

**ESTUDIO DE LA RELACIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES  
DE PRIMER SEMESTRE DE LAS INGENIERIAS EN LA UIS CON LOS RESULTADOS  
DE LAS PRUEBAS DE ESTADO A TRAVÉS DEL ANÁLISIS DE COMPONENTES  
PRINCIPALES Y TABLAS DE CONTINGENCIA**

**MIGUEL OSWALDO PÉREZ PULIDO  
HERNÁN ARTURO ROA FUENTES**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICAS  
SERVICIO SOCIAL EDUCATIVO Y TRABAJO DE GRADO II  
BUCARAMANGA  
2008**

**ESTUDIO DE LA RELACIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES  
DE PRIMER SEMESTRE DE LAS INGENIERIAS EN LA UIS CON LOS RESULTADOS  
DE LAS PRUEBAS DE ESTADO A TRAVÉS DEL ANÁLISIS DE COMPONENTES  
PRINCIPALES Y TABLAS DE CONTINGENCIA**

**Trabajo de grado para optar el título de  
Licenciado en Matemáticas**

**MIGUEL OSWALDO PÉREZ PULIDO  
HERNÁN ARTURO ROA FUENTES**

**DIRECTOR:  
GABRIEL YÁÑEZ CANAL  
P.H.D EN MATEMÁTICA EDUCATIVA**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICAS  
SERVICIO SOCIAL EDUCATIVO Y TRABAJO DE GRADO II  
BUCARAMANGA  
2008**

## **AGRADECIMIENTOS**

Agradecemos a Dios por la sabiduría que nos ha dado para alcanzar esta meta en nuestra realización personal

A la Universidad Industrial de Santander y a la Escuela de Matemáticas por aportar a nuestra formación profesional

Al Profesor Gabriel Yáñez Canal por su colaboración y orientación de este trabajo de grado

A la oficina de Admisiones y Registro académico de la Universidad Industrial de Santander por brindarnos su colaboración incondicional

A la licenciada Andrea González Villar, por su apoyo y disposición en la elaboración de este proyecto

A todas aquellas personas que influyeron en la elaboración de esta investigación.

## ÍNDICE GENERAL

	Pag
Presentación.....	1
<b>CAPÍTULO I: ACERCA DEL PROCESO DE ADMISIÓN EN LA UIS Y EL EXAMEN DE ESTADO .....</b>	<b>3</b>
1.1 Propósitos del Examen de Estado .....	4
1.2 Cómo se elaboran las pruebas de Estado.....	4
1.3 Tipos de preguntas del Examen.....	5
1.3.1 Preguntas de selección múltiple con única respuesta tipo I.....	5
1.3.2 Preguntas de selección múltiple con múltiple respuesta tipo IV....	6
1.3.3 Preguntas de selección múltiple con única respuesta tipo X.....	7
1.4 Tipos de resultados y su interpretación.....	9
1.4.1 Puntaje.....	10
1.4.2 Nivel de competencia.....	11
1.4.3 Desempeño por grupos de preguntas.....	11
1.4.4 Grado de profundización.....	11
1.4.5 Puesto .....	12
1.5 Política de Admisión a la UIS.....	12
1.6 Criterio de Admisión.....	13
1.7 Asignación de cupos por carrera.....	15
1.8 Puntaje de corte de los últimos procesos de Admisión.....	15
<b>CAPÍTULO II: MARCO TEÓRICO Y CONCEPTUAL.....</b>	<b>17</b>
2.1 Análisis de Componentes Principales.....	18
2.1.2 Cálculo de las Componentes Principales.....	21
2.1.2.1 Cálculo de la primer componente.....	21
2.1.2.2 Cálculo de la segunda Componente.....	24

2.1.2.3 Matriz de varianza y covarianza.....	27
2.1.2.4 Porcentajes de variabilidad.....	28
2.1.2.5 Representación gráfica.....	30
2.2 Tablas de contingencia.....	31
2.2.1 Estadístico Chi-Cuadrado.....	33
2.2.2 Estimaciones de riesgo.....	35
2.2.3 Intervalos de confianza para la estimación de riesgo.....	37
2.3 Análisis de Componentes Principales con SPSS.....	40
2.3.1 Presentación de los datos.....	40
2.3.2 Pasos a seguir a través de SPSS .....	43
2.3.2.1 Para Componentes Principales.....	43
2.3.2.2 Para tablas de contingencia y estimaciones de riesgo.....	53
<b>CAPITULO III: RESULTADOS Y ANÁLISIS DE LA INVESTIGACIÓN.....</b>	<b>57</b>
3.1 Análisis de Componentes Principales.....	58
3.2 Análisis con Tablas de Contingencia.....	71
3.2.1 Estudio de la asociación entre género y resultados en las notas de cálculo I con los niveles de “aprobó” y “no aprobó”.....	71
3.2.1.1 Resultados .....	72
3.2.1.2 Estimaciones de riesgo.....	76
3.2.2 Estudio de la asociación entre puntaje de Icfes en Matemáticas y resultados en las notas de cálculo I con los niveles de “aprobó” y “no aprobó”.....	77
3.2.1.1 Resultados.....	78
3.2.1.2 Estimaciones de Riesgo.....	82
3.2.3 Estudio de la asociación entre puntaje de Icfes en Lenguaje y Resultados en Cálculo I con los niveles de aprobó y no aprobó.....	83
3.2.3.1 Resultados .....	84
3.2.4 Estudio de la asociación entre puntaje de Icfes en Lenguaje y	87

Resultados en Química Básica con los niveles de aprobó y no aprobó	
3.2.4.1 Resultados	88
3.2.5 Estudio de la asociación entre la categoría del colegio de procedencia es decir “privado” o “público” y resultados en Cálculo I con las categorías de “aprobó” y “no aprobó” .....	91
3.2.5.1 Resultados .....	93
3.2.6 Estudio entre la categoría del colegio de procedencia es decir “público” o “privado” y resultados en Álgebra Lineal con las categorías de “aprobó” y “no aprobó” .....	95
3.2.6.1 Resultados .....	96
Conclusiones y Recomendaciones.....	100
Referencias Bibliográficas.....	103

## ÍNDICE DE FIGURAS

	Pag
Figura 1. Ejemplo pregunta de selección múltiple con única respuesta tipo I. ....	6
Figura 2. Ejemplo pregunta de selección múltiple con múltiple respuesta tipo IV.....	7
Figura 3. Gráfico de la pregunta de selección múltiple con múltiple respuesta tipo X .....	9
Figura 4. Formato de interpretación de resultados de las Pruebas de Estado .....	10
Figura 5. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella.....	20
Figura 6. Diagrama de Dispersión.....	30
Figura 7. Distribución $\chi^2$ .....	35
Figura 8. Gráfico de sedimentación.....	48
Figura 9. Plano Factorial.....	49
Figura 10. Diagrama de dispersión.....	53
Figura 11. Histograma con los estadísticos descriptivos de las variables en estudio .....	58
Figura 12. Frecuencias de los individuos en estudio por género.....	61
Figura 13. Frecuencia de los individuos en estudio por carreras.....	61
Figura 14. Frecuencia de los individuos en estudio por edades.....	62
Figura 15. Diagrama de sedimentación .....	66
Figura 16. Gráfico de dos componentes.....	68
Figura 17. Diagrama de Dispersión Factor 1 vs. Factor 2.....	70
Figura 18. Diagrama de frecuencias de género y resultados en cálculo I	74
Figura 19. Diagrama de frecuencias de ICFES en Matemáticas y	80

resultados en Cálculo I .....	
Figura 20. Diagrama de frecuencias de ICFES en lenguaje y resultados en Cálculo I .....	85
Figura 21. Diagrama de frecuencias de Icfes en lenguaje y Resultados en Química .....	90
Figura 22. Diagrama de frecuencias de tipo de colegio de procedencia con el resultado en Cálculo I .....	94
Figura 23. Diagrama de frecuencias de tipo de colegio con el resultado en Álgebra Lineal .....	97

## ÍNDICE DE TABLAS

	Pag
Tabla 1. Criterios de Admisión UIS (Fuente: Oficina de Admisión y registro académico UIS).....	14
Tabla 2. Cupos para Admitidos por carrera (Fuente: Oficina de Admisión y registro académico UIS).....	15
Tabla 3. Puntaje de cortes de los últimos dos años.....	16
Tabla 4. Ilustración de las Componentes Principales a partir de las variables originales .....	18
Tabla 5. Tabla de contingencia de dos cursos contra el nivel de aprobación o no aprobación de indicadores.....	31
Tabla 6. Tabla de contingencia de los cursos 9-10 y 9-12 con las frecuencias observadas y frecuencias esperadas .....	32
Tabla 7. Tabla de contingencia de los cursos 9-10 y 9-12 contra estudiantes que aprobaron todos los indicadores.....	37
Tabla 8. Estimación de riesgo .....	39
Tabla 9. Medidas corporales de gorriones (Fuente: Manly (1996, p.2).....	41
Tabla 10. Cálculos de la diagonal de matriz de varianzas de los pájaros.....	43
Tabla 11. Matriz de Correlación.....	46
Tabla 12. Matriz de Comunalidades.....	46
Tabla 13. Varianza total explicada .....	47
Tabla 14. Matriz de Componentes.....	49
Tabla 15. Matriz para el cálculo de las puntuaciones factoriales.....	50
Tabla 16. Coordenadas del diagrama de dispersión .....	52
Tabla 17. Tabla de contingencia.....	55
Tabla 18. Prueba de Chi-Cuadrado .....	56
Tabla 19. Estimaciones de riesgo.....	56

Tabla 20. Estadísticos descriptivos de las variables.....	60
Tabla 21. Prueba de Bartlett y KMO .....	63
Tabla 22. Matriz de correlaciones .....	64
Tabla 23. Matriz de Comunalidades.....	65
Tabla 24. Varianza explicada.....	65
Tabla 25. Matriz de Componentes.....	67
Tabla 26. Matriz de componentes para el cálculo de las puntuaciones en las componentes .....	69
Tabla 27. Ficha Técnica.....	72
Tabla 28. Asociación entre género y notas de cálculo 1 (Fuente: Oficina de Admisión y registro académico UIS).....	73
Tabla 29. Prueba de Chi- Cuadrado .....	75
Tabla 30. Estimación de riesgo de perder cálculo I entre los dos géneros.	76
Tabla 31. Ficha Técnica.....	78
Tabla 32. Clasificación de los estudiantes según le puntaje del Icfes en Matemáticas y la nota de Cálculo I .....	79
Tabla 33. Prueba de Chi- Cuadrado .....	81
Tabla 34. Estimación de riesgo de perder cálculo I entre los dos géneros...	82
Tabla 35. Ficha Técnica.....	84
Tabla 36. Clasificación de los estudiantes según le puntaje del Icfes en Matemáticas y la nota de Cálculo I .....	84
Tabla 37. Prueba de Chi- Cuadrado .....	86
Tabla 38. Ficha Técnica.....	88
Tabla 39. Clasificación de los estudiantes según el puntaje del Icfes en Lenguaje y la nota de Química .....	88
Tabla 40. Prueba de Chi- Cuadrado .....	91
Tabla 41. Ficha Técnica.....	92
Tabla 42. Clasificación de los estudiantes según colegio de procedencia con resultados en Cálculo I .....	93

Tabla 43. Prueba de Chi- Cuadrado .....	95
Tabla 44. Ficha Técnica.....	96
Tabla 45. Clasificación de los estudiantes según colegio de procedencia con resultados en Álgebra Lineal .....	96
Tabla 46. Estadístico Chi-Cuadrado.....	98

## RESUMEN

**TÍTULO:** Estudio de la relación del rendimiento académico de los estudiantes de primer semestre de las ingenierías en la UIS con los resultados de las pruebas de estado a través del análisis de componentes principales y tablas de contingencia\*

**AUTORES:** Miguel Oswaldo Pérez Pulido\*\*  
Hernán Arturo Roa Fuentes

**PALABRAS CLAVES:** Componentes Principales  
Estimaciones de Riesgo  
Tablas de Contingencia

### DESCRIPCIÓN:

La presente investigación, se centró en buscar la relación entre el puntaje del Icfes en las áreas de Matemáticas, Física y Lenguaje, con el rendimiento en asignaturas del ciclo básico de los ingenieros, tal como Cálculo I, Álgebra Lineal, Química I en la Universidad Industrial de Santander. Se concluyó que la relación entre puntaje Icfes y el rendimiento en la Universidad es totalmente independiente. Se realizó con el uso de la técnica estadística multivariada “Análisis de Componentes Principales” y efectuando contrastes con variables categóricas con análisis de estadística inferencial, usando tablas de contingencia.

Además se logró identificar algunas características que hacen de un estudiante un sujeto con “riesgo” de no cumplir satisfactoriamente con los resultados mínimos exigidos por la Universidad en las carreras de ingenierías y proponer posibles alternativas para aumentar sus posibilidades de alcanzar un mejor rendimiento académico. Esta investigación es apoyada con el software SPSS (*Statistical Package For Social Sciences, versión 12.0*).

La investigación se desarrollo con datos de estudiantes que ingresaron a la Universidad desde el año 2004 a 2006 a los diferentes programas de ingenierías de la Universidad Industrial de Santander. Estos fueron suministrados por la oficina de admisión y registro académicos de la UIS.

---

\* Trabajo de grado

\*\* Facultad de Ciencias. Licenciatura en Matemáticas  
Director Gabriel Yáñez Canal

## SUMMARY

**TITLE:** Study about the academic performance of the first level students in the engineering careers in the UIS and its relation with the state test results, through the analysis of the principal components and the contingency tables.

**AUTHORS:** Miguel Oswaldo Pérez Pulido.  
Hernán Arturo Roa Fuentes.

**KEY WORDS:** Principal components  
Risk estimation  
Contingency tables

### DESCRIPTION:

This research pretends to find the relation between the ICFES results in the Math, Physics and languages areas and the academic performance in the subjects of the basic cycle for the engineering careers like: calculus 1, linear algebra, chemistry 1, in the Universidad Industrial de Santander. It concluded that this relation is totally independent. And it was made with the multivariable statistics technique "Analysis of the Principal Components" and also in contrast with categorical variables, through the analysis of the inferential statistics using the contingency tables.

Moreover, it was possible to identify some traits which have an influence in the making of students not attain the minimum qualifications required by the university in the engineering programs. Also, we intend to propose new alternatives in order to increase the personal possibilities to have a better academic performance. This research was supported by SPSS software (Statistical Package For Social Sciences, version 12.0)

This work has been developed with the information of the students who have been admitted by university from 2004 to 2006, in the different UIS engineering careers. This information was provided by the UIS Admission and Academic Registry Office of UIS.

---

\* Degree work

\*\* Science Faculty. Mathematics Degree

Director Gabriel Yáñez Canal

## **PRESENTACIÓN**

Para ingresar a la Universidad Industrial de Santander se tiene en cuenta las pruebas de estado como el único referente para la admisión a los diferentes programas académicos ofrecidos. Los criterios de selección son relativos a cada carrera, siendo los puntajes de mayor exigencia los obtenidos en Matemáticas, Física y Lenguaje en las carreras de ingenierías.

