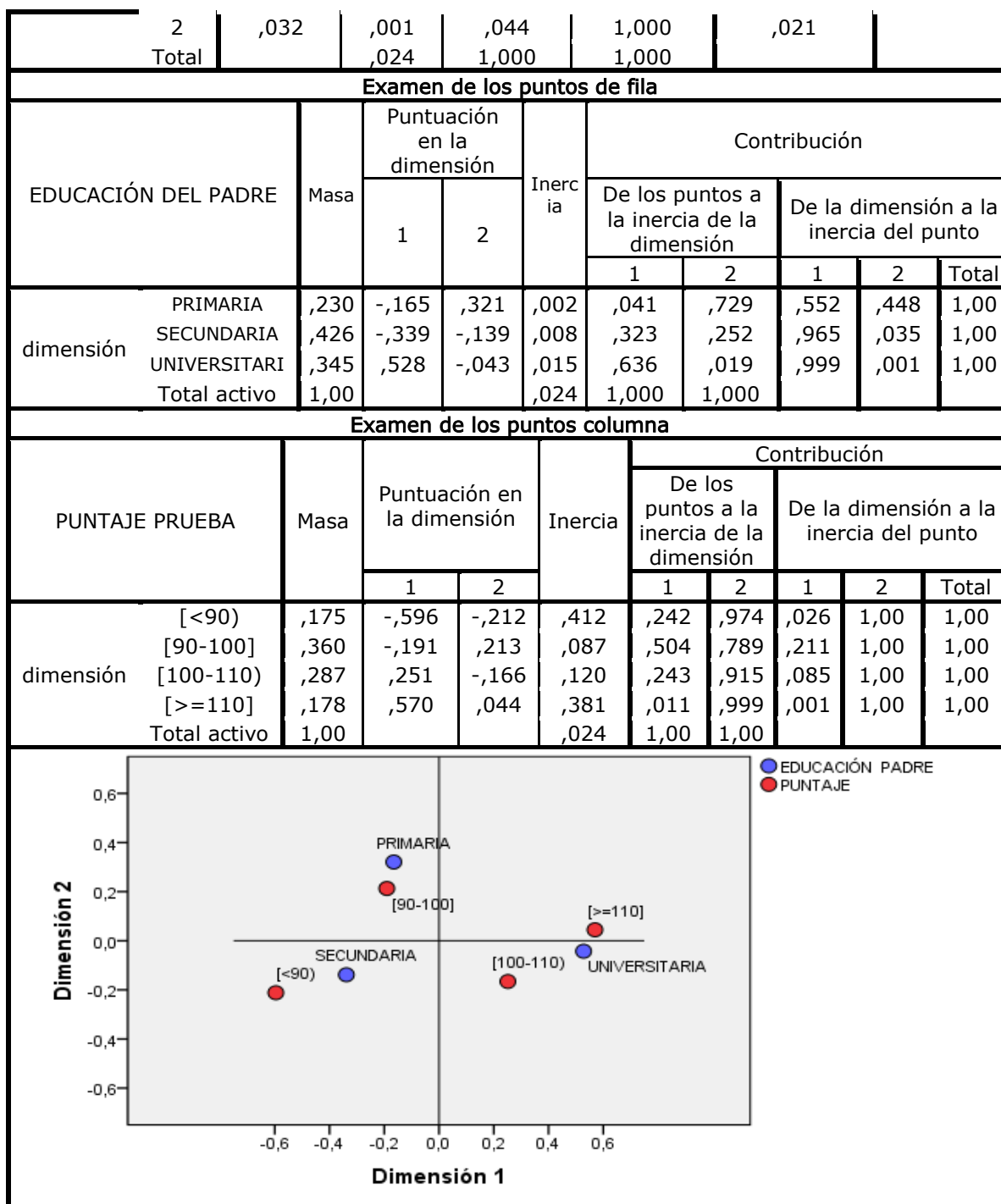


Las cuatro categorías de la variable puntaje quedan en orden ascendente hacia la derecha, las instituciones con carácter público se asocian mejor con los puntajes altos y muy altos y las de carácter privado con los puntajes medios y bajos.

□ EDUCACIÓN PADRE * PUNTAJE

Tabla 18. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Educación Padre.

Resumen						
Dimensión	Valor propio	Inercia	Proporción de inercia		Confianza para el Valor propio	
			Explicada	Acumulada	Desviación típica	Correlación
dimensión 1	,151	,023	,956	,956	,021	,025



En este caso, el primer eje explica el 95,6% de la inercia total lo que simplifica el análisis a uno solo de los ejes principales. Este eje está definido por los puntajes en orden creciente de izquierda a derecha,

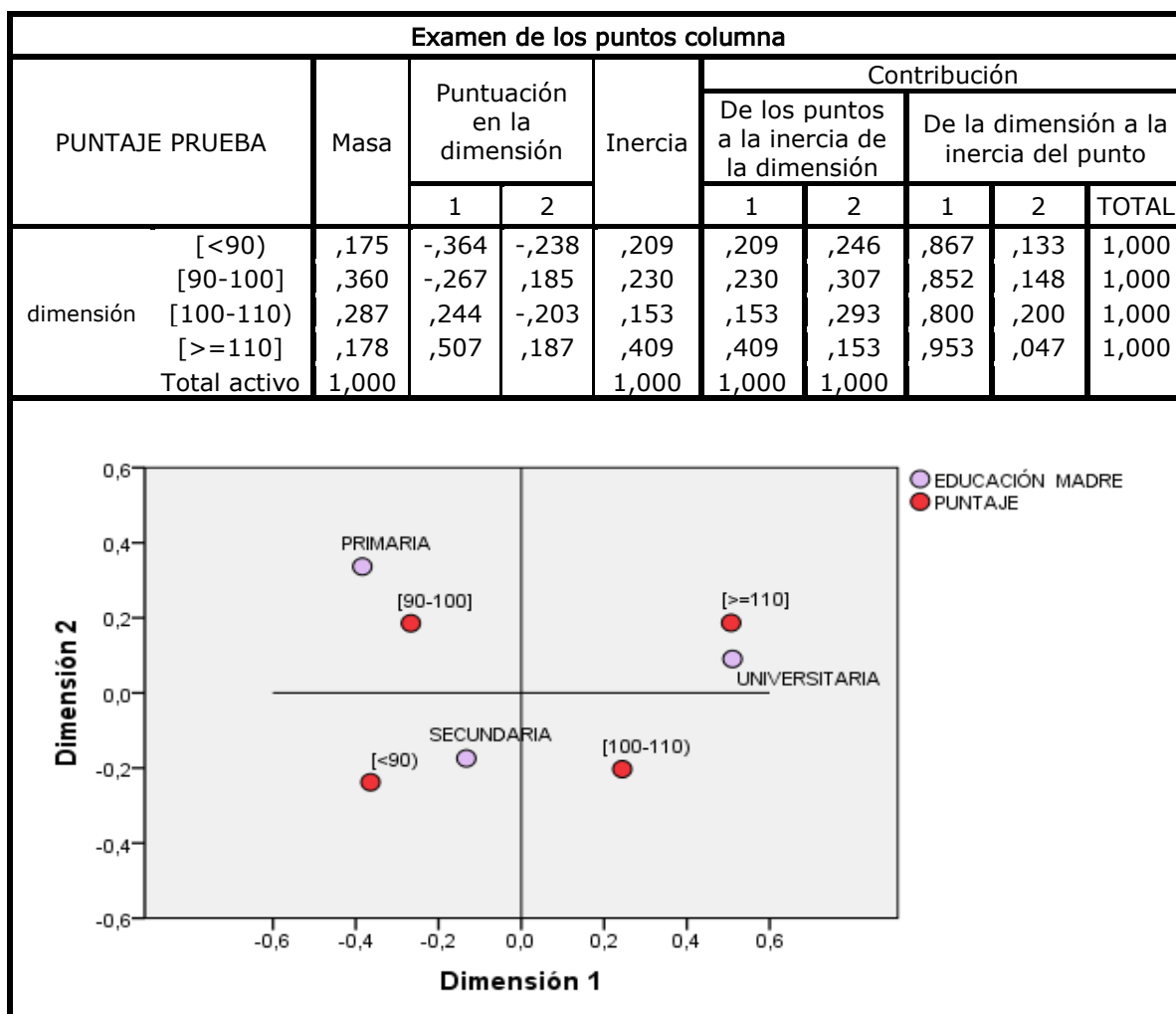
siendo [<90) con coordenada (puntuación en la dimensión) -0,596 el valor en el extremo izquierdo del eje y el nivel [≥ 110] con coordenada 0,570 el valor en el extremo derecho del eje. Las categorías educativas del padre están bien explicadas por este eje principal (valores de explicación superiores a 0,5).

En el gráfico al final de la Tabla 12 se observa como los estudiantes con puntajes superiores se asocian con padres que tienen educación universitaria y los de puntajes inferiores con padres que solo tienen la primaria o la secundaria.

□ EDUCACIÓN MADRE * PUNTAJE

Tabla 19. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Educación Madre.