A pesar que el mecanismo utilizado pretende seleccionar los estudiantes más capacitados para cada programa académico, es preocupante el índice de repitencia y deserción académica tan alto que existe en las materias del ciclo básico requeridos en las ingenierías, especialmente en los Cálculos, Químicas y Físicas. Se espera que los estudiantes admitidos tengan un buen rendimiento en la inducción de las carreras, debido a que en las pruebas de estado se evaluaron distintos temas básicos para el buen desempeño académico.

Una forma de examinar el mecanismo de admisión y su intención de aceptar aquellos candidatos más capacitados para responder a la exigencia académica de las carreras de ingenierías es realizando un análisis exploratorio que permita buscar la relación entre los resultados del ICFES y el desempeño académico en las materias del primer semestre de ingenierías.

Para este fin realizaremos una investigación cuantitativa apoyada en la técnica estadística multivariada Análisis de Componentes Principales y tablas de contingencia, ya que esta nos va a permitir encontrar qué relaciones existen entre las variables tanto numéricas como categóricas (Notas finales primer semestre del ciclo básico, los resultados pruebas de estado).

El trabajo sigue un esquema general iniciando en el capítulo 1 donde muestra la información necesaria acerca del examen de estado y el proceso de admisión en la Universidad Industrial de Santander. El capítulo II contempla el marco teórico y conceptual en el que se describen los procesos estadísticos a usar.

En el capítulo III se incluyen los resultados del análisis de la información obtenida mediante la técnica multivariada “Análisis de Componente Principales” y de tablas de contingencia con el cálculo de riesgo en los casos que sean necesarios.

La investigación finaliza con las conclusiones y referencias bibliográficas, seguida de algunos anexos.

**CAPÍTULO I: ACERCA DEL PROCESO DE ADMISIÓN EN LA UIS Y EL EXAMEN  
DE ESTADO**

## **1.1 PROPÓSITOS DEL EXAMEN DE ESTADO**

Dentro de los propósitos del examen de Estado para la educación colombiana en todos sus estamentos, tenemos los siguientes:

- Servir como un criterio para el Ingreso a la Educación Superior.
- Informar a los estudiantes acerca de sus competencias en cada una de las áreas evaluadas, con el ánimo de aportar elementos para la orientación de su opción profesional.
- Apoyar los procesos de auto evaluación y mejoramiento permanente de las instituciones escolares.
- Constituirse en base e instrumento para el desarrollo de investigaciones y estudios de carácter cultural, social y educativo.
- Servir de criterio para otorgar beneficios educativos.

## **1.2 CÓMO SE ELABORAN LAS PRUEBAS DE ESTADO**

Según Gaona y Pardo (1998), las pruebas desarrolladas para el Examen de Estado cumplen con una serie de condiciones que garantizan su calidad desde los puntos de vista académico y técnico. En términos generales, el cumplimiento de estas condiciones se puede apreciar en el proceso de desarrollo de pruebas, el cual implica llevar a cabo varias etapas y tomar decisiones sobre aspectos teóricos, metodológicos y prácticos.

El proceso se inicia con la conformación de equipos coordinados por profesionales del ICFES, en los cuales participan investigadores con experiencia y conocimiento en educación, en evaluación y en los campos en los que se desarrollan las pruebas, además de docentes de diferentes niveles educativos (básico, medio, universitario).

Los equipos inician una serie de discusiones tendientes a la fundamentación conceptual de la evaluación en general y de cada prueba, a la vez que se busca dar sentido y significado a todo el proceso y a los resultados que se obtendrán. Para ello se tienen en cuenta diferentes perspectivas sobre educación, evaluación, conocimiento, experiencias nacionales e internacionales en evaluaciones similares, enfoques disciplinares y pedagógicos, y teorías socioculturales, así como reglamentación educativa vigente en otros aspectos.

A partir de estas discusiones se definen los fundamentos generales de las pruebas y la estructura del examen, en los que se contemplan: la definición de los aspectos a evaluar, el número y formato de las preguntas y el tiempo aproximado para responder la prueba. También se establece un modelo psicométrico para el procesamiento y la interpretación de resultados de acuerdo con los propósitos de la evaluación y con las características de la población a evaluar. A partir de estos parámetros se procede a la construcción de los instrumentos y a la experimentación de los mismos.

Una vez experimentadas las pruebas, se procesan los resultados y, con base en ello, se realizan los ajustes pertinentes según los criterios académicos y técnicos, y finalmente se conforman las pruebas definitivas.

### **1.3 TIPOS DE PREGUNTAS DEL EXAMEN**

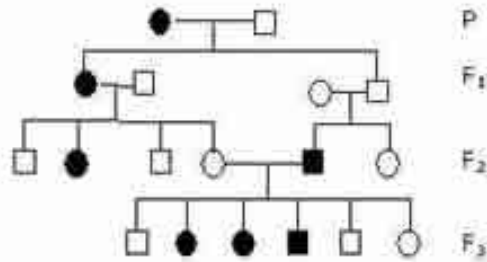
#### **1.3.1 Preguntas de selección múltiple con única respuesta Tipo I.**

Según Gaona y Pardo (1998), estas preguntas utilizadas en todas las pruebas, se desarrollan en torno a una idea o a un problema y constan de un enunciado y cuatro opciones de respuesta. Se recomienda leer cuidadosamente el enunciado y, después de analizarlo, escoger entre las opciones la que se considera correcta.

## EJEMPLO

En la figura, del cruce entre individuos de la  $F_2$  resultan tres hijos enfermos ● ó □ y tres normales. Si se asume que la enfermedad representada es de carácter recesivo, entonces los genotipos de los padres deben ser: (A = dominante y a = recesivo)

- A.  $Aa \times AA$
- B.  $Aa \times Aa$
- C.  $Aa \times aa$
- D.  $Aa \times AA$



La respuesta correcta es C y así debe usted marcarla en su Hoja de Respuestas:



Figura 1. Ejemplo pregunta de selección múltiple con única respuesta tipo I

### 1.3.2 Preguntas de selección múltiple con múltiple respuesta Tipo IV.

Estas preguntas, utilizadas en las pruebas de historia, geografía, filosofía, violencia y sociedad, constan de un enunciado, cuatro posibles consecuencias, aplicaciones o condiciones relacionadas con el enunciado (numeradas de 1 a 4) y cuatro opciones de respuesta (A, B, C, D), resultantes de cuatro posibles combinaciones así:

Si 1 y 2 son correctas, llene el óvalo A

Si 2 y 3 son correctas, llene el óvalo B

Si 3 y 4 son correctas, llene el óvalo C

Si 2 y 4 son correctas, llene el óvalo D

## EJEMPLO

Para Durkheim, la sociología no deber ser una filosofía de la historia que pretenda descubrir las leyes generales que guían la marcha del progreso de la humanidad, su tarea consiste en tomar los hechos sociales como objetos de investigación y, tratándolos como cosas, buscar la causa determinante de esos hechos entre los demás hechos sociales y no entre los hechos de la conciencia individual. Así, la sociología se convierte en una ciencia autónoma porque

1. consigue separarse de toda ciencia existente y abrirse un nuevo rumbo
2. encuentra que su objeto particular y concreto es el hecho social
3. descubre que el desarrollo de la sociedad depende del proceso histórico del hombre
4. rompe el esquema universalista de la filosofía en el estudio de los fenómenos sociales

Las opciones 2 y 4 son correctas por lo cual debe marcar en su Hoja de Respuestas:



Figura 2. Ejemplo pregunta de selección múltiple con múltiple respuesta tipo IV

### 1.3.3 Preguntas de selección múltiple con múltiple respuesta válida. Tipo X.

Estas preguntas utilizadas en las pruebas de matemáticas (núcleo común) y lenguaje (profundización), constan de:

- Una situación, que puede ser una gráfica, una tabla, un texto o una combinación de ellos.

- Un enunciado problema, que puede estar dado en forma afirmativa o interrogativa.
- Cuatro opciones de respuesta.

Recuerde que puede encontrar varias opciones válidas para solucionar el enunciado problema; usted debe seleccionar entre estas opciones sólo una: la que considere da respuesta manera más precisa o estructurada, a las condiciones particulares de la situación en el contexto evaluado.

#### EJEMPLO

Se pide realizar la gráfica (Figura 1) de la función ¿Es correcta la gráfica?

$$f(x) = -x^2 + 3x - 2$$

- A. si, porque el punto (2,0) pertenece a la gráfica y  $f(2) = 0$
- B. no, porque la gráfica no corresponde a una parábola y  $f(x)$  sí
- C. si, porque el rango de la gráfica y la función tienen valores reales positivos
- D. no, porque el signo de la mayor potencia de  $f(x)$  es negativo, luego su representación gráfica debe ir hacia abajo

En este ejemplo, las opciones B y D dan respuesta al problema, ya que ambas dan razones por las cuales la gráfica dibujada no corresponde a la función planteada en el enunciado; sin embargo, usted debe seleccionar sólo una, la opción que considere relaciona de manera más estructurada el concepto matemático, en este caso, el concepto de función cuadrática con las condiciones particulares de la situación. La clave es la opción D, dado que en ésta se realiza un análisis más detallado de las características de la función cuadrática  $f(x) = -x^2 + 3x - 2$  en relación con la gráfica; por lo tanto, usted debe rellenar en su hoja de respuestas el óvalo correspondiente a la letra D.

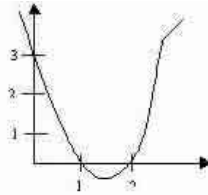



Figura 3. Gráfico de la pregunta de selección múltiple con múltiple respuesta tipo X

#### 1.4. TIPOS DE RESULTADOS Y SU INTERPRETACIÓN

Los resultados del examen de estado para ingreso a la educación superior tienen el propósito de brindar información completa sobre el desempeño de los estudiantes frente al examen, para contribuir con el proceso de reconocimiento del perfil de potencialidades que cada uno de ellos posee. Son 5 tipos de resultados que se resumen en distintas tablas para cada institución educativa. Se espera que esta información resulte útil como complemento de los procesos de autoevaluación y seguimiento periódico que se realizan en cada colegio. Los 5 tipos de resultados son los siguientes:

 <b>EXAMEN DE ESTADO</b> <small>PARA INGRESO A LA EDUCACION SUPERIOR</small>		<b>INFORME DE RESULTADOS</b>									30
<b>NUMERO DE REGISTRO</b> AC200123254471			<b>CIUDAD DEL COLEGIO</b> BUCARAMANGA			<b>CODIGO COLEGIO</b> 092619					
<b>NOMBRE</b> RUEDAD GARCIA ANTHONY JOB						<b>DOCUMENTO DE IDENTIDAD</b> T.I. 83121851540					
<b>FECHA DE EXAMEN</b> 18 Y 19 DE AGOSTO DE 2001											
<small>Para mayor información consulte el Marco de Interpretación de Resultados en la página web: <a href="http://www.icfes.gov.co">www.icfes.gov.co</a> o en el documento enviado a su colegio</small>											
<b>NUMEROS</b>	<b>PRUEBAS</b>	BIOLOGIA	MATEMATICAS	FILOSOFIA	FISICA	HISTORIA	QUIMICA	LENGUAJE	GEOGRAFIA	INGLÉS	
	<b>PUNTAJE</b>	44	38	42	46	41	42	46	44	31	
	<b>GRUPOS DE PREGUNTAS</b>	G1	SB	A	SB	SB	SA	SB	B	SB	SA
		G2	SA	M	SB	SB	SB	SA	SA	B	SA
		G3	A	SB	SA	SA	M	SA	SB	SB	SA
		G4		SA	SA	SA	SB	SB		SA	SB
		G5			SA					SA	SB
	<b>NIVEL DE COMPETENCIA</b>	C1	B	B	B	A	B	A	B	B	B
		C2	B	A	B	B	B	B	B	B	A
		C3	B	B	A	B	B	B	A	B	A
<b>COMPLETOS</b>	<b>GRADO DE PROFUNDIZACION</b>	GB				GB		I			
	<b>PUNTAJE INTERDISCIPLINAR</b>	MEDIO AMBIENTE: 57									


 <b>EXAMEN DE ESTADO</b> PARA INGRESO A LA EDUCACION SUPERIOR		<b>INTERPRETACION DE RESULTADOS</b> <small>0000309</small>									
<b>NÚCLEO COMÚN</b>	<b>PUNTAJE</b>	Es un resultado que representa la competencia en cada una de las pruebas del Núcleo Común en una escala que oscila entre 0 y 100 puntos aproximadamente. Los rangos de puntaje son: BAJO <= 30, MEDIO 31 A 70 Y ALTO > 70.									
	<b>GRUPOS DE PREGUNTAS</b>	El resultado por Grupos de Preguntas corresponde al desempeño relativo del examinado en cada uno de los temas que aborda cada prueba en el Núcleo Común, permite identificar cuáles son fortalezas o debilidades. En estos grupos, su desempeño se identifica en las siguientes convenciones: <b>Desempeño relativo significativamente alto (SA)</b> <b>Desempeño relativo medio (M)</b> <b>Desempeño relativo significativamente bajo (SB)</b> <b>Desempeño relativo alto (A)</b> <b>Desempeño relativo bajo (B)</b> Si usted quiere mejorar los resultados debe reforzar todos los temas y prestar especial atención a aquellos en los que obtenga Bajo o Significativamente Bajo.									
		Los temas para las distintas pruebas del examen son los siguientes:									
			<b>BIOLOGÍA</b>	<b>MATEMÁTICAS</b>	<b>FILOSOFÍA</b>	<b>FÍSICA</b>	<b>HISTORIA</b>	<b>QUÍMICA</b>	<b>LENGUAJE</b>	<b>GEOGRAFÍA</b>	<b>IDIOMA</b>
		<b>G1</b>	CELULAR	CONTEO	EPISTEMOLÓGICO	MECÁNICA CLÁSICA	CULTURAL	ASPECTOS ANALÍTICOS DE SUSTANCIAS	IDENTIFICACIÓN Y FUNCIÓN DE ELEMENTOS SEMÁNTICOS	CULTURAL	CONSTRUCCIÓN DE ORACIONES
		<b>G2</b>	ORGANISMO	MEDICIÓN	ANTROPOLÓGICO	EVENTOS ONDULATORIOS	POLÍTICO	ASPECTOS FÍSICO-QUÍMICOS DE SUSTANCIAS	CONFIGURACIÓN DEL SENTIDO GLOBAL DEL TEXTO	POLÍTICO	INTERPRETACIÓN DE GRÁFICAS
	<b>G3</b>	ECOSISTEMA	VARIACIÓN	ÉTICO	TERMODINÁMICA	ECONÓMICO	ASPECTOS ANALÍTICOS DE MEZCLAS	DEL SENTIDO DEL TEXTO HACIA OTROS TEXTOS	ECONÓMICO	CONSTRUCCIÓN DE PÁRRAFOS	
	<b>G4</b>		ALEATORIEDAD	ESTÉTICO	ELECTRO-MAGNETISMO	SOCIAL	ASPECTOS FÍSICO-QUÍMICOS DE MEZCLAS		SOCIAL	SITUACIONES COMUNICATIVAS	
	<b>G5</b>			ONTOLÓGICO					FÍSICO	COMPRESIÓN DE TEXTOS	
	<b>NIVEL DE COMPETENCIA</b>	<b>C1</b> <b>C2</b> <b>C3</b>	Describe la ejecución del examinado en cada una de las competencias evaluadas en las diferentes pruebas del Núcleo Común. Las convenciones correspondientes a los resultados de nivel de competencia se deben leer de la siguiente forma: Nivel bajo (A), Nivel Medio (B), Nivel Alto (C)								
<b>COMPLEJIDAD</b>	<b>GRADO DE PROFUNDIZACION</b>	Corresponde a la evaluación de competencias en contextos disciplinares con una mayor complejidad que en el Núcleo Común del examen. Las convenciones correspondientes a los resultados de grado de profundización se deben leer de la siguiente forma: Grado Básico (GB), Grado Bajo (I), Grado Medio (II), Grado Alto (III)									
	<b>PUNTAJE INTERDISCIPLINAR</b>	Es un resultado que representa la competencia del estudiante en la prueba interdisciplinaria seleccionada, este puntaje fluctúa en una escala que oscila entre 0 y aproximadamente 100 puntos: BAJO <= 30, MEDIO 31 A 70 Y ALTO >70									

Figura 4. Formato de interpretación de resultados de las Pruebas de Estado

**1.4.1 Puntaje:** es un resultado cuantitativo expresado en una escala que va de 0 a aproximadamente 100 puntos, el cual se produce para cada prueba del núcleo común, así como para la prueba interdisciplinaria elegida y puede interpretarse de acuerdo con los tres siguientes rangos:

BAJO entre 0 y 30 puntos

MEDIO entre 31 y 70 puntos

ALTO entre 71 o más puntos

Podría plantearse una subdivisión del rango medio en dos partes así: medio bajo, que iría desde 31 hasta 45 puntos, y medio alto, que iría desde 46 hasta 70 puntos.