Resumen										
Dimensión	Valor propio	Inercia	Proporción de inercia		Confianza para el Valor propio					
			Explicada	Acumulada	Desviación típica	Correlación				
1	,112	,012	,885	,885	,021	,012				
dimensión 2	,040	,002	,115	1,000	,021					
Total		,014	1,000	1,000						
Examen de los puntos fila										
EDUCACIÓN DE LA MADRE		Masa	Puntuación en la dimensión		Inercia	Contribución				
						De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
			1	2		1	2	1	2	Total
dimensión	PRIMARIA	,195	-,384	,336	,258	,547	,783	,217	1,000	1,00
	SECUNDARIA	,522	-,133	-,175	,083	,395	,616	,384	1,000	1,00
	UNIVERSITARIA	,283	,510	,091	,660	,058	,989	,011	1,000	1,00
	Total activo	1,000			,014	1,000	1,000			

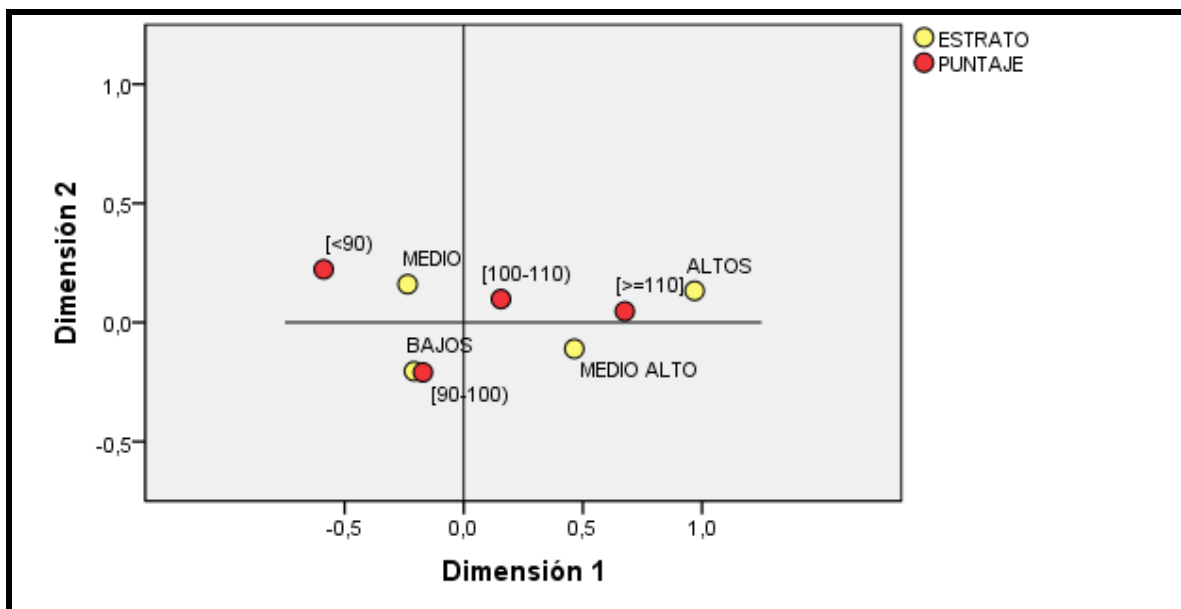


El primer eje explica el 88.5% de la inercia total y al igual que con la variable anterior, este eje está definido por los puntajes en orden creciente de izquierda a derecha. La mayor puntuación en la dimensión 1 la presenta el nivel universitario (0,51) y la menor (-0,384) el nivel primaria. De forma similar a la educación del padre, los estudiantes con mejores puntajes se asocian con madres que tienen formación universitaria y los estudiantes con puntajes medios y bajos con madres que tienen estudios de primaria o secundaria.

□ ESTRATO * PUNTAJE

Tabla 20. Contribuciones a la inercia de cada dimensión y contribuciones de los perfiles fila y columna para la variable Estrato.

Resumen									
Dimensión	Valor propio	Inercia	Proporción de inercia		Confianza para el Valor propio				
			Explicada	Acumulada	Desviación típica		Correlación		
1	,159	,025	,965	,965	,021		,035		
2	,028	,001	,029	,994	,021				
3	,012	,000	,006	1,000					
Total		,026	1,000	1,000					
Examen de los puntos de fila									
ESTRATO SOCIAL	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
BAJOS	,308	-,208	-,204	,003	,084	,466	,849	,142	,991
MEDIO	,433	-,236	,160	,004	,151	,402	,925	,074	,999
MEDIO ALTO	,167	,464	-,110	,006	,226	,074	,977	,010	,986
ALTOS	,091	,969	,133	,014	,539	,058	,993	,003	,997
Total activo	1,00			,026	1,00	1,00			
Examen de los puntos columna									
PUNTAJE PRUEBA	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
[<90]	,175	-,588	,222	,380	,380	,313	,974	,024	,998
[90-100]	,360	-,171	-,210	,066	,066	,572	,793	,207	1,000
[100-110]	,287	,156	,098	,044	,044	,100	,872	,061	,932
[>=110]	,178	,676	,047	,510	,510	,014	,996	,001	,997
Total activo	1,00			1,00	1,00	1,00			



En la parte superior de la Tabla 20 –Resumen- se muestran las contribuciones de cada una de las 3 dimensiones a la inercia total. Se observa que solamente la primera dimensión contribuye con el 96,5% a dicha inercia y que con la segunda dimensión suman una contribución del 99,4% por lo que se puede decir que la primera dimensión captura la dependencia observada en las Tabla 15.

En la parte media de la misma tabla –Examen de los puntos fila- se observa que la primera dimensión discrimina por estratos, siendo los puntos fila que más contribuyen a la inercia de la primera dimensión el estrato altos (0,539) y el estrato medio alto (0,226). En el gráfico al final se advierte que los puntajes superiores se asocian con los estratos más altos y los puntajes inferiores con los estratos más bajos.

4.1.5 Análisis de Correspondencias Múltiple (ACM)

Seguidamente se realiza el ACM, tomando como variables activas las que en los análisis anteriores exhibieron dependencia con la variable *Puntaje: género, tipo, educación del padre, educación de la madre y estrato*; y como variables suplementarias, *título y trabaja* que manifestaron independencia.

Para hacer estos análisis se utilizó el programa Software Process Analysis Design And Enactment® (SPAD 5.6).

En la Tabla 21 se muestran las contribuciones de cada una de las dimensiones a la inercia total. Se observan de igual forma los valores correspondientes a las coordenadas de los puntos filas, sus pesos y las contribuciones absolutas y relativas (cosenos cuadrados).

Tabla 21. Valores propios, coordenadas y contribuciones.

EIGENVALUES					
COMPUTATIONS PRECISION SUMMARY : TRACE BEFORE DIAGONALISATION..				2.0000	
				SUM OF EIGENVALUES.....	2.0000
HISTOGRAM OF THE FIRST 12 EIGENVALUES					
+-----+-----+-----+-----+-----+-----+					
NUMBER	EIGENVALUE	PERCENTAGE	CUMULATED		
			PERCENTAGE		
+-----+-----+-----+-----+-----+-----+					
1	0.3279	16.39	16.39	*****	
2	0.2497	12.48	28.88	*****	
3	0.2105	10.52	39.40	*****	
4	0.1748	8.74	48.14	*****	
5	0.1692	8.46	56.60	*****	
6	0.1644	8.22	64.82	*****	
7	0.1566	7.83	72.65	*****	
8	0.1479	7.39	80.04	*****	
9	0.1255	6.28	86.32	*****	
10	0.1077	5.38	91.70	*****	
11	0.0968	4.84	96.55	*****	
12	0.0691	3.45	100.00	*****	