**1.4.2 Nivel de competencia:** este resultado indica el nivel alcanzado en cada una de las competencias evaluadas en las pruebas del núcleo común (interpretativa, argumentativa y propositiva). Los resultados se interpretan de acuerdo con las siguientes convenciones:

BAJO A

MEDIO B

ALTO C

**1.4.3 Desempeño por grupos de preguntas:** describe la forma como se están abordando los tópicos evaluados en las pruebas del núcleo común, sopesando los resultados en los tópicos propios de cada prueba. Estos resultados se interpretan a partir de las siguientes categorías de desempeño:

DESEMPEÑO RELATIVO SIGNIFICATIVAMENTE ALTO (SA)

DESEMPEÑO RELATIVO ALTO (A)

DESEMPEÑO RELATIVO MEDIO (M)

DESEMPEÑO RELATIVO BAJO (B)

DESEMPEÑO RELATIVO SIGNIFICATIVAMENTE BAJO (SB)

**1.4.4 Grado de profundización:** indica hasta qué punto el estudiante profundizó en cada una de las 3 pruebas que eligió. Se interpreta de acuerdo con las convenciones descritas abajo. Grado Básico indica no haber alcanzado ningún grado de profundización y los grados siguientes muestran, en orden trascendente, el éxito del estudiante al abordar cada prueba.

GRADO BÁSICO GB

GRADO BAJO (I)

GRADO MEDIO (II)

GRADO ALTO (III)

**1.4.5 Puesto:** es un dato que sintetiza el perfil de puntajes de un estudiante en las pruebas del núcleo común. Hay en total 1261 puestos, siendo 1 el puesto que indica un perfil de puntajes en el rango alto de la escala o muy cercanos a este rango, y 1261 el puesto que indica un perfil de puntajes en el rango bajo de la escala o muy cercanos a este rango. Las condiciones que definen cada puesto se mantienen de una aplicación a otra del examen, ya que son independientes de la distribución de los puntajes en la población. Por tanto, no hay límite para el número de personas que puede ocupar un determinado puesto. Todos los estudiantes de una institución, de una región o incluso del país, podrían ocupar, por ejemplo, el puesto 1, siempre y cuando sus puntajes cumplan con las condiciones establecidas para asignar dicho puesto.

## **1.5 POLITICA DE ADMISIÓN A LA UIS.**

Según Admisiones UIS (2000), la universidad diseñó sus propias pruebas para evaluar la suficiencia académica de los aspirantes y las aplicó hasta el año 1984. A partir de 1985 se continuaron aplicando las pruebas de la Universidad y se consideraron algunas áreas del Examen para complementar el proceso de selección de estudiantes. En el proceso de admisión de noviembre de 1991, se aprobó como único criterio de selección de aspirantes, los resultados del Examen de Estado, decisión apoyada en un estudio donde se encontró una alta correlación entre los resultados obtenidos por los aspirantes en las pruebas de la UIS y en las realizadas en el Examen de Estado, resultado coincidente con estudios similares efectuados en otras universidades.

Con la implantación en el país del nuevo Examen de Estado a partir del año 2000, la Universidad modificó su sistema de admisión para adaptarlo a la conceptualización, enfoque metodológico, modelo matemático y escala de

calificación de las nuevas pruebas, con el propósito de lograr la utilización apropiada de los resultados entregados por este nuevo Examen de Estado.

La política de admisión de la Universidad, establecida en los artículos 90.91 y 92 del estatuto General, plantea que la Universidad será accesible a los estudiantes que demuestren poseer las capacidades y calidades requeridas y cumplan con las condiciones académicas exigidas, sin discriminación de sexo, raza, etnia, condición económica, política y social. La Universidad es autónoma para establecer su proceso de admisión. El proceso de admisión a los programas académicos de pregrado presencial, se maneja en forma centralizada y es coordinado por la Dirección de Admisiones y Registro Académico.

La Dirección de admisiones y Registro académico está adscrita a la Vicerrectoría académica y tiene como propósito dirigir y coordinar el desarrollo permanente de mejoras en los procesos de admisión y matrícula, de todos los programas de pregrado y postgrado ofrecidos por la universidad, así como efectuar los registros de documentos e información académica; dirigir el desarrollo de procesos y actividades de modernización del manejo de la información y colaborar con instituciones de educación media, con el fin de asesorar y apoyar programas de orientación de los bachilleres en el proceso de selección o escogencia de profesión.

## **1.6. CRITERIOS DE ADMISIÓN**

La Universidad tiene como criterio de admisión a los programas ofrecidos el desempeño académico del aspirante evaluado mediante el Examen de Estado, excepto para la carrera de Licenciatura en Música, en la cual se tendrá en cuenta para la admisión únicamente la prueba de Aptitud y Conocimientos musicales. La Universidad realiza dos procesos de selección: uno para los aspirantes inscritos

con Examen de Estado anterior y otro para los aspirantes inscritos con el nuevo Examen de Estado, admitiendo siempre los aspirantes que obtengan los mejores puntajes. Los cupos disponibles para cada programa académico se asignan proporcionalmente al número de aspirantes inscritos con cada tipo de Examen de Estado. El proceso de selección se hace de acuerdo a los criterios de cada carrera, cuando el criterio está conformado por más de un área del examen de estado, se toma el menor puntaje porque representa el desempeño integral del aspirante en ese conjunto de áreas. De esta forma se continúa hasta el sexto criterio para el proceso de selección de los aspirantes. A continuación presentamos los criterios de admisión para las ingenierías que son nuestro objeto de estudio:

<b>FACULTAD</b>	<b>CRITERIOS DE SELECCIÓN</b>
<b>INGENIERIAS FISICO QUIMICAS</b>  Ingeniería Metalúrgica Ingeniería de Petróleos Ingeniería Química Geología	<b>Criterios de selección para los programas de la Facultad de Ing. Físico Químicas:</b>  1. Matemáticas, Física, Química y Lenguaje  2. Matemáticas 3. Física, Química y Lenguaje 4. Lenguaje 5. Química 6. Filosofía.
<b>INGENIERIAS FISICO MECANICAS</b>  Ingeniería de Sistemas Ingeniería Civil Ingeniería Eléctrica Ingeniería Electrónica Ingeniería Industrial Ingeniería Mecánica Diseño Industrial	<b>Criterios de selección para los programas de la Facultad de Ing. Físico Mecánicas:</b>  1. Matemáticas, Física y Lenguaje  2. Matemáticas 3. Física y Lenguaje 4. Lenguaje 5. Física 6. Filosofía y Química

Tabla 1. Criterios de Admisión UIS (Fuente: Registro y Admisión Académico UIS)

### **1.7. ASIGNACIÓN DE CUPOS POR CARRERAS.**

Según el acuerdo No 207 de octubre de 2007, la asignación de cupos a cada carrera con base en las inscripciones para el primer periodo académico del año 2008, se distribuyeron así:

<b>CARRERA</b>	<b>CUPO</b>
Ingeniería de Sistemas	<b>83</b>
Ingeniería Civil	<b>83</b>
Ingeniería Eléctrica	<b>60</b>
Ingeniería Electrónica	<b>72</b>
Ingeniería Industrial	<b>86</b>
Ingeniería Mecánica	<b>86</b>
Ingeniería Metalúrgica	<b>58</b>
Ingeniería de petróleos	<b>58</b>
Ingeniería Química	<b>70</b>

Tabla 2. Cupos para admitidos por carrera (Fuente: Registro y Admisión Académico UIS)

Los cupos disponibles para cada programa académico, serán distribuidos proporcionalmente de acuerdo con el número de aspirantes inscritos con cada tipo de Examen de Estado. Los números de cupos descritos anteriormente, son constantes en los semestres anteriores para el ingreso a los programas académicos de la Universidad.

### **1.8 PUNTAJES DE CORTE DE LOS ÚLTIMOS PROCESOS DE ADMISIÓN.**

A continuación se presentan los últimos cortes para ingresar a los diferentes programas académicos de la Universidad. Esto no garantiza, que si no se tiene

este puntaje no sea probable el ingreso a la Universidad; ya que este varía de acuerdo a los puntajes de los aspirantes que se presenten en dicho semestre.

<b>CARRERAS</b>	<b>PRIMER CRITERIO</b>	<b>JUN 2005</b>	<b>NOV 2005</b>	<b>JUN 2006</b>	<b>NOV 2006</b>	<b>JUN 2007</b>
Ingeniería Civil	Matemáticas, Física, Lenguaje	48.38	53.89	52.45	55.80	52.10
Ingeniería Sistemas	Matemáticas, Física, Lenguaje	48	53.45	51.81	53.42	50.97
Ingeniería Eléctrica	Matemáticas, Física, Lenguaje	46.64	51.11	51.53	53.30	50.97
Ingeniería Electrónica	Matemáticas, Física, Lenguaje	48.38	57.98	53.45	56.39	50.97
Ingeniería Industrial	Matemáticas, Física, Lenguaje	54.20	58.70	56.00	57.14	55.21
Ingeniería Mecánica	Matemáticas, Física, Lenguaje	48.11	55.73	51.81	55.21	51.20
Ingeniería Metalúrgica	Matemáticas, Física, Lenguaje	43.63	46.83	48.75	49.26	48.40
Ingeniería Petróleos	Matemáticas, Física, Lenguaje	48.38	54.54	53.00	57.14	53.30
Ingeniería Química	Matemáticas, Física, Lenguaje	50.27	55.96	53.00	55.80	50.97

Tabla 3. Puntajes de Cortes en los últimos dos años (Fuente: Registro y Admisión Académico UIS)

## **CAPÍTULO II: MARCO TEÓRICO Y CONCEPTUAL**

## 2.1 ANÁLISIS DE COMPONENTES PRINCIPALES

Los orígenes del análisis de Componentes Principales (ACP) se remontan al año 1901, en el que Karl Pearson publicó un trabajo sobre el ajuste de un sistema de puntos de un multiespacio a un plano. Este enfoque fue retomado en 1933 por Harold Hotelling, quien fue el primero en formular el ACP tal como se ha difundido en nuestros días. El enfoque de Hotelling se centraba en el análisis de los Componentes Principales que sintetizan la mayor variabilidad del sistema de puntos, ello implica quizás el calificativo de “principal”.

La técnica de ACP puede ser considerada desde diversas perspectivas, según el énfasis del objetivo del investigador.

En primer lugar, Peña (2002, p.134) dice que “puede ser considerada como una técnica multivariada descriptiva, que permite tratar matrices de grandes dimensiones, así la podemos entender como una técnica exploratoria, en el sentido de que, explorar los datos, sin restricción alguna sobre ellos, será el primer paso para entenderlos”. Otra forma de verla es como una técnica estadística que nos sirve para probar la hipótesis que se formulan de los datos, buscando reducir el número de dimensiones originales a un conjunto de  $k$  variables, logrando la mayor interpretabilidad de los datos. Finalmente podemos ver el ACP como una técnica geométrica, la cual busca colocar de forma óptima, en algún sentido, a los ejes de coordenadas.

El Análisis de Componentes Principales transforma el conjunto de variables originales en un conjunto más pequeño de variables, y que contienen la mayor parte de la variabilidad presente en el conjunto inicial.

$X_1,$	$X_2,$	.....	$X_p$	<i>Variables originales</i>
$Y_1,$	$Y_2,$	.....	$Y_p$	<i>Componentes Principales</i>

Tabla 4. Ilustración de las componentes Principales a partir de las variables originales

Dichas componentes son combinaciones lineales de las variables originales, derivadas en orden decreciente de importancia, de tal manera que el primer componente sea el que explique la mayor cantidad posible de la variación total contenida en los datos originales.

Explicaremos matemáticamente el significado del Análisis de Componentes Principales. Supongamos que se disponen valores de  $p$ -variables en  $n$  elementos de una población dispuestos en una matriz  $X$  de dimensiones  $n \times p$ , donde las columnas contienen las variables y en las filas los elementos. “Para encontrar un espacio de dimensión  $q$  donde ( $q < p$ ), donde las variables y los individuos queden en este, se empieza por hallar un subespacio de dimensión uno: es decir, una línea recta que contenga al origen, la cual se ajusta lo mejor posible a los datos” Díaz (2002, p.199). Según gráfica 3.1.1, la proyección de un vector cualquiera  $OQ_i$  (individuos) sobre la recta  $CP_1$  es el vector  $OP_i$ . Sea  $a = (a_{11} \dots a_{1p})$  un vector unitario del subespacio  $CP_1$ , entonces la proyección  $OP_i$  es el producto escalar entre  $OQ_i$  y  $a$ . De esta forma, el producto de  $X$  (matriz de datos) y  $a$  es la proyección de cada una de las filas de  $X$  sobre  $CP_1$ .