LOADINGS, CONTRIBUTIONS AND SQUARED COSINES OF ACTIVE CATEGORIES

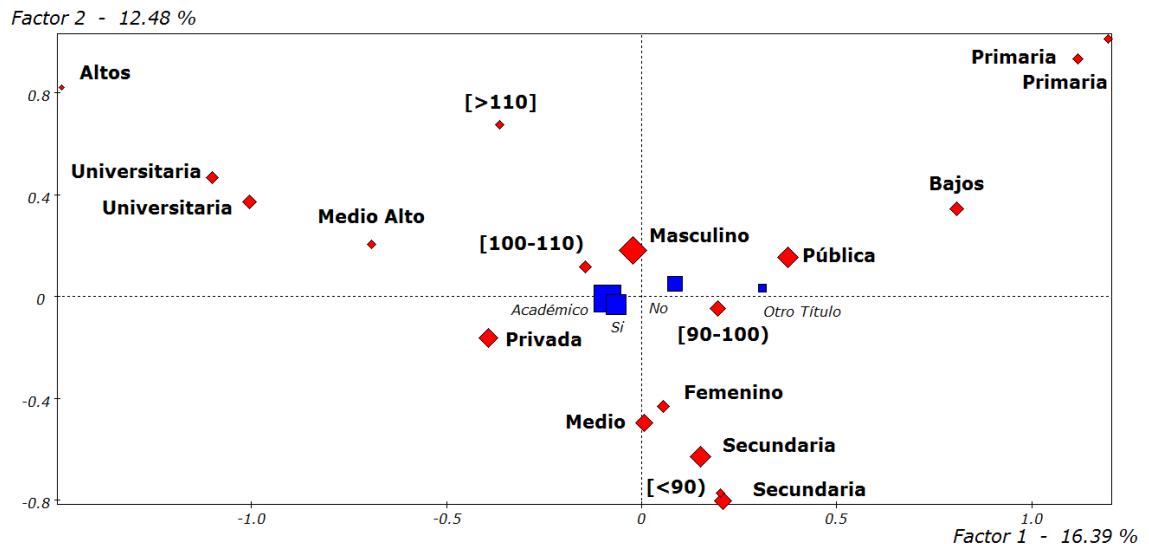
AXES 1 TO 5

CATEGORIES				LOADINGS					CONTRIBUTIONS					SQUARED COSINES				
IDEN - LABEL	REL. WT.	DISTO		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1 . GENERO																		
1 - Masculino	11.75	0.42	-0.02	0.18	-0.26	-0.37	0.04		0.0	1.5	3.8	9.2	0.1	0.00	0.08	0.16	0.33	0.0
2 - Femenino	4.92	2.39	0.05	-0.43	0.63	0.88	-0.10		0.0	3.7	9.2	21.9	0.3	0.00	0.08	0.16	0.33	0.0
								CUMULATED CONTRIBUTION = 0.1 5.2 13.0 31.1 0.4										
2 . TIPO																		
1 - Privada	8.14	1.05	-0.39	-0.16	0.63	-0.27	0.06		3.8	0.9	15.4	3.3	0.2	0.15	0.02	0.38	0.07	0.0
2 - Pública	8.53	0.95	0.37	0.15	-0.60	0.25	-0.06		3.6	0.8	14.7	3.1	0.2	0.15	0.02	0.38	0.07	0.0
								CUMULATED CONTRIBUTION = 7.5 1.7 30.1 6.4 0.3										
4 . ED_PADRE																		
1 - Primaria	3.83	3.35	1.12	0.93	0.47	-0.10	0.02		14.6	13.3	4.1	0.2	0.0	0.37	0.26	0.07	0.00	0.0
2 - Secundaria	7.09	1.35	0.21	-0.80	-0.29	-0.15	0.03		0.9	18.4	2.9	0.9	0.0	0.03	0.48	0.06	0.02	0.0
3 - Universitaria	5.74	1.90	-1.00	0.37	0.04	0.25	-0.05		17.7	3.2	0.1	2.1	0.1	0.53	0.07	0.00	0.03	0.0
								CUMULATED CONTRIBUTION = 33.2 34.9 7.0 3.2 0.1										
5 . ED_MADRE																		
1 - Primaria	3.25	4.13	1.19	1.01	0.65	-0.17	-0.01		14.1	13.3	6.5	0.5	0.0	0.35	0.25	0.10	0.01	0.0
2 - Secundaria	8.71	0.91	0.15	-0.63	-0.35	-0.01	0.04		0.6	13.9	5.1	0.0	0.1	0.02	0.43	0.13	0.00	0.0
3 - Universitaria	4.71	2.54	-1.10	0.47	0.20	0.13	-0.07		17.4	4.1	0.9	0.5	0.2	0.48	0.09	0.02	0.01	0.0
								CUMULATED CONTRIBUTION = 32.1 31.3 12.4 1.0 0.3										
6 . ESTRATO																		
1 - Bajos	5.14	2.24	0.81	0.34	-0.06	0.21	-0.15		10.2	2.4	0.1	1.3	0.7	0.29	0.05	0.00	0.02	0.0
2 - Medio	7.21	1.31	0.01	-0.50	0.01	-0.39	-0.14		0.0	7.2	0.0	6.2	0.8	0.00	0.19	0.00	0.11	0.0
3 - Medio Alto	2.79	4.97	-0.69	0.21	-0.14	1.04	1.21		4.1	0.5	0.3	17.2	24.1	0.10	0.01	0.00	0.22	0.2
4 - Altos	1.52	9.93	-1.49	0.82	0.39	-0.78	-1.04		10.3	4.1	1.1	5.3	9.8	0.22	0.07	0.02	0.06	0.1
								CUMULATED CONTRIBUTION = 24.6 14.2 1.5 30.0 35.4										
8 . PUNTAJE																		
1 - [<90)	2.92	4.70	0.20	-0.77	1.18	-0.62	0.21		0.4	7.0	19.3	6.5	0.8	0.01	0.13	0.30	0.08	0.0
2 - [90-100)	6.00	1.78	0.19	-0.05	0.22	0.45	0.33		0.7	0.1	1.4	6.9	3.9	0.02	0.00	0.03	0.11	0.0
3 - [100-110)	4.78	2.49	-0.14	0.12	-0.46	0.34	-1.19		0.3	0.3	4.9	3.1	39.9	0.01	0.01	0.09	0.05	0.5
4 - [>110]	2.96	4.63	-0.36	0.67	-0.86	-0.83	1.04		1.2	5.4	10.4	11.8	19.0	0.03	0.10	0.16	0.15	0.2
								CUMULATED CONTRIBUTION = 2.5 12.7 35.9 28.3 63.5										

En la parte superior de la Tabla 21 se aprecia que con una dimensión se capta el 16,39% de la variabilidad total de los datos, y conjuntamente con la segunda dimensión (12,48%) explican solo el 28,87%, indicando que este plano no percibe de forma satisfactoria la asociación completa entre las variables de estudio. Con tres dimensiones ese valor llega a 39,40%, siendo necesario un mínimo de 5 dimensiones para explicar un poco más del 50% de la variabilidad observada en los datos.

En la Figura 1 se presenta el primer plano factorial con las proyecciones de las distintas modalidades de las variables activas.

*Figura 1. Plano factorial eje 1 * eje 2.*

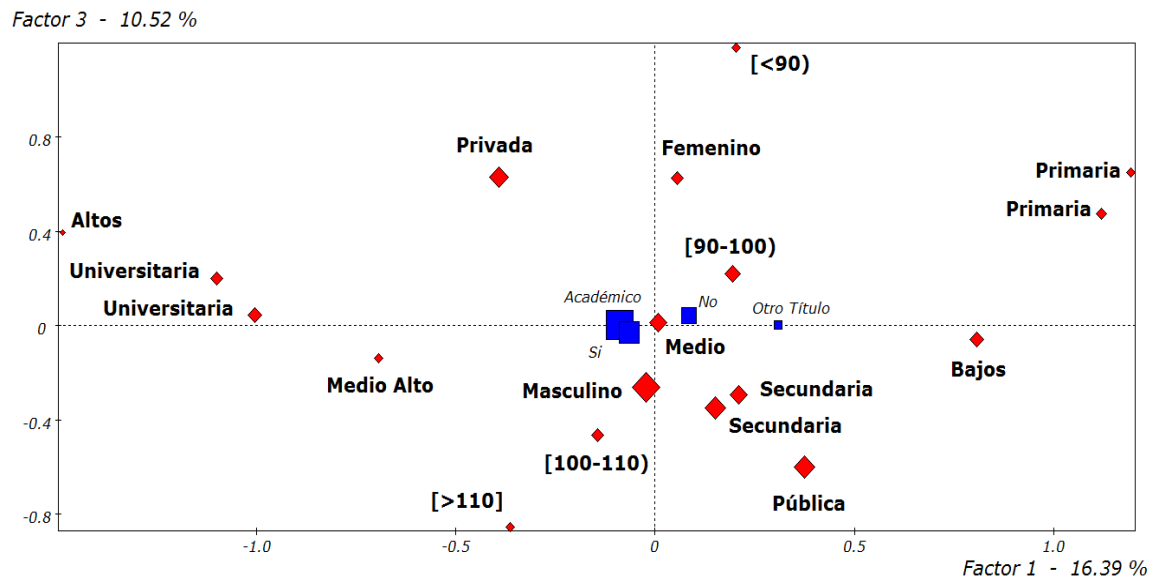


Al primer eje contribuyen de forma conjunta las modalidades de primaria, universitaria (tanto de la educación del padre como de la madre), estratos altos y estratos bajos explicando un 84,3% de su inercia total. Opone los estudiantes cuyos padres tienen formación universitaria y viven en estratos altos, de los estudiantes que tienen padres con formación primaria y viven en los estratos bajos.

Al segundo eje aportan las modalidades de primaria y secundaria (tanto en la educación del padre como de la madre), estrato medio, puntajes inferiores a 90 y puntajes superiores a 110 capturando el 84,6% de su inercia total. Contrapone a los estudiantes con puntajes en el extremo superior de la escala, que viven en estratos altos y tienen padres con formación académica de primaria, de los estudiantes que tienen puntajes en el extremo inferior de la escala, viven en estratos medios y sus padres tienen estudios de secundaria.

Las Figuras 2 y 3 muestran el plano factorial conformado por los ejes 1*3 y 1*4 respectivamente.

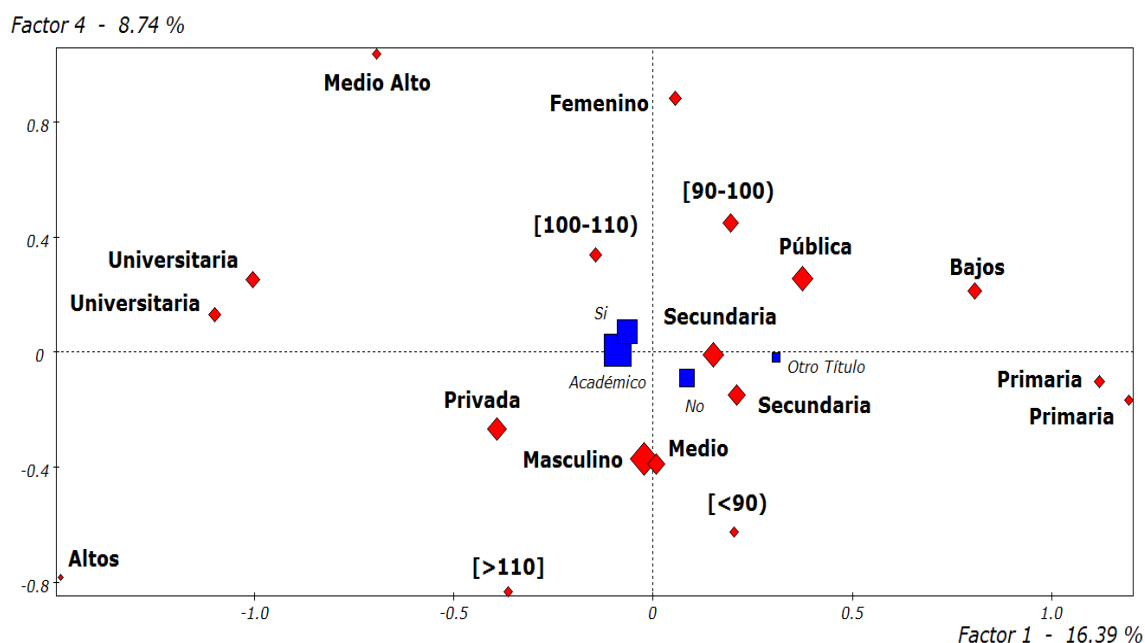
Figura 2. Plano factorial eje 1 * eje 3.



La mayor contribución a la inercia del eje tres (35,9%) la aportan las modalidades de la variable *puntaje*, seguida por las de *tipo* (30,1%), *género* (13%) y *educación_madre* (12,4%).

Esta dimensión separa a los estudiantes con puntajes [<90), del género femenino que estudian en instituciones privadas y tienen madres con educación primaria, de los estudiantes hombres con puntajes muy altos que cursan estudios en IES públicas y sus madres recibieron educación secundaria.

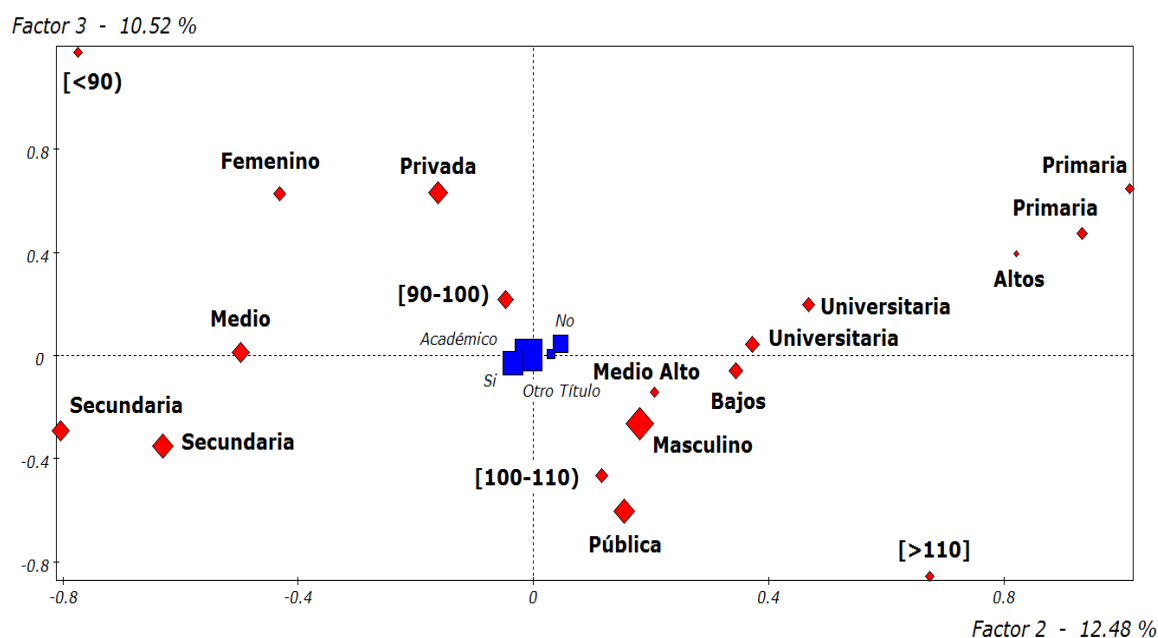
Figura 3. Plano factorial eje 1 * eje 4.



Al eje cuatro contribuyen en casi una tercera parte (32,1%) de su inercia las modalidades: femenino, estrato medio alto y puntaje [>110]. Este eje opone a los estudiantes del género femenino de estrato medio alto y puntajes en los niveles intermedios, de los estudiantes hombres pertenecientes a estratos altos y con puntajes mayores a 110.

La Figura 4 presenta el plano factorial con la combinación de los ejes dos y tres que acumulan un 23% de la inercia total. En ella se puede observar que se distinguen dos grupos: un grupo de estudiantes mujeres con puntaje bajo, que son de estrato medio y sus padres tienen estudios de secundaria, y el otro grupo, de estudiantes con puntajes superiores a 110, que son de estrato alto y sus padres tienen formación primaria.

Figura 4. Plano factorial: eje 2 * eje 3.



Analizadas las contribuciones de las categorías a los ejes, la inercia captada por los ejes, la calidad de las representaciones y la interpretación de los diferentes planos factoriales, del análisis de correspondencias se puede deducir que:

- Los puntajes superiores a la media se asocian más con estudiantes del sexo masculino y los inferiores con los de sexo femenino.
- Los resultados altos y muy altos favorecen más a los estudiantes de instituciones públicas que privadas.
- Los estudiantes que tienen padres con estudios universitarios se asocian más con los que obtienen puntajes altos y muy altos que los que tienen padres con menos estudios.

- Los estudiantes de los estratos medios y altos se asocian con los puntajes más altos.
- Los estudiantes que obtienen puntajes por encima de la media se pueden asociar con estudiantes del género masculino que viven en los estratos medios y altos, tienen padres con formación universitaria y estudian en IES de carácter público.

4.2 RESULTADOS DE LAS INSTITUCIONES

Ahora el interés se centra en observar que variables podrían influir en el puntaje promedio que logran los estudiantes de una institución. Se presenta a continuación la forma como fueron categorizadas las variables y seguidamente el ACS y ACM.

4.2.1. Niveles de las variables categóricas

Las variables que se tienen en cuenta en esta segunda parte del análisis son las que refieren a algunos aspectos de las IES tales como: la acreditación del programa, el carácter privado o público de la institución, el número de estudiantes que presentan la prueba, la ubicación o no en una ciudad principal del país y el puntaje promedio del grupo de estudiantes evaluados.

Al igual que para los resultados de los estudiantes, estas variables se categorizan para el análisis considerando no solo su distribución, sino también, la representatividad de cada una de las modalidades planteadas.

- ☐ *ACREDITACIÓN*¹²: Refiere si el programa ofrecido por la institución está acreditado. (1) No, (2) Si.
- ☐ *TIPO*: corresponde al tipo de Institución de Educación Superior según el origen de los recursos económicos que obtiene: privada (1), pública (2).
- ☐ *ESTUDIANTES*: Corresponde al número de estudiantes que presentan la prueba. (1) menos de 30 estudiantes, (2) entre 30 y 60 estudiantes y (3) más de 60 estudiantes.
- ☐ *CIUDAD*: Si la institución está ubicada en una de las siete ciudades principales del país (Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Cartagena y Pereira). (1) No y (2) Si.
- ☐ *PUNTAJE_PROM*: Puntuación promedio obtenida por el grupo de estudiantes que presentan la evaluación Saber Pro: (1) puntajes inferiores a 95 puntos, (2) puntajes entre 95 y 100 puntos, (3) puntajes entre 100 y 105 puntos y (4) los puntajes superiores a 105 puntos.

¹² Programas académicos que han demostrado altos niveles de calidad, razón por la cual el Ministerio de Educación Nacional, previo concepto del Consejo Nacional de Acreditación les otorga la correspondiente acreditación.

La Tabla 22 presenta a continuación el resumen de estas variables con su correspondiente categorización.

Tabla 22 . Variables categóricas y sus niveles para resultados de las instituciones.

VARIABLE	No.	VALOR	
ACREDITACIÓN	<i>1</i>	<i>NO</i>	
	<i>2</i>	<i>SI</i>	
TIPO	<i>1</i>	<i>PRIVADA</i>	
	<i>2</i>	<i>PÚBLICA</i>	
ESTUDIANTES	<i>1</i>	<i>[<30)</i>	
	<i>2</i>	<i>[30 - 60)</i>	
	<i>3</i>	<i>[>=60]</i>	
CIUDAD	<i>1</i>	<i>NO</i>	
	<i>2</i>	<i>SI</i>	
PUNTAJE	<i>1</i>	<i>[<90]</i>	<i>BAJO</i>
	<i>2</i>	<i>[90 - 100)</i>	<i>MEDIO</i>
	<i>3</i>	<i>[100 - 110)</i>	<i>ALTOS</i>
	<i>4</i>	<i>[>= 110]</i>	<i>MUY ALTOS</i>

4.2.2 Descripción de la variable PUNTAJE_PROM

Se pueden ver en la Tabla 23 los principales estadísticos de tendencia central y variabilidad para la variable *Puntaje_prom*, así como también

el histograma de frecuencias, diagramas de caja y pruebas de normalidad.