En la figura 3, muestra una serie de puntos, y una recta que pasa muy cerca de todos los puntos, en que las distancias entre ellos se mantienen aproximadamente en su proyección sobre la recta. La condición esencial que debe cumplir la recta es que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo mas pequeñas posibles.

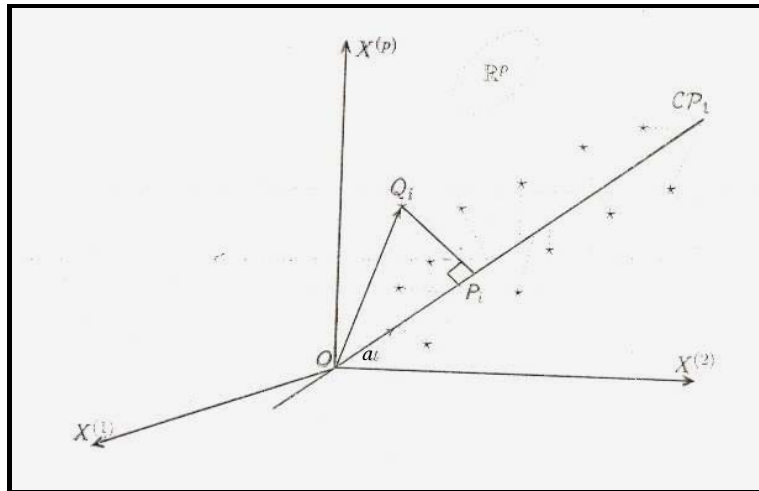


Figura 5. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella. Fuente: Díaz (2002, p.200)

Según lo anterior se muestra que al proyectar cada uno de los puntos sobre la recta trazada, se forma un triángulo rectángulo, donde su hipotenusa es la distancia del punto al centro  $OQ_i$ , y los dos catetos son la distancia entre el punto y su proyección  $Q_iP_i$ , y el otro es la proyección del punto sobre la recta  $OP_i$ . Aplicando el teorema de Pitágoras entre los lados del triángulo  $OP_iQ_i$ , resulta la siguiente expresión al sumar cada uno de los triángulos rectángulos determinados por sus proyecciones sobre el subespacio  $CP_1$ .

$$\sum_{i=1}^n (Q_iP_i)^2 = \sum_{i=1}^n (OQ_i)^2 - \sum_{i=1}^n (OP_i)^2$$

Las proyecciones  $OP_i$  reflejan la información recogida en el subespacio  $CP_1$  de cada punto. Se quiere maximizar esta cantidad de información. Como los puntos  $Q_i$  están a una misma distancia fija del origen  $O$ , es decir son constantes, minimizar  $\sum_{i=1}^n (Q_iP_i)^2$ , la suma al cuadrado de las distancias a la recta de todos los

puntos, es equivalente a maximizar  $\sum_{i=1}^n (OP_i)^2$ , la suma al cuadrado de los valores de las proyecciones. La cantidad a maximizar es:

$$\sum_{i=1}^n (OP_i)^2 = (Xa)'(Xa) = a'X'Xa$$

Con la restricción que  $a'a = 1$

El mejor subespacio bidimensional que ajusta los  $n$  puntos es el generado por  $a_1$  y  $a_2$ , donde  $a_2$  es el segundo vector en la base para este subespacio ortogonal a  $a_1$  que maximiza a  $a_2'X'Xa_2$ . Luego queda claro que en una Proyección de un punto en una dirección, maximizar la varianza de la proyección equivale a minimizar las distancias.

## 2.1.2 CÁLCULO DE LAS COMPONENTES PRINCIPALES

### 2.1.2.1 Cálculo de la Primer Componente

El primer componente Principal, según Peña (2002) se define como la combinación lineal de las variables originales que tienen varianza máxima. Los valores en este primer componente de los  $n$  individuos se representarán por un vector  $Y_1$ , dado por:

$$Y_1 = Xa_1$$

Para una variable Multivariante, se define la varianza como:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

Como las variables originales tienen media cero también  $Y_1$  tendrá media nula. Su varianza será:

$$\text{var}(Y_1) = \frac{1}{n} Y_1' Y_1 = \frac{1}{n} a_1' X' X a_1 = a_1' \sum a_1$$

⋮

En general

$$\text{var}(y_i) = a_i' \sum a_i$$

Donde la matriz  $\sum$  corresponde a la matriz de varianzas y covarianzas muestrales de las  $p$ -variables. Es obvio que podemos maximizar la varianza sin límite aumentando el módulo del vector  $a_1$ . Para que la maximización tenga solución debemos imponer una restricción al módulo del vector  $a_1$ , y, sin pérdida de generalidad, impondremos que  $a_1' a_1 = 1$  o dicho de forma general  $a_j' a_j = 1 \Rightarrow \|a_j\| = 1$ , esto para evitar la indeterminación del sistema. Se puede observar que la incógnita es precisamente  $a_1$  (el vector desconocido que nos da la combinación óptima). La forma de introducir esta restricción es mediante el multiplicador de Lagrange lo cual es una técnica para maximizar una función con restricciones.

$$L(a_1) = a_1' \sum a_1 - \lambda (a_1' a_1 - 1)$$

Para maximizar a  $L(a_1)$  se deriva con respecto a  $a_1$ , igualando a cero y resolviendo

$$\begin{aligned} \frac{\partial L}{\partial a_1} &= 2 \sum a_1 - 2 \lambda a_1 = 0 \\ \Rightarrow \sum a_1 - \lambda a_1 &= 0 \end{aligned}$$

Luego

$$\Rightarrow (\Sigma - \lambda I)a_1 = 0 \text{ ó } \Sigma a_1 = \lambda a_1$$

Para obtener una solución diferente a la trivial ( $a_1 = 0$ ), se debe cumplir que  $(\Sigma - \lambda I) = 0$  y puesto que la expresión entre paréntesis es una matriz, tenemos que su determinante es el que debe ser igual a cero

$$|\Sigma - \lambda I| = 0$$

El desarrollo del determinante  $|\Sigma - \lambda I| = 0$  conduce a una ecuación característica que permite obtener los valores de  $\lambda$ , que se identifican con los valores asociados a la matriz de varianza y covarianza  $\Sigma$ . Puesto que  $\Sigma$  es una matriz semidefinida positiva, todos sus autovalores serán positivos y serán ordenados tal que  $\lambda_1 < \lambda_2 < \lambda_3 \dots < \lambda_p$ . Luego teniendo este resultado, podemos decir lo siguiente:

$$\text{var}(Y_1) = a_1' \Sigma a_1 = a_1' \lambda I a_1 = \lambda a_1' a_1 = \lambda I = \lambda$$

Lo cual quiere decir que para maximizar la varianza de  $Y_1$  se tiene que tomar el mayor valor propio, digamos en este caso  $\lambda_1$ , y el correspondiente vector propio  $a_1$ . En realidad,  $a_1$  es un vector que nos da la combinación de las variables originales que tiene mayor varianza, esto es, si

$$Y_1 = a_1' X = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

### 2.1.2.2 Cálculo del segundo Componente Principal

El segundo Componente principal se determina hallando un segundo vector  $a_2$  ortogonal a  $a_1$ , tal que que  $Y_2 = a_2'X$  tenga la segunda varianza más grande entre todos los vectores que satisfacen:

$$a_1 a_2' = 0 \text{ (Ortogonalidad entre los dos vectores)}$$

$$a_2' a_2 = 1 \text{ (Ortonormalidad)}$$

Para obtener el segundo Componente principal nuevamente planteamos la maximización de la varianza de  $Y_2 = a_2'X$ ,  $V(Y_2) = a_2' \Sigma a_2$  sujeto a las restricciones de que  $a_2' a_2 = 1$  (módulo unitario) y además está incorrelacionado con  $Y_1$ ; Según Entralgo (2003), la covarianza entre las dos componentes es cero;  $Cov(Y_1, Y_2) = 0$  reemplazando  $Y_1$ ,  $Y_2$ , y efectuando el proceso tenemos lo siguiente:

$$Cov(Y_1, Y_2) = Cov(a_1'X, a_2'X) = E[a_2'(X - \mu)(X - \mu)'a_1] = a_2'E[(X - \mu)(X - \mu)']a_1$$

Luego  $Cov(Y_1, Y_2) = a_2' \Sigma a_1$  sustituyendo la expresión  $\Sigma a_1 = \lambda a_1$  tenemos:

$$Cov(Y_1, Y_2) = a_2' \lambda a_1, \quad Cov(Y_1, Y_2) = \lambda a_2' a_1 = 0 \text{ y como } \lambda \neq 0 \text{ finalmente tenemos que } a_2' a_1 = 0, \text{ lo que significa que los vectores } a_1 \text{ y } a_2 \text{ son perpendiculares.}$$

De esta manera se agrega otra condición de que  $a_2' a_1 = 0$  (los dos vectores son ortogonales). Generamos nuevamente la función Lagrangeana:

$$L(a_2) = a_2' \Sigma a_2 - \lambda(a_2' a_2) - \beta(a_2' a_1)$$

Se deriva respecto a  $a_2$ :

$$\frac{\partial L(a_2)}{\partial a_2} = 2\sum a_2 - 2\lambda a_2 - \beta a_1 = 0$$

Si se multiplica por  $a_1'$ , entonces

$$2a_1'\Sigma a_2 - \beta = 0$$

Porque

$$a_1'a_2 = a_2'a_1 = 0 \quad \text{y} \quad a_1'a_1 = 1$$

Luego

$$\beta = 2a_1'\Sigma a_2 = 2a_2'\Sigma a_1 = 0$$

ya que

$$\text{Cov}(Y_1, Y_2) = 0$$

Finalmente la expresión queda:

$$\frac{\partial L(a_2)}{\partial a_2} = 2\sum a_2 - 2\lambda a_2 - \beta a_1 = 2\Sigma a_2 - 2\lambda a_2 = (\Sigma - \lambda I)a_2 = 0$$

$$(\Sigma - \lambda I)a_2 = 0 \quad \text{ó} \quad \Sigma a_2 = \lambda a_2$$

Usando el mismo procedimiento que en el primer Componente, elegimos  $\lambda$  como el segundo mayor autovalor de la matriz  $\Sigma$  con su autovector asociado  $a_2$ . En realidad,  $a_2$  es un vector que nos da la combinación de las variables originales que tiene mayor varianza, esto es, si

$$Y_2 = a_2'X = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$$

Resumiendo, todos los componentes Y (en total p) se pueden expresar como el producto de una matriz formada por los vectores propios, multiplicada por el vector x que contiene las variables originales  $x_1, \dots, x_p$

$$Y = Ax$$

donde

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_p \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{p1} & a_{p2} & \cdot & a_{pp} \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_p \end{pmatrix}$$

Como también, decimos de las varianzas:

$$\text{Var}(Y_1) = \lambda_1$$

$$\text{Var}(Y_2) = \lambda_2$$

·

·

·

$$\text{Var}(Y_p) = \lambda_p$$

La matriz de varianza y Covarianza de los Componentes Principales será una matriz diagonal, denotada por L, y se escribe:

$$L = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_p \end{pmatrix}$$

Porque  $Y_1, \dots, Y_p$  se han construido como variables incorrelacionadas.

Se tiene que

$$L = \text{Var}(Y) = A\text{Var}(X)A = A\Sigma A$$

ó también:

$$L = A\Sigma A$$

Al premultiplicar por A y post multiplicar por A', tenemos que la matriz de varianza y covarianza de las variables originales se puede factorizar como:

$$\Sigma = ALA'$$

### 2.1.2.3 Matriz de varianza y covarianza

Según Peña (2002) Supóngase que los valores de p-variables  $X_1, X_2, X_3, \dots, X_p$  se obtienen sobre una muestra de n-individuos, la matriz X de tamaño  $(n \times p)$  representa tales datos. Sean  $\bar{X}_j$  la media muestral de la variable  $X_j$ ;  $S_{kj}$  la covarianza muestral entre las variables  $X_j$  y  $X_k$  y la matriz  $\Sigma = S_{jk}$  corresponde a la matriz de varianzas y covarianzas muestrales de las p-variables. Para variables escalares la variabilidad respecto a la media se mide habitualmente por la varianza, o su raíz cuadrada, la desviación típica. La relación lineal entre dos variables se mide por la covarianza. La *covarianza* entre las variables  $X_j$  y  $X_k$  se calcula por:

$$S_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

y mide su dependencia lineal.

Para una variable multivariante se define la matriz de varianzas y covarianzas como:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

que es una matriz cuadrada y simétrica que contiene en la diagonal las varianzas de las variables y fuera de la diagonal las covarianzas entre las variables. En efecto, al multiplicar los vectores:

$$\begin{bmatrix} x_{i1} - \bar{x}_1 \\ \cdot \\ \cdot \\ \cdot \\ x_{ip} - \bar{x}_p \end{bmatrix} \begin{bmatrix} x_{i1} - \bar{x}_1, & \cdot & \cdot & \cdot & , x_{ip} - \bar{x}_p \end{bmatrix} = \begin{bmatrix} (x_{i1} - \bar{x}_1)^2 & \cdot & \cdot & \cdot & (x_{i1} - \bar{x}_1)(x_{ip} - \bar{x}_p) \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ (x_{ip} - \bar{x}_p)(x_{i1} - \bar{x}_1) & \cdot & \cdot & \cdot & (x_{ip} - \bar{x}_p)^2 \end{bmatrix}$$

Se obtiene la matriz de cuadrados y productos cruzados de las  $p$ -variables en el elemento  $i$ . Al sumar para todos los elementos y dividir por  $n$  se obtienen las varianzas en la diagonal y las covarianzas fuera de ella. La matriz de varianzas y covarianzas, que llamaremos para simplificar matriz de covarianzas, es la matriz simétrica de orden  $p$  con forma:

$$\Sigma = \begin{bmatrix} s_1^2 & \cdot & \cdot & \cdot & s_1^p \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ s_p^1 & \cdot & \cdot & \cdot & s_p^2 \end{bmatrix}$$

#### 2.1.2.4 Porcentajes de Variabilidad

Como observamos anteriormente, cada valor propio corresponde a la varianza de cada componente, que se definía por medio de su respectivo vector propio, es decir,  $Var(Y_i) = \lambda_i$

Al realizar la suma de todos estos valores propios, obtenemos la varianza de los componentes, es decir

$$\sum_{i=1}^p Var(Y_i) = \sum_{i=1}^p \lambda_i = traza(L)$$

Donde la matriz L es diagonal. Pero por propiedades, tenemos que:

$$\text{traza}(L) = \text{traza}(A'\Sigma A) = \text{traza}(\Sigma A' A) = \text{traza}(\Sigma)$$

Porque  $AA' = I$  al ser A ortogonal, dando como resultado

$$\text{traza}(L) = \text{traza}(\Sigma) = \sum_{i=1}^p \text{var}(x_i)$$

Es decir, la suma de las varianzas de las variables originales y la suma de las varianzas de las componentes Principales son iguales. Esto permite hablar del porcentaje de varianza total que recoge un componente principal:

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} = \frac{\lambda_i}{\sum_{i=1}^p \text{Var}(x_i)}$$

Si esta expresión la multiplicamos por 100, obtenemos el porcentaje. Así también se podrá expresar el porcentaje de variabilidad recogido por los primeros  $m$  componentes:

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \text{Var}(x_i)}$$

Donde  $m < p$

### 2.1.2.5 Representación Gráfica. Matriz Factorial

En el análisis de componentes principales es importante conocer la correlación de cada variable con las componentes. Su obtención es muy sencilla. Así, el coeficiente de correlación  $r_{hj}$  entre la componente  $h$ -ésima y la variable  $j$ -ésima viene dada por:

$$r_{hj} = u_{hj} \sqrt{\lambda_h}$$

Donde  $\lambda_k$  es el valor propio y  $u_{kj}$  es el vector propio.

A estos coeficientes de correlación se les denomina cargas factoriales. Estas cargas son hallados así: teniendo en cuenta la tabla de los vectores propios y los valores propios, aplicamos la fórmula anterior, dando resultados:

$$r_{11} = u_{11} \sqrt{\lambda_1}, \quad r_{12} = u_{12} \sqrt{\lambda_2} \dots \quad r_{hj} = u_{hj} \sqrt{\lambda_h}$$

De esta manera se representan las cargas factoriales de cada componente principal, que son las coordenadas para representar las variables en un diagrama de dispersión y mirar el comportamiento de los individuos en este plano factorial.

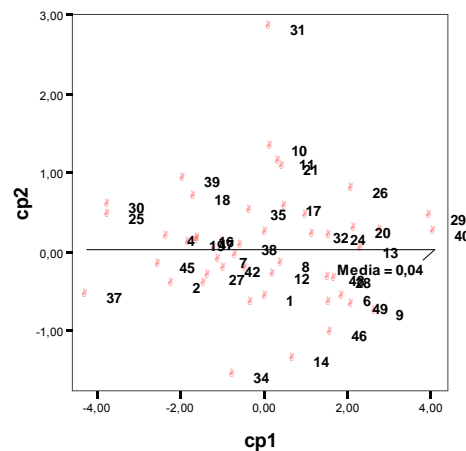


Figura 6. Diagrama de dispersión

## 2.2 TABLAS DE CONTINGENCIA

Cuando se trabaja con variables categóricas, los datos suelen organizarse en tablas de doble entrada en las que cada entrada representa un criterio de clasificación (una variable categórica). Como resultado de esta clasificación, las frecuencias (el número o porcentaje de casos) aparecen organizadas en casillas que contienen información sobre la relación existente entre ambos criterios. A estas tablas de frecuencias se les llama *tablas de contingencia*. Universidad de Cádiz (2002).

Este procedimiento de las tablas de contingencia permite obtener tablas de dos dimensiones, pero también se puede incluir una tercera variable que se llamará variable de segmentación, esto con el fin de definir subgrupos y obtener así tablas multidimensionales.

Las tablas de contingencia serán de tipo  $h \times k$  siendo  $k$  el número de columnas que forman las frecuencias observadas y  $h$  el número de filas.

A continuación presentamos para ilustrar por medio del siguiente ejemplo los datos del rendimiento en matemáticas de dos cursos (9-10) y (9-12) del Instituto Técnico Superior Dámaso Zapata, en el segundo periodo académico del año 2007, en los niveles de aprobó todos los indicadores de logro o no aprobó todos los indicadores. Para ilustrar, tenemos:

		Aprobados	No Aprobados	Total
CURSOS	9-10	2	40	42
	9-12	10	30	40
Total		12	70	82

Tabla 5. Tabla de Contingencia de dos cursos contra el nivel de aprobación o no aprobación

Esta es una tabla de tamaño  $2 \times 2$  ( $h = 2; k = 2$ ). La tabla 5 presenta los valores observados de los dos cursos, pero falta los valores esperados teóricamente. Estos se calculan de la siguiente forma donde  $e_{ij}$  es el valor esperado asumiendo independencia en cada celda.

$$(frecuencia\ esperada)_{ij} = \frac{(total\ de\ la\ fila\ i) * (total\ de\ la\ columna\ j)}{n^\circ\ total\ de\ casos}$$

		NIVEL APROBACIÓN			Total
		TODOS LOS LOGROS DE MATEMÁTICAS			
			Aprobados	No Aprobados	
GRADO	9-10	Frecuencia Observada	2	40	42
		Frecuencia esperada	6.14	35.85	41.9
	9-12	Frecuencia Observada	10	30	40
		Frecuencia esperada	5.85	34.14	39.9
Total		Frecuencia Observada	12	70	82
		Frecuencia esperada	11.9	69.99	81.89

Tabla 6. Tabla de Contingencia de los dos cursos 9-10 y 9-12 con las frecuencias observadas y frecuencias esperadas

$$e_{11} = \frac{42 * 12}{82} = 6.14$$

$$e_{12} = \frac{42 * 70}{82} = 35.85$$

$$e_{21} = \frac{40 * 12}{82} = 5.85$$

$$e_{22} = \frac{40 * 70}{82} = 34.14$$

Para mirar si existe diferencia entre los dos cursos con el nivel de aprobación de todos los logros de matemáticas, se analiza por medio de algún estadístico. En nuestro caso usaremos el Chi-cuadrado

### 2.2.1 Estadístico Chi cuadrado

Este estadístico permite contrastar la hipótesis de que los dos criterios de clasificación utilizados (las dos variables categóricas) son independientes. Para ello, compara las frecuencias observadas (las frecuencias de hecho obtenidas) con las frecuencias esperadas (las frecuencias que teóricamente deberíamos haber encontrado en cada casilla si los dos criterios de clasificación fueran independientes). Obtenidas las frecuencias esperadas para cada casilla, el estadístico  $\chi^2$  o Chi-cuadrado de Pearson se obtiene de la siguiente manera:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

( $n_{ij}$  Se refiere a las frecuencias observadas y  $m_{ij}$  a las esperadas). De la ecuación se desprende que el estadístico  $\chi^2$  valdrá cero cuando las variables sean completamente independientes (pues las frecuencias observadas y las esperadas serán iguales), y que el valor del estadístico  $\chi^2$  será tanto mayor sea la diferencia entre las frecuencias observadas y las esperadas (diferencia que será tanto mayor cuanto mayor sea la relación entre las variables).

Tomando el ejemplo de la tabla de contingencia 6, estimamos el estadístico Chi-Cuadrado así:

$$\chi^2 = \frac{(2 - 6.14)^2}{6.14} + \frac{(40 - 35.85)^2}{35.85} + \frac{(10 - 5.85)^2}{5.85} + \frac{(30 - 34.14)^2}{34.14} = 6.71$$

El estadístico  $\chi^2$  sigue el modelo de distribución de probabilidad  $\chi^2_{(n-1)(k-1)}$  con los grados de libertad resultantes de multiplicar el número de filas menos uno (n-1) por el número de columnas menos uno (k-1), Para el ejemplo, tenemos:  $gl = ((2-1)(2-1)) = 1$ . Por tanto, podemos utilizar la distribución  $\chi^2$  para establecer el grado de relación existente entre el valor del estadístico  $\chi^2$  y la hipótesis de independencia. Si el valor calculado es mayor que el valor de las tablas (ver figura 5), significará que las diferencias entre las frecuencias observadas y las frecuencias teóricas o esperadas son muy elevadas y por tanto diremos con un determinado nivel de confianza que existe dependencia entre los factores o variables analizadas. Según Hamdan (2005), Si los datos son compatibles con la hipótesis de independencia, la probabilidad asociada al estadístico  $\chi^2$  será alta (mayor de 0.05). Si esa probabilidad es muy pequeña (menor que 0.05), consideramos que los datos se muestran incompatibles con la hipótesis de independencia y concluiremos que las variables estudiadas están relacionadas.

Para que las probabilidades de la distribución  $\chi^2$  constituyan una buena aproximación a la distribución del estadístico  $\chi^2$  conviene que se cumplan algunas condiciones, entre ellas: Las frecuencias esperadas no sean demasiado pequeñas. Suele asumirse que, si existe frecuencias esperadas menores que 5, estas no deben superar el 20 por ciento del total de frecuencias esperadas. En nuestro caso, todas son mayores de 5, toman valores altos en los valores esperados.

Resumiendo, tenemos:

$\chi^2 > \chi^2_{(n-1)(k-1)} \Rightarrow$  Rechazar hipótesis nula (dependencia entre las variables)

$\chi^2 < \chi^2_{(n-1)(k-1)} \Rightarrow$  Aceptar hipótesis nula (independencia entre las variables)

Según el resultado del Chi-cuadrado, vemos que el estadístico Chi cuadrado de Pearson toma un valor de 6.71, para un nivel de significancia de 95%, siendo mayor que el Chi-cuadrado tabulado (3.85) (ver figura 5, con 1 grado de libertad y un nivel de significancia de 95%); Rechazándose la hipótesis de independencia concluyendo que las variables grado y nivel están relacionadas. (Son dependientes).

$\nu$	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,15}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	23,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	48,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	116,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

Figura 7. Distribución  $\chi^2$

## 2.2.2 Estimaciones de Riesgo

Pita (1997, p.2) sostuvo que

*“El término de riesgo implica que la presencia de una característica o factor aumenta la probabilidad de consecuencias adversas. La cuantificación del grado*

de riesgo constituye un elemento esencial y fundamental en la formulación de políticas y prioridades que no deben dejar hueco a la intuición ni a la casualidad”.

Hay diferentes maneras de cuantificar ese riesgo.

- a. *Riesgo Absoluto*: Mide la incidencia del daño en la población total
- b. *Riesgo Relativo*: Compara la frecuencia con que ocurre el daño entre los que tienen el factor de riesgo y los que no lo tienen.

De acuerdo con el ejemplo de la tabla 5, estimaremos el riesgo relativo. El riesgo relativo de *no aprobación* del curso 9-10 frente al curso 9-12 es:

$$RR = \frac{40/42}{30/40} = \frac{0.95}{0.75} = 1.26$$

Es decir, el riesgo relativo mide la fuerza de asociación entre el grado (9-10, 9-12) y la aprobación o no aprobación de los indicadores de logros. En el caso de no aprobación, el riesgo relativo se estima dividiendo el número de estudiantes de un grado que no aprobaron entre el número total de los que aprobaron y no aprobaron. Este resultado dividido entre los estudiantes del otro grado que no aprobaron entre el número total de aprobados y no aprobados.

De la misma forma se calcula el riesgo de *aprobación* del curso de 9-10 frente a 9-12:

$$RR = \frac{2/42}{10/40} = \frac{0.04}{0.25} = 0.16$$

Para cuantificar la importancia de la asociación, la forma es calcular el cociente de esos dos riesgos relativos calculados anteriormente, dando como resultado .16.

Este valor se interpreta de la misma manera que el índice de riesgo relativo (pues es una estimación del mismo), admitiendo otra interpretación y nos indica cuanto más probable es que ocurra que el curso de 9-10 aprueben todos los indicadores respecto a 9-12. Es decir, le va mejor al curso de 9-12, ya que tiene más probabilidad de pasar todos los indicadores que el curso de 9-10 y una menor de no aprobar. Este cociente es conocido como el *odds ratio*<sup>1</sup> (OR), siendo otra forma de representar un riesgo. Molinero (2001). Este cálculo se puede hacer de esta manera:

		INDICADOR		Total
		Aprobó	No Aprobó	
GRADO	9-10	2	40	42
	9-12	10	30	40
Total		12	70	85

Tabla 7. Tabla de contingencia de grado 9-10, 9-12 contra estudiantes que aprobaron todos los indicadores o no aprobaron.

$$OR = \frac{2 * 30}{10 * 40} = \frac{60}{400} = 0.15$$

Un índice de riesgo de 1 indica que la probabilidad de encontrarnos con el factor desencadenante es la misma en las dos situaciones estudiadas.

### 2.2.3 Intervalos de Confianza para la estimación del riesgo

Cuando calculamos el riesgo relativo, debemos expresar si dicho riesgo es diferente de 1. Si al construir el intervalo de confianza del 95%, el intervalo no incluye el valor 1, concluimos que el riesgo es estadísticamente significativo

<sup>1</sup> No hay un término en castellano que sea bien aceptado

$p < 0.05$ . Si el intervalo de confianza del 99% no incluye el valor 1, el riesgo es significativo  $p < 0.01$  (Pita, 1997). El cálculo de dicho intervalo de confianza del 95% para el riesgo relativo se realiza del siguiente modo:

$$(RR) \exp[\pm 1.96 \text{ Error Estándar del Ln RR}]$$

donde:

- RR es la estimación puntual del riesgo relativo
- Exp es la base del logaritmo natural elevada a la cantidad entre paréntesis
- Error Estándar =  $\sqrt{\frac{1}{a} - \frac{1}{(a+b)} + \frac{1}{c} - \frac{1}{(c+d)}}$  donde a,b,c y d representan los valores numéricos de la tabla 2x2 que contiene la información de la situación

Si usamos el ejemplo anterior tenemos:

Logaritmo natural de 0.16 = -1.83

$$\text{Error Estándar} = \sqrt{\frac{1}{2} - \frac{1}{42} + \frac{1}{10} - \frac{1}{40}} = 0.74$$

El intervalo de confianza del 95% del logaritmo de  $RR = -1.83 \pm 1.96 * 0.74 = (-0.37; -3.28)$

El antilogaritmo de estos límites es:  $e^{-0.37}$  a  $e^{-3.28} = 0.04$  a  $0.7$

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para GRADO (9-10 / 9-12 )	,136	,028	,662
Para la cohorte NIVEL = A	,177	,042	,752
Para la cohorte NIVEL = NA	1,302	1,069	1,585
N de casos válidos	83		

Tabla 8. Estimación de riesgo

De esta manera se calcula para el nivel de no aprobación. Para el cálculo del 95% para el OR sería:

$$(OR) \exp \left[ \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right]$$

donde

- OR es la estimación puntual del Odds ratio
  - exp es la base del logaritmo natural elevada a la cantidad entre paréntesis
- a,b,c,d representan los valores numéricos de la tabla 2x2

El cálculo del intervalo de confianza del riesgo relativo y del odds ratio es fundamental al realizar el análisis de cualquier estudio. Dicho cálculo nos indica la significancia estadística, si el intervalo no engloba el valor 1 y la precisión del intervalo que está directamente relacionada con el tamaño muestral del estudio

## 2.3 ANÁLISIS DE COMPONENTES PRINCIPALES CON SPSS

### 2.3.1 Presentación de los datos

En esta sección, se desarrolla un caso particular, que busca ilustrar de una forma didáctica los conceptos trabajados hasta el momento de la técnica de Componentes Principales. Los cálculos se desarrollan con el programa SPSS (*Statistical Package for Social Sciences*, versión 12).