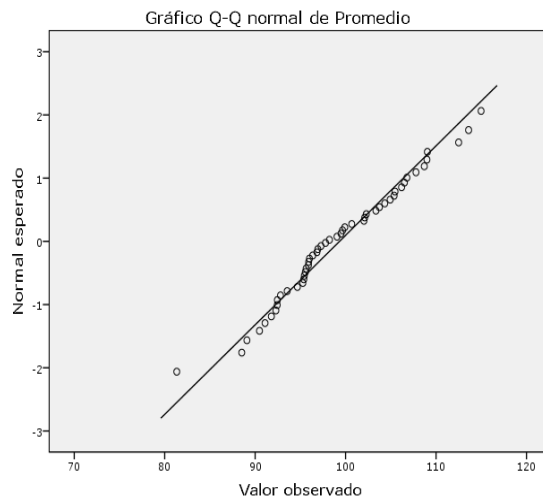
Tabla 23. Descriptivos, histogramas, diagramas de caja y pruebas de normalidad para la variable PUNTAJE_PROM.

ESTADÍSTICOS DESCRIPTIVOS			
		Estadístico	Error típico
Puntaje Promedio	Media	99,3284	0,99883
	Mediana	97,9950	
	Varianza	49,883	
	Desviación típica	7,06276	
	Mínimo	81,33	
	Máximo	114,98	
	Asimetría	0,148	0,337
	Curtosis	-0,127	0,662

Histograma

Diagrama de caja

PRUEBAS DE NORMALIDAD



Pruebas de normalidad

Kolmogorov-Smirnov ^a			Shapiro-Wilk		
Estadístico	gl	Sig.	Estadístico	gl	Sig.
Puntaje_prom	,094	50 ,200 [*]	,982	50	,624

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

La información que resume la tabla señala una distribución con media igual a 99,33, desviación 7 y un valor en la mediana (97,99) muy cerca a la media, la mitad de las instituciones (25) tuvieron una puntuación promedio superior a este valor, siendo el puntaje máximo alcanzado de 114,98 y el mínimo de 81,33.

En el histograma de frecuencias se advierte que casi el 25% de las IES tienen un *puntaje_prom* entre 95 y 97,5. El diagrama de cajas deja ver una distribución simétrica que corrobora el coeficiente de asimetría (0,148) con un valor muy cercano a cero.

Las pruebas de normalidad indican que no hay argumentos para rechazar la hipótesis que los datos se comportan como una normal y el gráfico Q-Q muestra una nube de puntos que se ajusta relativamente a la diagonal.

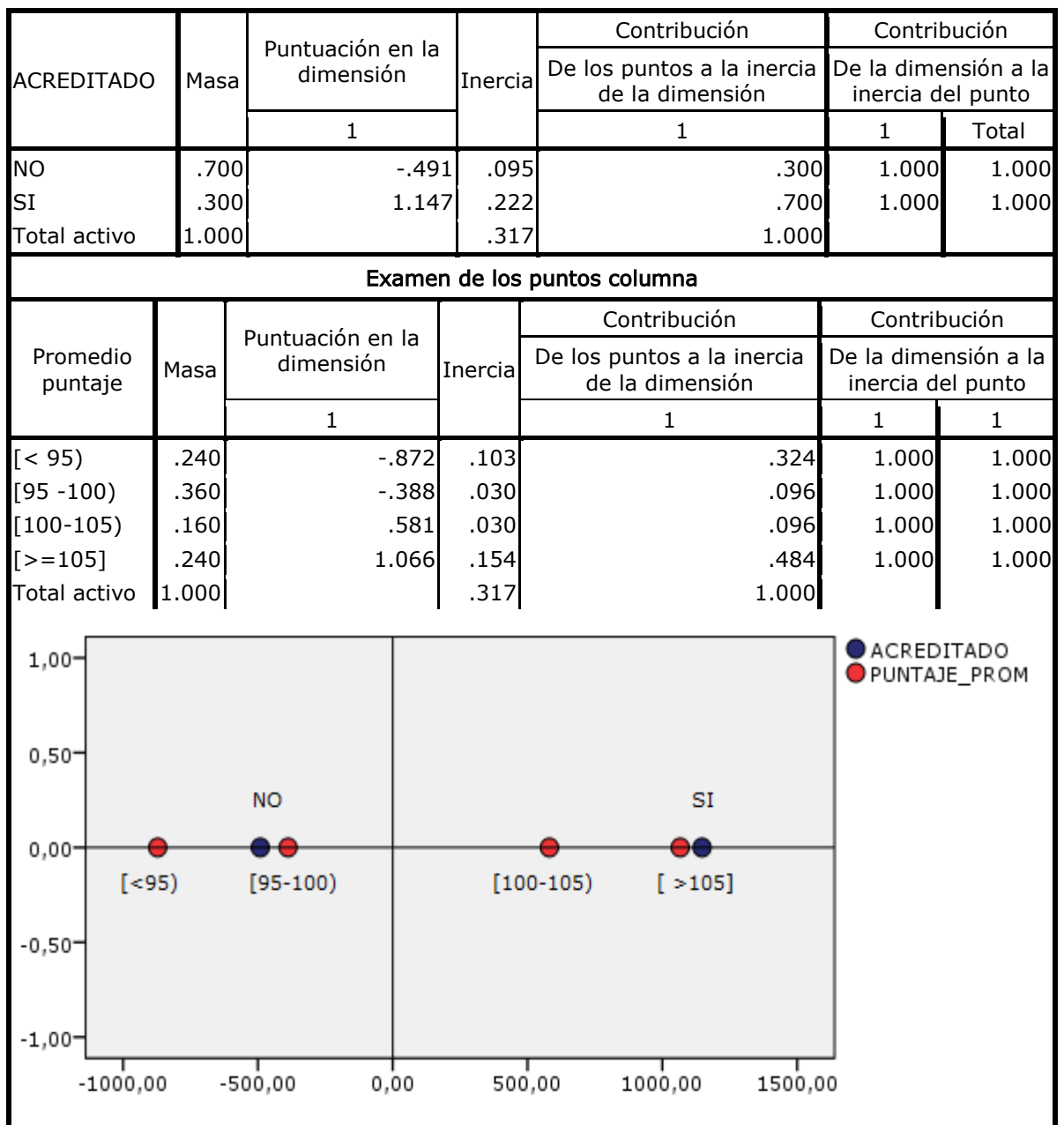
4.2.3 Análisis de Correspondencia Simple (ACS)

Las tablas unificadas que se presentan a continuación para cada una de las variables consideradas en esta sección en relación con la variable puntaje_prom, contienen la respectiva tabla de contingencia, los resultados de la prueba de independencia, como también los cuadros resúmenes de los exámenes de los puntos fila y los puntos columna y el respectivo diagrama con los valores de las puntuaciones en la dimensión.

□ ACREDITADO * PUNTAJE_PROM

*Tabla 24. Tabla unificada para la variable acreditado * puntaje_prom.*

Tabla de contingencia y perfiles fila							
			Promedio puntaje				Total
			[< 95]	[95 -100)	[100-105)	[>=105]	
ACREDITADO	NO	Recuento	12	15	4	4	35
		Perfil fila	34,3%	42,9%	11,4%	11,4%	100,0%
	SI	Recuento	0	3	4	8	15
		Perfil fila	,0%	20,0%	26,7%	53,3%	100,0%
Total		Recuento	12	18	8	12	50
		Perfil fila	24,0%	36,0%	16,0%	24,0%	100,0%
Resumen							
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el valor propio
					Explicada	Acumulada	Desviación típica
1	.563	.317			1.000	1.000	.107
Total		.317	15.873	.001	1.000	1.000	
Examen de los puntos de fila							



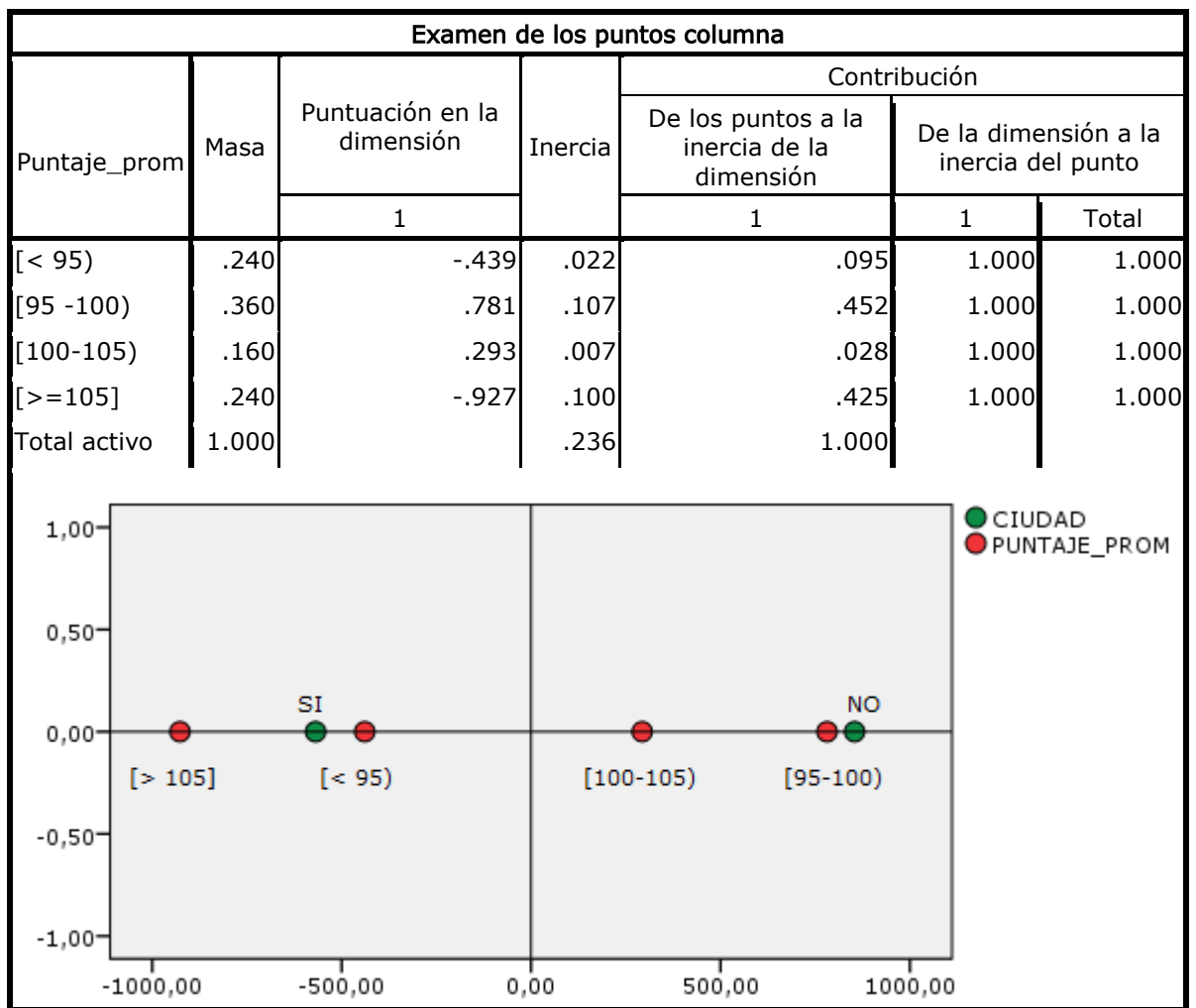
La prueba de independencia con el estadístico Ji-cuadrado avala con un valor de significancia de 0,001 la dependencia entre estas variables concluyendo que el carácter acreditado o no del programa de Ing. Civil tiene alguna influencia sobre el puntaje promedio que alcanzan las IES.

Las cuatro modalidades de *puntaje_prom* se ubican en orden ascendente de izquierda a derecha en la dimensión. Se puede decir que los puntajes altos y muy altos se asocian con las instituciones que sí tienen sus programas acreditados y por el contrario, los puntajes medios y bajos con las instituciones que no lo tienen.

CIUDAD * PUNTAJE_PROM

Tabla 25. Tabla unificada para la variable ciudad * puntaje_prom.

Tabla de contingencia y perfiles fila							
		Promedio puntaje				Total	
		[< 95)	[95 -100)	[100-105)	[>=105]		
CIUDAD	NO	Recuento	3	12	4	1	20
		Perfil fila	15,0%	60,0%	20,0%	5,0%	100,0%
	SI	Recuento	9	6	4	11	30
		Perfil fila	30,0%	20,0%	13,3%	36,7%	100,0%
Total		Recuento	12	18	8	12	50
		Perfil fila	24,0%	36,0%	16,0%	24,0%	100,0%
Resumen							
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio
					Explicada	Acumulada	Desviación típica
1	.486	.236			1.000	1.000	.114
Total		.236	11.806	.008	1.000	1.000	
Examen de los puntos de fila							
CIUDAD	Masa	Puntuación en la dimensión	Inercia	Contribución			
				De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto	
		1		1		1	1
NO	.400	.854	.142	.600		1.000	1.000
SI	.600	-.569	.094	.400		1.000	1.000
Total activo	1.000		.236	1.000			



Un valor de significancia de 0,008 en la prueba ji-cuadrado respalda la dependencia entre estas variables determinando que la ubicación o no de la institución en una de las siete ciudades principales de Colombia, afecta los resultados consolidados de sus estudiantes.

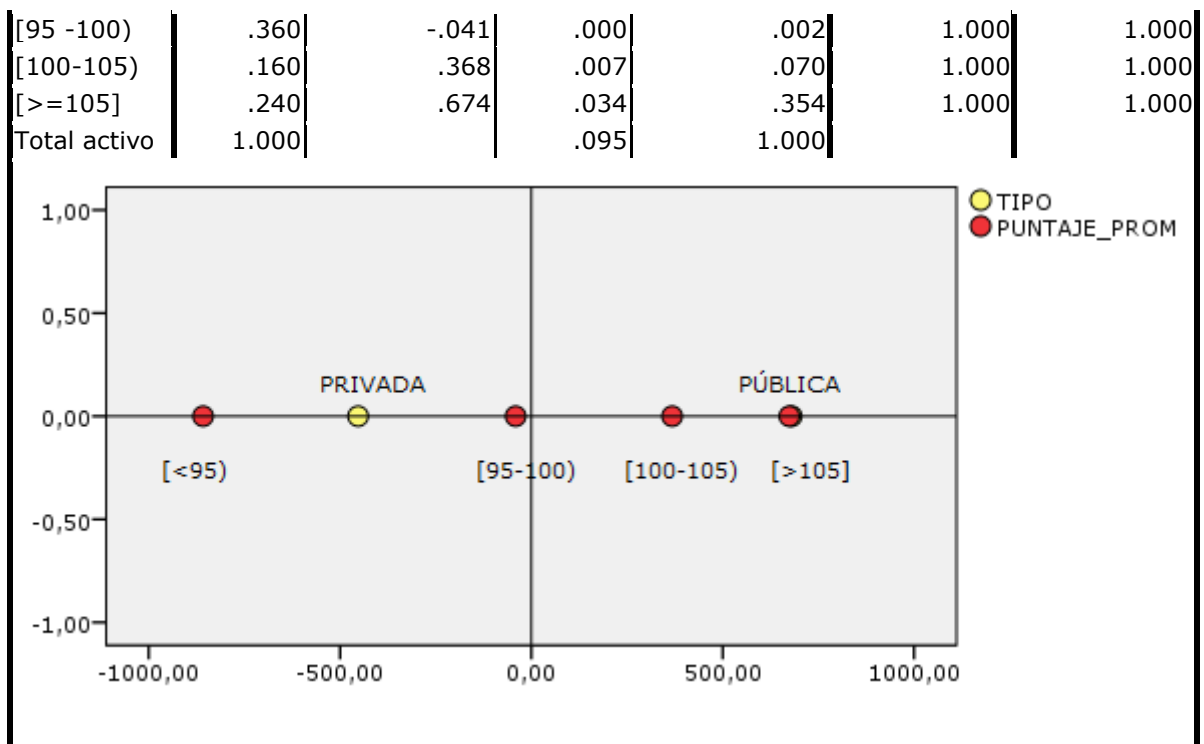
En el diagrama al final de la Tabla 25 se percibe que los resultados en las dos categorías extremas de la variable puntaje_prom se relacionan más con las instituciones que se encuentran en alguna de las ciudades principales y los puntajes mayores y menores en cinco puntos a la

media, se asocian con las instituciones que no están ubicadas en estas ciudades.

□ TIPO * PUNTAJE_PROM

Tabla 26. Tabla unificada para la variable tipo* puntaje_prom.

Tabla de contingencia y perfiles fila							
			Promedio puntaje				Total
			[< 95)	[95 -100)	[100-105)	[>=105]	
Tipo	PÚBLICA	Recuento	2	7	4	7	20
		Perfil fila	10,0%	35,0%	20,0%	35,0%	100,0%
	PRIVADA	Recuento	10	11	4	5	30
		Perfil fila	33,3%	36,7%	13,3%	16,7%	100,0%
Total		Recuento	12	18	8	12	50
		Perfil fila	24,0%	36,0%	16,0%	24,0%	100,0%
Resumen							
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio
					Explicada	Acumulada	Desviación típica
1	.308	.095			1.000	1.000	.124
Total		.095	4.745	.191	1.000	1.000	
Examen de los puntos de fila							
Tipo	Masa	Puntuación en la dimensión	Inercia	Contribución			
				De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto	
		1		1		1	Total
PÚBLICA	.400	.680	.057	.600		1.000	1.000
PRIVADA	.600	-.453	.038	.400		1.000	1.000
Total activo	1.000		.095	1.000			
Examen de los puntos columna							
Promedio puntaje	Masa	Puntuación en la dimensión	Inercia	Contribución			
				De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto	
		1		1		1	Total
[< 95)	.240	-.858	.054	.574		1.000	1.000



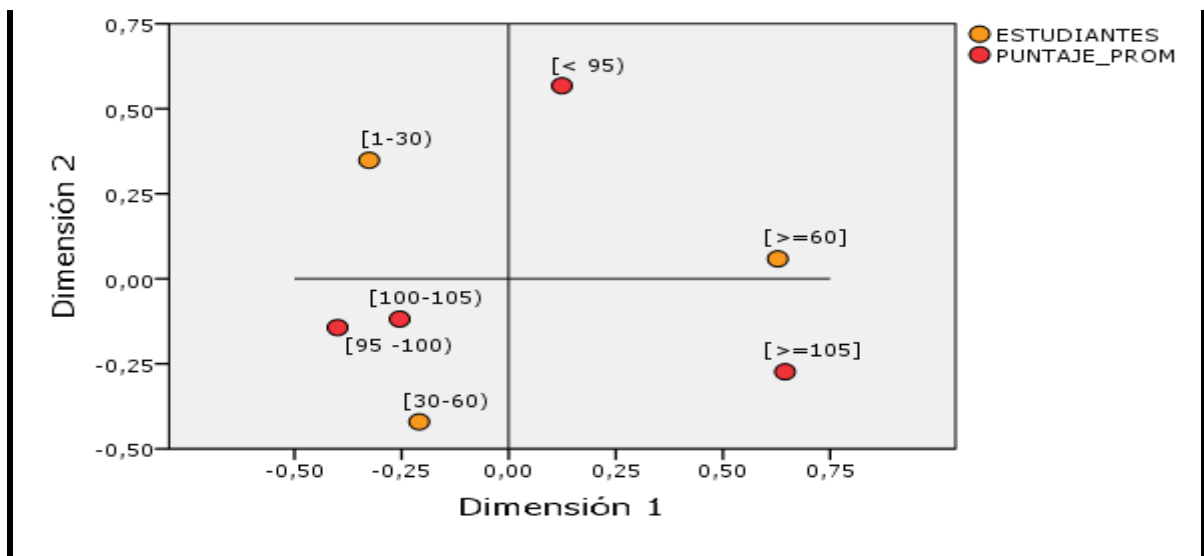
La prueba ji-cuadrado muestra con un valor_p igual a 0,191, que las variables tipo y puntaje_prom son independientes, señalando que el carácter privado o público de la IES no tiene influencia sobre los resultados del rendimiento académico de la institución.