Los datos de la tabla son algunas medidas corporales de pájaros. Fueron extraídos de (Díaz, p. 219). El objetivo es estudiar el efecto de la selección natural en tales aves. Las variables son:

$X_1$ : Longitud total

$X_2$ : Extensión de las alas

$X_3$ : Longitud de cabeza hasta el pico

$X_4$ : Longitud del húmero

$X_5$ : Longitud de la quilla (*Esternón o pecho*)

Se midieron 49 pájaros moribundos después de una tempestad, 21 de los cuales sobrevivieron. A continuación se presenta el análisis de los datos a través de los siguientes pasos:

#### Datos de los pájaros

OBS	X1	X2	X3	X4	X5
1	156	245	31,6	18,5	20,5
2	154	240	30,4	17,9	19,6
3	153	240	31	18,4	20,6
4	153	236	30,9	17,7	20,2

5	155	243	31,5	18,6	20,3
6	163	247	32	19	20,9
7	157	238	30,9	18,4	20,2
8	155	239	32,8	18,6	21,2
9	164	248	32,7	19,1	21,1
10	158	238	31	18,8	22
11	158	240	31,3	18,6	22
12	160	244	31,1	18,6	20,5
13	161	246	32,3	19,3	21,8
14	157	245	32	19,1	20
15	157	235	31,5	18,1	19,8
16	156	237	30,9	18	20,3
17	158	244	31,4	18,5	21,6
18	153	238	30,5	18,2	20,9
19	155	236	30,3	18,5	20,1
20	163	246	32,5	18,6	21,9
21	159	236	31,5	18	21,5
22	155	240	31,4	18	20,7
23	156	240	31,5	18,2	20,6
24	160	242	32,6	18,8	21,7
25	152	232	30,3	17,2	19,8
26	160	250	31,7	18,8	22,5
27	155	237	31	18,5	20
28	157	245	32,2	19,5	21,4
29	165	245	33,1	19,8	22,7
30	153	231	30,1	17,3	19,8
31	162	239	30,3	18	23,1
32	162	243	31,6	18,8	21,3
33	159	245	31,8	18,5	21,7
34	159	247	30,9	18,1	19
35	155	243	30,9	18,5	21,3
36	162	252	31,9	19,1	22,2
37	152	230	30,4	17,3	18,6
38	159	242	30,8	18,2	20,5
39	155	238	31,2	17,9	19,3
40	163	249	33,4	19,5	22,8
41	163	242	31	18,1	20,7
42	156	237	31,7	18,2	20,3
43	159	238	31,5	18,4	20,3
44	161	245	32,1	19,1	20,8
45	155	235	30,7	17,7	19,6
46	162	247	31,9	19,1	20,4
47	153	237	30,6	18,6	20,4
48	162	245	32,5	18,5	21,1
49	164	248	32,3	18,8	20,9

Tabla 9 Medidas corporales de gorriones. Fuente: Manly (1998, p. 2)

## Varianza de los datos de los pájaros

S11	S22	S33	S44	S55
3,9204	13,49	0,0198292	0,01	0,10660225
15,8404	1,76	1,1218701	0,25	1,50430225
24,8004	1,76	0,2108497	0	0,05130225
24,8004	28,37	0,3126864	0,49	0,39250225
8,8804	2,80	0,001666	0,04	0,27720225
25,2004	32,19	0,2924823	0,36	0,00540225
0,9604	11,07	0,3126864	0	0,39250225
8,8804	5,41	1,7977884	0,04	0,13950225
36,2404	44,54	1,5396251	0,49	0,07480225
0,0004	11,07	0,2108497	0,16	1,37710225
0,0004	1,76	0,0253395	0,04	1,37710225
4,0804	7,15	0,1290129	0,04	0,10660225
9,1204	21,84	0,7069721	0,81	0,94770225
0,9604	13,49	0,2924823	0,49	0,68310225
0,9604	40,02	0,001666	0,09	1,05370225
3,9204	18,72	0,3126864	0,16	0,27720225
0,0004	7,15	0,0035027	0,01	0,59830225
24,8004	11,07	0,9200334	0,04	0,00540225
8,8804	28,37	1,3437069	0,01	0,52780225
25,2004	21,84	1,0832986	0,04	1,15240225
1,0404	28,37	0,001666	0,16	0,45360225
8,8804	1,76	0,0035027	0,16	0,01600225
3,9204	1,76	0,001666	0,04	0,05130225
4,0804	0,45	1,3014618	0,16	0,76300225
35,7604	86,98	1,3437069	1,44	1,05370225
4,0804	75,23	0,0579925	0,16	2,80060225
8,8804	18,72	0,2108497	0,01	0,68310225
0,9604	13,49	0,5488088	1,21	0,32890225
49,2804	13,49	2,6922781	1,96	3,51000225
24,8004	106,64	1,8473803	1,21	1,05370225
16,1604	5,41	1,3437069	0,16	5,16880225
16,1604	2,80	0,0198292	0,16	0,22420225
1,0404	13,49	0,1161558	0,01	0,76300225
1,0404	32,19	0,3126864	0,09	3,33610225
8,8804	2,80	0,3126864	0,01	0,22420225
16,1604	113,92	0,194319	0,49	1,88650225
35,7604	128,29	1,1218701	1,21	4,95730225
1,0404	0,45	0,4345232	0,04	0,10660225
8,8804	11,07	0,0671762	0,25	2,33020225
25,2004	58,88	3,7667679	1,21	3,89470225
25,2004	0,45	0,2108497	0,09	0,01600225
3,9204	18,72	0,0579925	0,04	0,27720225
1,0404	11,07	0,001666	0	0,27720225

9,1204	13,49	0,4106455	0,49	0,00070225
8,8804	40,02	0,5763599	0,49	1,50430225
16,1604	32,19	0,194319	0,49	0,18190225
24,8004	18,72	0,7381966	0,04	0,18190225
16,1604	13,49	1,0832986	0,01	0,07480225
36,2404	44,54	0,7069721	0,16	0,00540225
<b>13,0812</b>	<b>25,16</b>	<b>0,6187422</b>	<b>0,31673</b>	<b>0,96276552</b>

Tabla 10. Cálculos de la diagonal de matriz de varianzas de los pájaros

## 2.3.2 Pasos a seguir a través de SPSS

### 2.3.2.1 Para Componentes principales

#### *Paso 1. Obtención de los datos en SPSS*

datos pájaros - Editor de datos SPSS

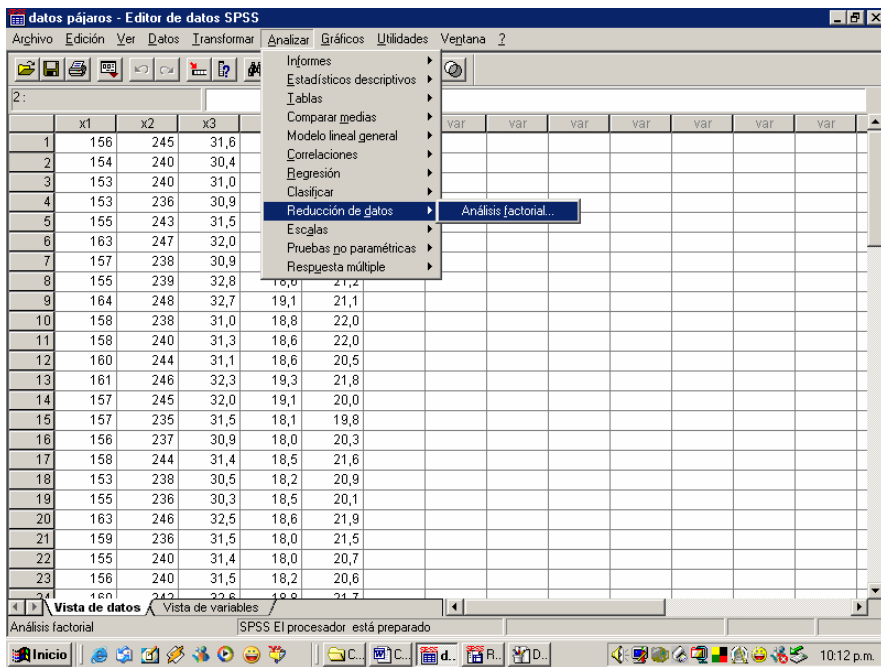
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

2:

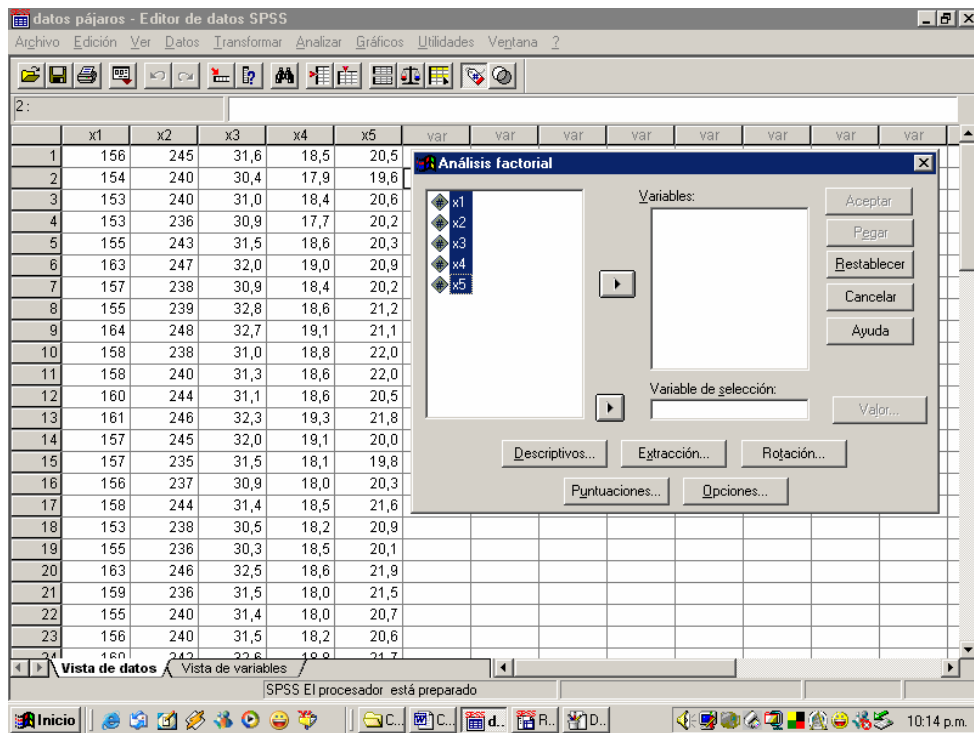
	x1	x2	x3	x4	x5	var	var	var	var	var	var	var	var
1	156	245	31,6	18,5	20,5								
2	154	240	30,4	17,9	19,6								
3	153	240	31,0	18,4	20,6								
4	153	236	30,9	17,7	20,2								
5	155	243	31,5	18,6	20,3								
6	163	247	32,0	19,0	20,9								
7	157	238	30,9	18,4	20,2								
8	155	239	32,8	18,6	21,2								
9	164	248	32,7	19,1	21,1								
10	158	238	31,0	18,8	22,0								
11	158	240	31,3	18,6	22,0								
12	160	244	31,1	18,6	20,5								
13	161	246	32,3	19,3	21,8								
14	157	245	32,0	19,1	20,0								
15	157	235	31,5	18,1	19,8								
16	156	237	30,9	18,0	20,3								
17	158	244	31,4	18,5	21,6								
18	153	238	30,5	18,2	20,9								
19	155	236	30,3	18,5	20,1								
20	163	246	32,5	18,6	21,9								
21	159	236	31,5	18,0	21,5								
22	155	240	31,4	18,0	20,7								
23	156	240	31,5	18,2	20,6								
24	160	242	32,6	18,9	21,7								

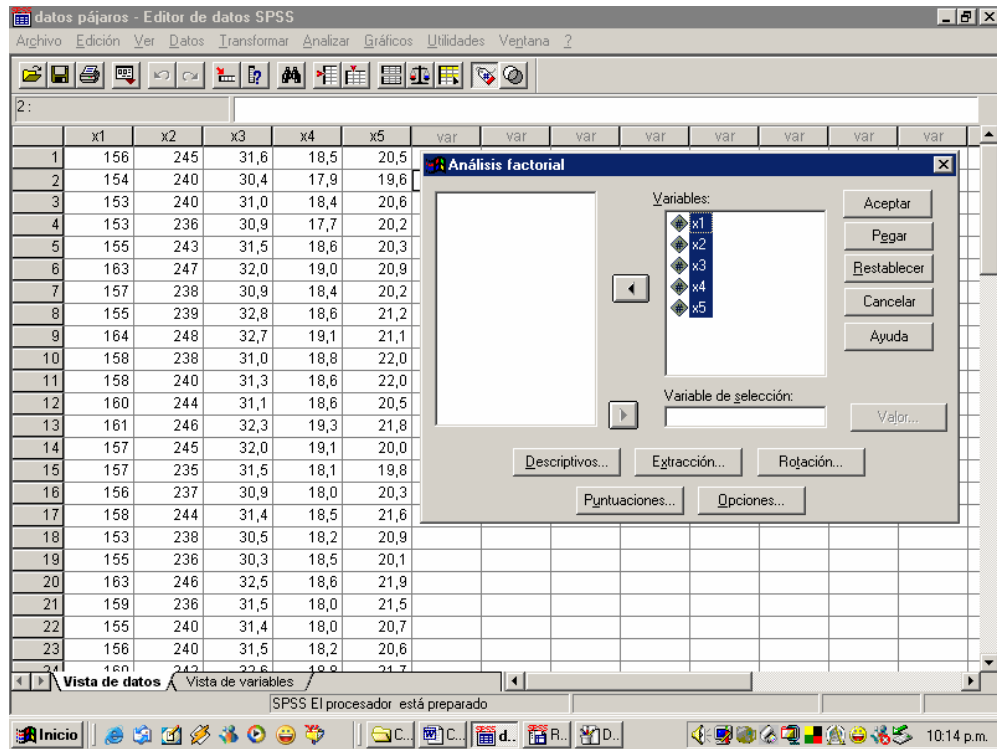
Vista de datos Vista de variables

SPSS El procesador está preparado

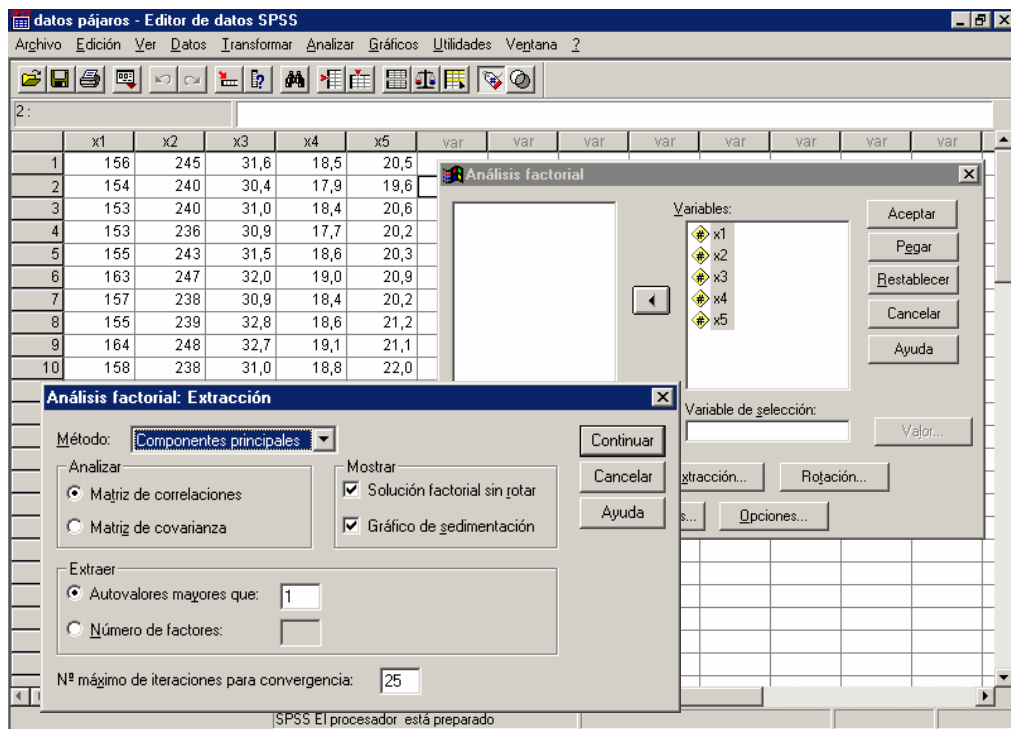


Paso 2. Identificar las variables que se van a trabajar





Paso 3. Elegir el método de extracción de las Componentes Principales.