□ ESTUDIANTES * PUNTAJE_PROM

Tabla 27. Tabla unificada para la variable estudiante * puntaje_prom.

Tabla de contingencia y de perfiles fila							
			Promedio puntaje				Total
			[< 95)	[95 -100)	[100-105)	[>=105]	
Estudiantes	[1-30)	Recuento	5	7	3	3	18
		Perfil fila	27,8%	38,9%	16,7%	16,7%	100,0%

	[30-60)	Recuento	3	7	3	4	17		
		Perfil fila	17,6%	41,2%	17,6%	23,5%	100,0%		
	[>=60]	Recuento	4	4	2	5	15		
		Perfil fila	26,7%	26,7%	13,3%	33,3%	100,0%		
Total		Recuento	12	18	8	12	50		
		Perfil fila	24,0%	36,0%	16,0%	24,0%	100,0%		
Resumen									
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio		
					Explicada	Acumulada	Desviación típica		
1	.171	.029			.727	.727	.141		
2	.105	.011			.273	1.000	.134		
Total		.040	2.018	.918 ^a	1.000	1.000			
Examen de los puntos de fila									
Estudiantes	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
		1	2		1	2	1	2	Total
[1-30)	.360	-.326	.349	.011	.223	.417	.588	.412	1.000
[30-60)	.340	-.209	-.421	.009	.087	.573	.287	.713	1.000
[>=60]	.300	.628	.059	.020	.690	.010	.995	.005	1.000
Total activo	1.000			.040	1.000	1.000			
Examen de los puntos columna									
Promedio puntaje	Masa	Puntuación en la dimensión		Inercia	Contribución				
					De los puntos a la inercia de la dimensión		De los puntos a la inercia de la dimensión		
		1	2		1	2	1	2	Total
[< 95)	.240	.124	.568	.009	.022	.737	.072	.928	1.000
[95 -100)	.360	-.400	-.143	.011	.335	.071	.927	.073	1.000
[100-105)	.160	-.254	-.118	.002	.060	.021	.883	.117	1.000
[>=105]	.240	.645	-.274	.019	.583	.171	.901	.099	1.000
Total activo	1.000			.040	1.000	1.000			



El test de ji-cuadrado corrobora la independencia entre estas variables indicando que el número de estudiantes que presentan la prueba no influye en el puntaje promedio de la institución.

4.2.4 Análisis de Correspondencias Múltiple (ACM)

El análisis de correspondencia simple permitió identificar las variables *acreditado* (programa acreditado) y *ciudad* (entre las siete ciudades principales) como las que mostraban dependencia con el puntaje promedio que obtenía la institución, por consiguiente para el ACM estas son las variables activas, y como suplementarias las variables *tipo* y *estudiantes*.

En la Tabla 28 se presentan los valores asumidos por los cinco valores propios y el peso que tiene cada uno en el total. La cuarta columna,

porcentaje acumulado, indica cuánta información se está representando al incorporar un eje factorial al análisis.

Tabla 28. Valores propios, coordenadas y contribuciones.

EIGENVALUES
COMPUTATIONS PRECISION SUMMARY : TRACE BEFORE DIAGONALISATION.. 1.6667
SUM OF EIGENVALUES..... 1.6667
HISTOGRAM OF THE FIRST 5 EIGENVALUES
+-----+-----+-----+-----+
NUMBER EIGENVALUE PERCENTAGE CUMULATED
PERCENTAGE
+-----+-----+-----+-----+
1 0.6111 36.67 36.67 *****
2 0.4176 25.06 61.73 *****
3 0.3333 20.00 81.73 *****
4 0.1788 10.73 92.45 *****
5 0.1258 7.55 100.00 *****
+-----+-----+-----+-----+
RESEARCH OF IRREGULARITIES (SECOND DIFFERENCES)
LOADINGS, CONTRIBUTIONS AND SQUARED COSINES OF ACTIVE CATEGORIES
AXES 1 TO 5
+-----+-----+-----+-----+
CATEGORIES LOADINGS CONTRIBUTIONS SQUARED COSINES
+-----+-----+-----+-----+
IDEN - LABEL REL. WT. DISTO 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
+-----+-----+-----+-----+
3 . ACREREDITADO
1 - No acred 23.33 0.43 0.52 -0.25 0.00 0.12 -0.27 10.5 3.6 0.0 2.0 14.0 0.64 0.15 0.00 0.04 0.18
2 - Si acred 10.00 2.33 -1.22 0.59 0.00 -0.29 0.64 24.4 8.4 0.0 4.6 32.6 0.64 0.15 0.00 0.04 0.18
+-----+-----+-----+-----+ CUMULATED CONTRIBUTION = 34.8 12.0 0.0 6.6 46.6 +-----+-----+
4 . CIUDAD
1 - No ppal 13.33 1.50 0.89 0.58 0.00 0.51 0.35 17.2 10.6 0.0 19.3 12.9 0.53 0.22 0.00 0.17 0.08
2 - Si ppal 20.00 0.67 -0.59 -0.38 0.00 -0.34 -0.23 11.5 7.0 0.0 12.9 8.6 0.53 0.22 0.00 0.17 0.08
+-----+-----+-----+-----+ CUMULATED CONTRIBUTION = 28.7 17.6 0.0 32.2 21.6 +-----+-----+
6 . PUNTAJE
1 - [<95] 8.00 3.17 0.36 -1.57 -0.48 0.01 0.58 1.7 47.2 5.6 0.0 21.5 0.04 0.78 0.07 0.00 0.11
2 - [95-100) 12.00 1.78 0.75 0.57 0.72 -0.61 -0.05 11.1 9.2 18.8 24.6 0.3 0.32 0.18 0.29 0.21 0.00
3 - [100-105) 5.33 5.25 -0.24 1.05 -1.99 -0.01 -0.39 0.5 14.0 63.1 0.0 6.4 0.01 0.21 0.75 0.00 0.03
4 - [>105] 8.00 3.17 -1.33 0.02 0.72 0.90 -0.24 23.1 0.0 12.5 36.6 3.7 0.56 0.00 0.16 0.26 0.02
+-----+-----+-----+-----+ CUMULATED CONTRIBUTION = 36.5 70.4100.0 61.2 31.9 +-----+-----+

Es así, como con los dos primeros valores propios se obtiene el primer plano factorial en el cual está representado el 61,73% del total de la inercia y con el tercer valor, explican el 81,73% de la variabilidad de los datos.

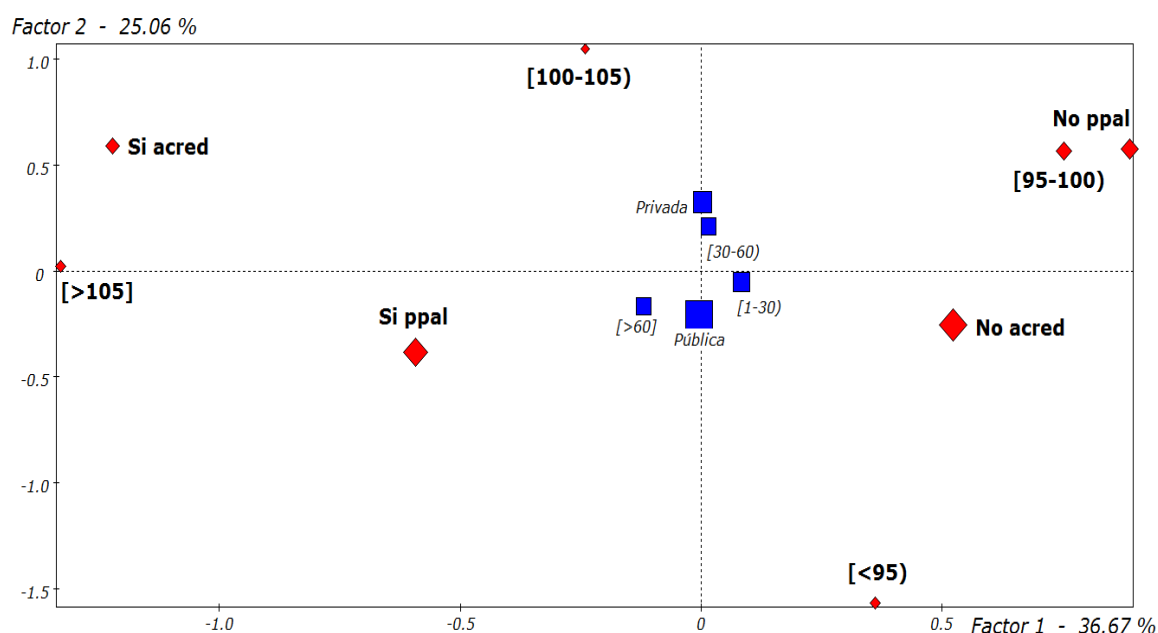
Al realizar el cociente de la inercia total y el total de valores propios, es decir, $1,667/5=0,3334$, y compararlo con el valor de los autovalores, se conservan los ejes que presenten sus autovalores por encima del promedio. En este caso, ello ocurre para los 3 primeros ejes factoriales, que serían en caso de ser interpretables, los que aporten información interesante respecto a los puntajes promedios de las Instituciones de Educación Superior.

Al primer eje contribuyen en mayor medida las modalidades de las variables puntaje y acreditado con un 36,5% y 34,8% respectivamente, seguida por las de la variable ciudad en un 28,7%.

La modalidad que mayor contribución hace a la inercia del segundo eje (47,2%) es puntajes [<95], y sumados con los puntajes entre [100-105) captan conjuntamente un 61,2% de la inercia.

La Figura 5 presenta el primer plano factorial conformado por los ejes 1 y 2.

Figura 5. Plano factorial: eje 1 * eje 2.

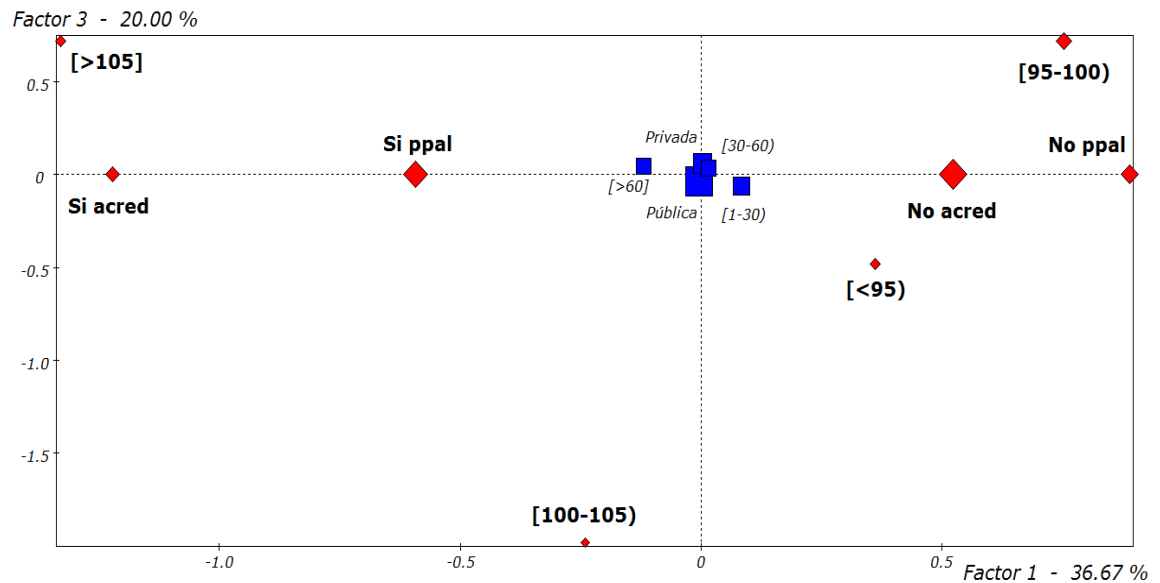


Se observa que la primera dimensión separa claramente las instituciones con puntajes promedio superiores a la media que tienen programas acreditados y están ubicadas en las primeras siete ciudades del país, de las que tienen puntajes promedios inferiores a la media, sus programas no están acreditados y no se encuentran ubicadas geográficamente en las principales ciudades.

La segunda dimensión opone principalmente las instituciones con puntajes promedio entre [100-105) de las que se encuentran en el menor nivel de la escala de puntaje_prom, [<95].

En la Figura 6 se aprecia el plano factorial formado por los ejes uno y tres.

Figura 6. Plano factorial: eje 1 * eje 3.



El tercer eje factorial recibe toda la contribución de la variable *puntaje_prom*. Siendo la modalidad [100-105) la que aporta la mayor parte con un 63,1%. Este eje opone las modalidades [>105] y [95-100) de las demás.

Realizado los respectivos análisis de correspondencia se puede deducir que:

- Las instituciones con los puntajes promedio inferiores a la media se relacionan con las IES que no tienen el programa de Ingeniería civil acreditado, y viceversa.

- Las instituciones ubicadas en los dos niveles intermedios de puntaje se asocian con las que no están ubicadas en alguna de las siete ciudades principales de Colombia.
- Las instituciones que en promedio tienen los puntajes más altos (sobre la media) se asocian con las que tienen el programa académico acreditado y están situadas en una de las siete ciudades principales de Colombia.
- Por el contrario, las instituciones que en promedio tienen puntajes por debajo de la media nacional, se relacionan con las que no tienen acreditado el programa y están ubicadas en alguna de las siete ciudades principales del país.

CONCLUSIONES

En este último capítulo se expone una síntesis de los principales resultados observados en la realización del presente estudio. En alguno de los apartados se presentan algunas comparaciones con los resultados del año 2009 que fue el de inicio de obligatoriedad de estas pruebas.

- Para el año 2010 el número de estudiantes de Ingeniería Civil que presentaron las pruebas Saber Pro a nivel nacional fue de 2263 que se corresponde con un aumento del 69,25% en comparación con los 1337 que fueron evaluados en el año inmediatamente anterior. Este aumento parece indicar que a pesar de que en el año 2009 la prueba ya era obligatoria, algunas facultades de Ingeniería Civil hicieron caso omiso de ella.
- El puntaje promedio que obtuvieron los estudiantes de este programa para el año 2010 fue de 100,06, completando el cuarto año consecutivo donde el programa alcanza únicamente el valor de la media normalizada que tiene el ICFES.
- Del total de estudiantes evaluados, los de instituciones públicas representan un poco más de la mitad (51,2%) disminuyendo este porcentaje en 5,5 puntos con respecto al año 2009 (56,7%). Sin embargo el puntaje promedio que alcanzaron (102,11) tuvo un pequeño aumento con respecto al 2009 que fue de 100,8. Un poco

mayor es el aumento que presentaron las instituciones privadas que obtuvieron 97,91 sobre los 93,0 del año anterior.

- Con respecto al género se conservan relativamente los mismos porcentajes que se han observado en el tiempo para este programa, siendo la mayor participación de estudiantes hombres con un 70,9% en el último año.
- El análisis de independencia de las variables consideradas para los resultados de los estudiantes señala que ni la situación laboral, ni el título de bachiller obtenido influyen en el puntaje que obtienen en la prueba.
- El análisis de correspondencias para los estudiantes permite concluir que los puntajes superiores a las media se asocian con estudiantes del género masculino, que estudian Ingeniería Civil en instituciones públicas, tienen padres con estudios universitarios y viven en los estratos sociales 4, 5 o 6.
- El análisis de independencia para los resultados de las instituciones indica que el número de estudiantes de la universidad que presentan la prueba y el carácter oficial o público la IES no se relacionan con el puntaje promedio que obtiene la institución.
- El análisis de correspondencias para las instituciones permite inducir que las IES con los puntajes promedio sobre la media se asocian con las que tienen el programa académico de Ingeniería

Civil acreditado y están situadas en una de las siete ciudades principales de Colombia.

- Indudablemente falta mucho análisis que realizar alrededor de las pruebas Saber Pro: el análisis de otras variables que puedan explicar los resultados y la implementación de otros modelos estadísticos que vayan más allá de lo meramente descriptivo, son algunos de ellos. Hacemos votos para que colegas de promociones futuras se animen a continuar con este tipo de investigación aplicada que, sin duda alguna, nos atrevemos a considerar de gran utilidad para el desarrollo de la educación superior en nuestro país.

BIBLIOGRAFÍA

Cárdenas Acevedo, L. J., & Rincón Delgado, Y. K. (2011). *Determinantes de la calidad de la educación pública universitaria en Colombia*. Bucaramanga, Colombia: Trabajo de grado de Economía. Universidad Industrial de Santander.

Componente de comprensión lectora guía de orientación. Recuperado el 1 de noviembre de 2011 de http://www.icfes.gov.co/index.php?option=com_docman&task=doc_view&gid=2348

Díaz Monroy, L. G. (2007). *Estadística multivariada: inferencia y métodos*. Bogotá, Colombia: Universidad Nacional de Colombia.

Díaz Monroy, L. G., & Morales Rivera, M. A. (2009). *Análisis estadístico de datos categóricos*. Bogotá, Colombia: Universidad Nacional de Colombia.

"Guía de acceso a base de datos ICFES". Recuperada el 1 de noviembre de 2011 de <http://www.icfes.gov.co/investigacion/index.php/bdicfes/acceso-a-bd-icfes>

Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Fundación BBVA.

ICFES. (Diciembre de 2011). *www.icfes.gov.co*. Recuperado el 30 de Enero de 2012

Marco De Fundamentación Conceptual Especificaciones De Prueba Ecaes Ingeniería Civil, Recuperado el 1 de noviembre de http://www.icfes.gov.co/index.php?option=com_docman&task=doc_view&gid=715

Peña, D. (2002). *Análisis de datos multivariantes*. Madrid, España: McGraw Hill.