#### Paso 4. Analizar los resultados

	x1	x2	x3	x4	x5
Correlación x1	1,000	,735	,662	,645	,605
x2	,735	1,000	,674	,769	,529
x3	,662	,674	1,000	,763	,526
x4	,645	,769	,763	1,000	,607
x5	,605	,529	,526	,607	1,000

². Determinante = ,037

Tabla 11. Matriz de correlación

En esta matriz se observa cuales variables se relacionan más. Vemos que la variable  $X_1$  y  $X_2$  tienen el valor más alto (0.735). Esto nos indica el grado de afinidad entre la longitud total y la extensión de las alas.

#### Matriz de Comunalidades

	Inicial	Extracción
x1	1,000	,738
x2	1,000	,771
x3	1,000	,734
x4	1,000	,801
x5	1,000	,572

Método de extracción: Análisis de Componentes principales.

Tabla 12. Matriz de Comunalidades

Esta matriz expresa los porcentajes de variabilidad de cada variable que es explicada por las componentes extraídas, en este caso una. La variabilidad de la longitud total del ave ( $X_1$ ) está explicada en un 73.8% por la componente extraída en el procedimiento.

## Porcentajes individuales y acumulados de la varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,616	72,320	72,320	3,616	72,320	72,320
2	,532	10,630	82,950			
3	,386	7,728	90,678			
4	,302	6,031	96,709			
5	,165	3,291	100,000			

Método de extracción: Análisis de Componentes principales.

Tabla 13. Varianza total explicada

El porcentaje de variación total explicado por la componente es de 72.3%. De acuerdo a esta tabla, la información es valiosa, ya que nos muestra el porcentaje de la varianza de cada componente y nos ayuda a determinar cuantos componentes retener. Es inmediato en este caso que el primer Componente es más importante que los demás; pues como se observa en la tabla anterior, esta reúne casi las tres cuartas partes (72,32%) de la variabilidad total. Este se halla de la siguiente manera:

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_p} = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \text{Var}(x_i)}$$

$$\frac{3.616}{3.16 + 0.532 + 0.386 + 0.302 + 0.165} * 100\% = \frac{3.616}{5.001} * 100\% = 72.32\%$$

## Grafico de Sedimentación

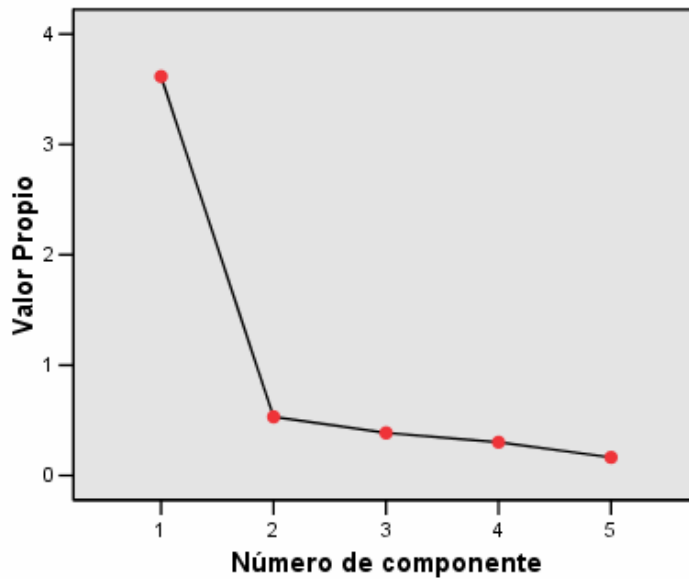


Figura 8. Gráfico de sedimentación

Por medio del *gráfico de autovalores* o *grafico de sedimentación* (ver figura 6), consiste en representar el porcentaje de variación explicada contra el número del componente. En el eje de las ordenadas se registra el porcentaje de variación explicada. En el eje de las abscisas se coloca el número del componente según su orden de importancia de acuerdo a la variación explicada.

Al analizar este gráfico se busca el punto de quiebre, donde el cambio de la pendiente se hace mayor, y la abscisa correspondiente a este punto indica el número de componentes a retener. En nuestro caso, observando la gráfica, se produce en la segunda componente.

## Matriz de Componentes

	Componente	
	1	2
x1	,859	,037
x2	,878	-,218
x3	,857	-,237
x4	,895	-,135
x5	,756	,639

Método de extracción: Análisis de componentes principales.  
<sup>a</sup>. 2 componentes extraídos

Tabla 14. Matriz de Componentes

En esta matriz se recoge la carga o ponderación de cada factor en cada una de las variables. Con estos valores, podemos llevar las variables a un plano factorial de componente 1 vs. componente 2

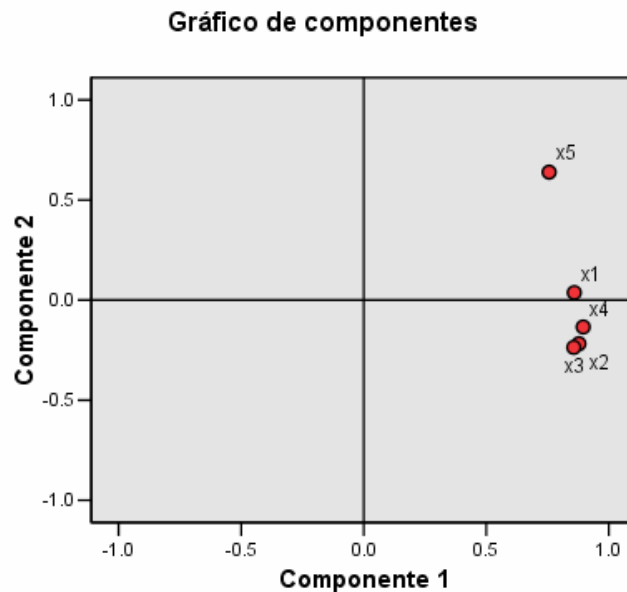


Figura 9. Plano Factorial de las dos componentes

En la figura anterior se observa las variables en el primer plano factorial. Se ve la influencia de cada variable sobre cada una de las componentes. El primer eje está asociado con las variables  $X_1$  y  $X_4$  y el segundo cargan más alto las variables  $X_5$  y  $X_3$ .

### Matriz de coeficientes para el cálculo de las puntuaciones en la Componentes

*Matriz de coeficientes para el cálculo de las puntuaciones en las componentes*

	Componente	
	1	2
x1	,238	,070
x2	,243	-,411
x3	,237	-,445
x4	,248	-,253
x5	,209	1,202

Método de extracción: Análisis de componentes principales.

Tabla 15. Matriz para el cálculo de las puntuaciones factoriales

La primera componente con las variables ( $X_1, X_2, X_3, X_4, X_5$ ) es:

$$Y_1 = 0.238X_1 + 0.243x_2 + 0.237X_3 + 0.248X_4 + 0.209X_5$$

que es un indicador del tamaño de los pájaros, ya que todos los coeficientes de la componente principal son positivos, y además alrededor del 72.3% de la variación en los datos está relacionada con diferencias de tamaño; dicho en otras palabras, el primer componente reúne las variables que determinan el tamaño de las aves

La segunda componente principal es:

$$Y_2 = 0.07X_1 - 0.41x_2 - 0.44X_3 - 0.253X_4 + 1.20X_5$$

Las variables  $X_2$  (extensión de las alas),  $X_3$  (Longitud de pico y cabeza) y  $X_4$  (Longitud del húmero) contrastan con la variable  $X_5$  (longitud de la quilla); es decir al aumentar el primer grupo de medidas  $X_5$  disminuye y viceversa. En consecuencia,  $Y_2$  reúne las variables que registran la forma de las aves. El valor tan bajo de  $X_1$  en  $Y_2$  significa que el tamaño de los pájaros afecta poco a  $Y_2$ .

### Coordenadas del diagrama de Dispersión

OBS	CP1	CP2
1	0,04	-0,83
2	-1,14	-0,61
3	-0,60	0,03
4	-1,21	0,24
5	-0,15	-0,92
6	1,02	-0,82
7	-0,55	-0,16
8	0,24	-0,23
9	1,43	-1,08
10	0,11	1,80
11	0,20	1,56
12	0,15	-0,43
13	1,26	0,01
14	0,39	-1,91
15	-0,73	-0,61
16	-0,81	0,20
17	0,30	0,74
18	-0,87	0,92
19	-0,93	0,13
20	1,16	0,37
21	-0,23	1,46
22	-0,50	0,14
23	-0,34	-0,11
24	0,85	0,25
25	-1,96	0,62
26	1,13	1,08
27	-0,69	-0,47
28	0,92	-0,51
29	2,12	0,58
30	-1,95	0,79
31	0,08	3,89
32	0,64	0,29
33	0,55	0,57
34	-0,37	-2,18

35	-0,16	0,68
36	1,49	0,34
37	-2,23	-0,77
38	-0,28	0,06
39	-1,00	-1,24
40	2,16	0,31
41	0,04	0,31
42	-0,49	-0,34
43	-0,21	-0,34
44	0,85	-0,92
45	-1,32	-0,26
46	0,87	-1,44
47	-0,81	0,16
48	0,83	-0,49
49	1,14	-0,96

Tabla 16. Coordenadas del Diagrama de Dispersión

### Diagrama de Dispersión

En la figura se han ubicado las 49 aves de acuerdo con sus coordenadas respecto a las dos primeras componentes  $CP_1$  y  $CP_2$ . Observemos algunos comportamientos de las aves: se observa que el ave número 31 tiene el valor más alto respecto a al quilla o estenon (pecho), nótese que la segunda componente está altamente influenciada por esta variable con un peso de 0.639 (ver tabla 3.1.10). Las aves numeradas como 30, 25, 37 tienen los valores más bajos respecto a la variable  $X_4$  (longitud del húmero), la más importante en el primer componente con un peso de 0.895, mientras que las aves numeradas con 29 y 40 tienen los valores más altos en esta misma variable.

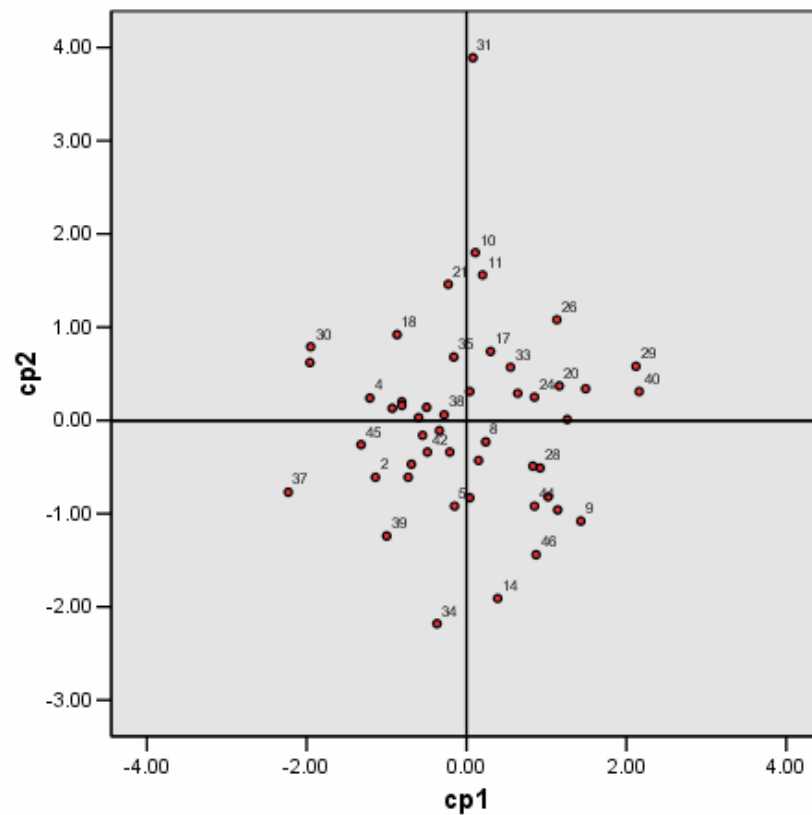
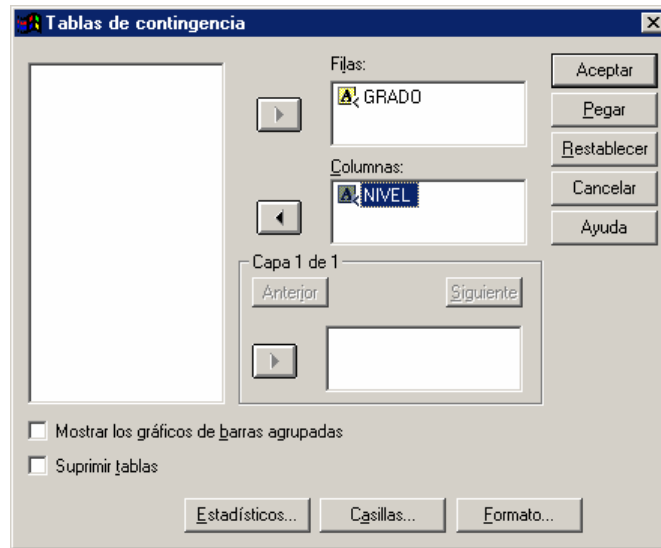


Figura 10. Diagrama de dispersión

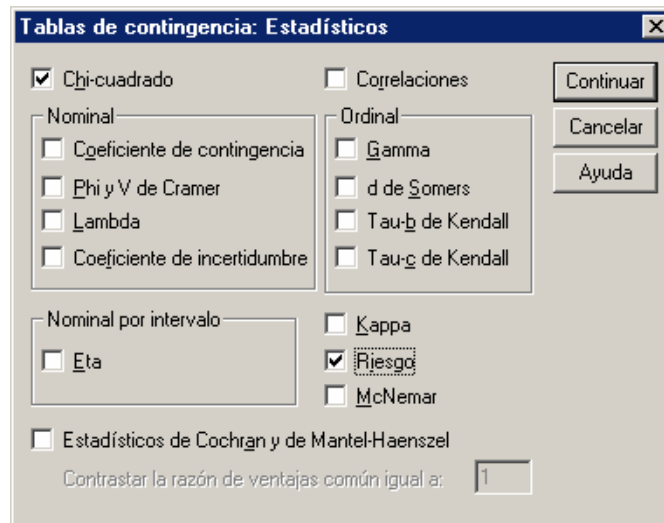
### 2.3.2.2 Para Tablas de Contingencia y Estimaciones de Riesgo

Según Vicens y Medina (2005), Para analizar si existe relación de dependencia o no entre variables por medio de SPSS, tendremos que seleccionar en el menú “analizar” la opción “estadístico” y dentro de esta opción “tabla de contingencia”. El cuadro de diálogo que aparece es el siguiente:



Debemos seleccionar las variables con las cuales queremos realizar el análisis. Siempre introduciremos la variable dependiente en las filas (grado) y la variable independiente en las columnas (nivel).

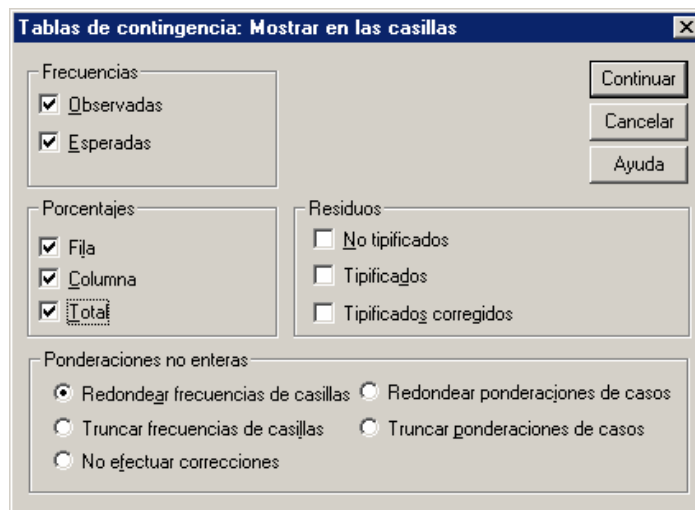
A continuación introduciremos el “estadístico” a usar, en este caso el Chi-Cuadrado:



En este caso el Chi-Cuadrado nos permite contrastar la relación de dependencia o independencia entre las variables de estudio. También señalaremos la etiqueta

“riesgo”, que nos permitirá realizar un análisis sobre las estimaciones de riesgo de los datos analizados.

La etiqueta “casilla” especificamos que valores queremos que aparezca en cada celda de las tablas de contingencia. Podemos seleccionar las frecuencias observadas, las esperadas, los porcentajes de cada una de las celdas sobre la fila correspondiente, columna y sobre el total. Esta información resulta muy útil para describir características de la muestra.



Los resultados o salidas como los llama SPSS son los siguientes:

			NIVEL		Total
			A	NA	
GRADO	9-10	Recuento	2	40	42
		Frecuencia esperada	6,1	35,9	42,0
		% de GRADO	4,8%	95,2%	100,0%
		% de NIVEL	16,7%	57,1%	51,2%
		% del total	2,4%	48,8%	51,2%
	9-12	Recuento	10	30	40
		Frecuencia esperada	5,9	34,1	40,0
		% de GRADO	25,0%	75,0%	100,0%
% de NIVEL		83,3%	42,9%	48,8%	
	% del total	12,2%	36,6%	48,8%	
Total	Recuento	12	70	82	
	Frecuencia esperada	12,0	70,0	82,0	
	% de GRADO	14,6%	85,4%	100,0%	
	% de NIVEL	100,0%	100,0%	100,0%	
	% del total	14,6%	85,4%	100,0%	

Tabla 17. Tabla de contingencia

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	6,717 <sup>b</sup>	1	,010		
Corrección por continuidad	5,195	1	,023		
Razón de verosimilitud	7,207	1	,007		
Estadístico exacto de Fisher				,012	,010
N de casos válidos	82				

Tabla 18. Prueba Chi-Cuadrado

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para GRADO (9-10 / 9-12 )	,150	,031	,736
Para la cohorte NIVEL = A	,190	,044	,816
Para la cohorte NIVEL = NA	1,270	1,049	1,538
N de casos válidos	82		

Tabla 19. Estimaciones de riesgo

## **CAPÍTULO III: RESULTADOS Y ANÁLISIS DE LA INVESTIGACIÓN**

### 3.1 ANÁLISIS DE COMPONENTES PRINCIPALES

A continuación presentaremos los resultados obtenidos a partir del procesamiento del software SPSS (*Statistical Package for Social Sciences*, versión 12). al aplicar la Técnica Estadística Multivariada de Análisis de Componentes Principales con el fin de buscar la relación entre puntajes ICFES y el rendimiento académico de los ingenieros en el primer semestre en la UIS. Primero analizaremos los estadísticos descriptivos para cada una de las variables de estudio.

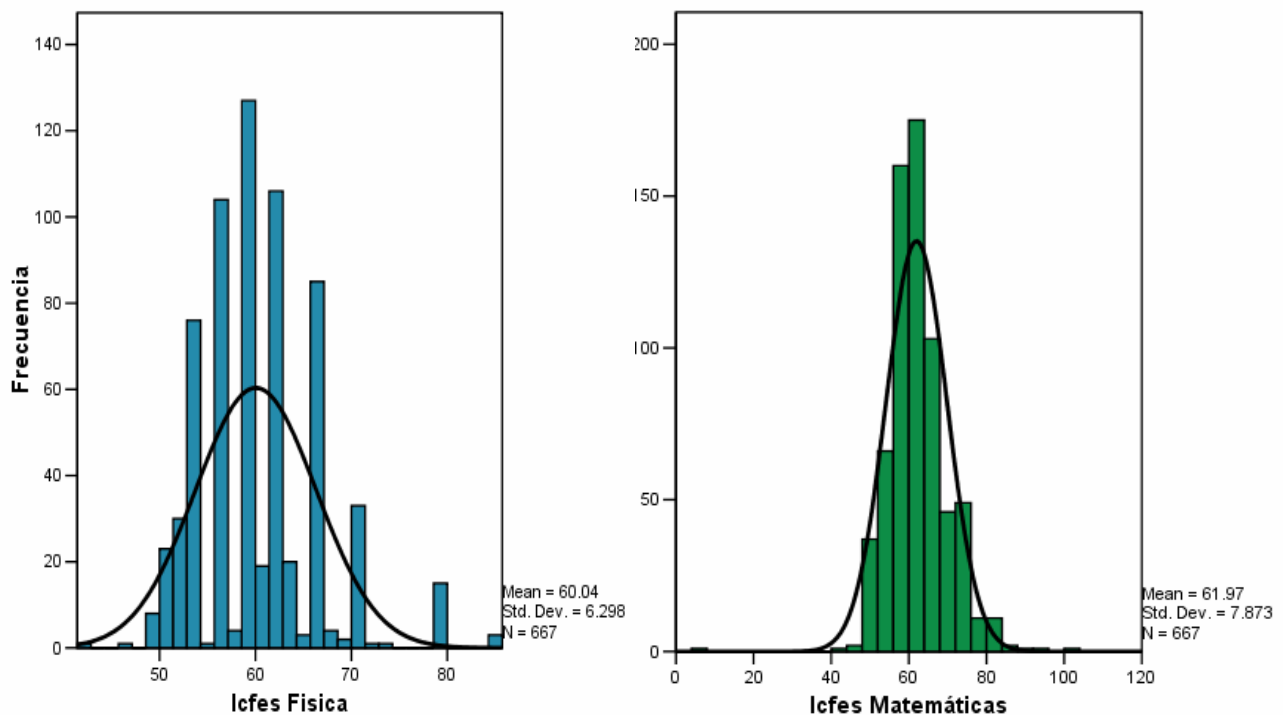
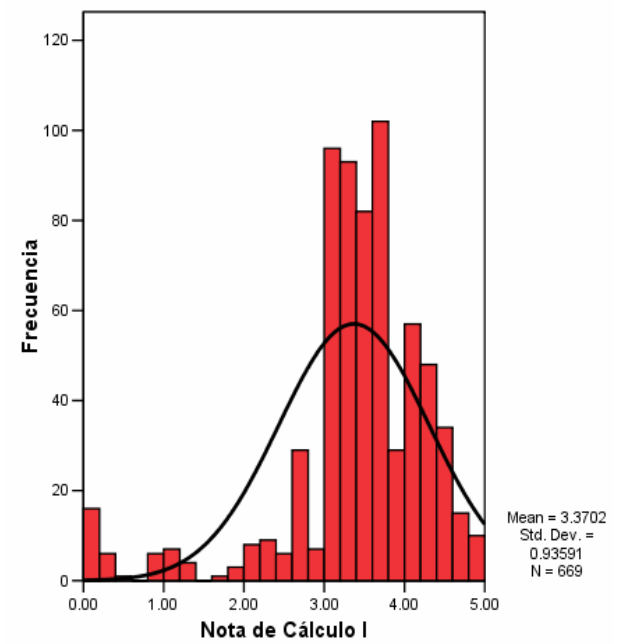
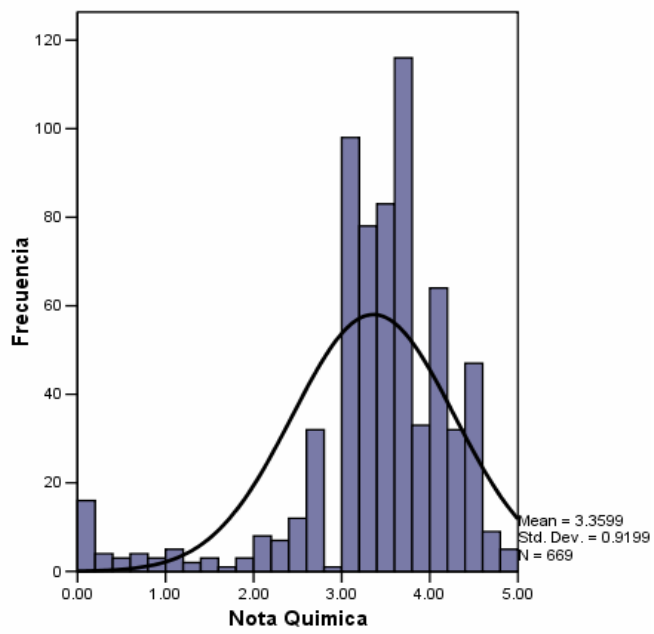
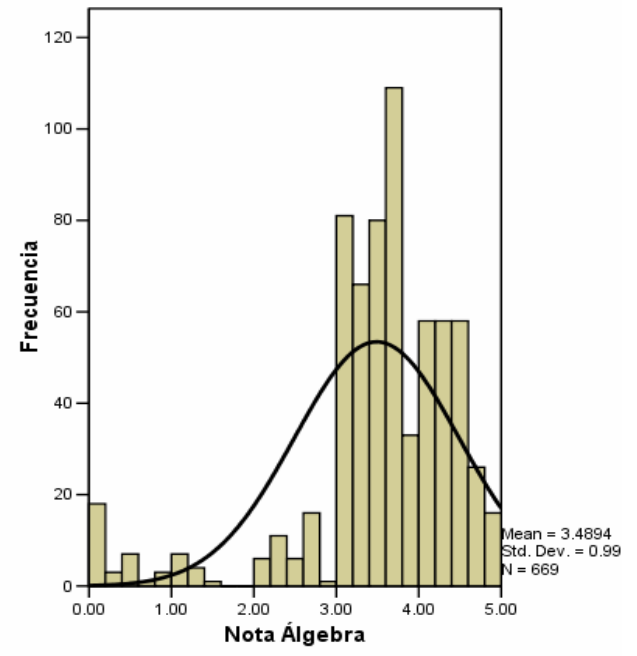
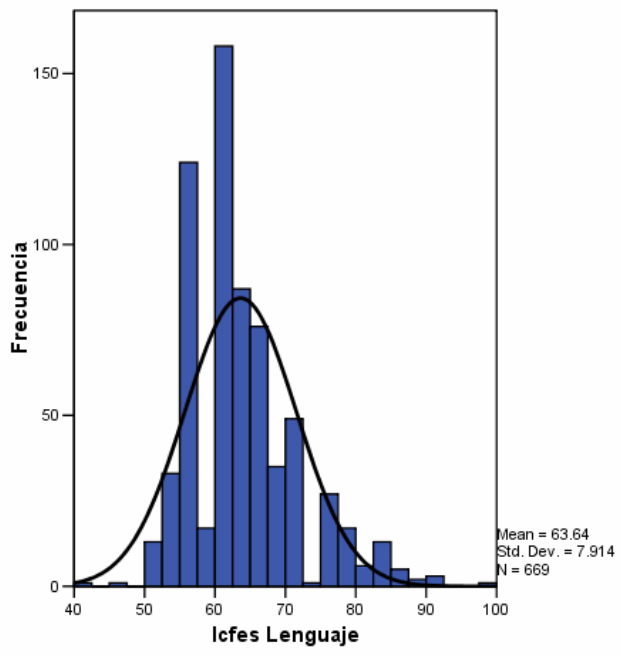


Figura 11. Histogramas con los estadísticos descriptivos de las variables en estudio



Observando la tabla 4.1.1, vemos que en general las notas definitivas promedio para las materias de Álgebra lineal, Cálculo I y Química son de: 3.38, 3.50 y 3.36, respectivamente. Las medias para los puntajes de ICFES son: Icfes en Matemáticas (61.97); Icfes en Física (60.04); Icfes en Lenguaje (63.63).

	Media	Desviación típica	Número de individuos
notacalc	3,3743	,93197	667
notaalg	3,4943	,99398	667
notaquim	3,3631	,91713	667
icfesmat	61,97	7,873	667
icfesfis	60,04	6,298	667
icfesleng	63,63	7,899	667

Tabla 20. Estadísticos descriptivos de las variables

La muestra se trabajó con 667 individuos que ingresaron a la Universidad durante los años 2004 a 2006 a los programas de ingenierías Físico Químicas y Físico Mecánicas de donde 209 son del género femenino y 460 son del género masculino. Los puntajes del ICFES se ubican por encima de 60 puntos, lo cual nos dice que se encuentran en el nivel medio según la escala propuesta por el ICFES (Esta escala se estima de la siguiente manera: bajo: menores de 30. Medio: 31 a 70. Alto: mayores de 70). Estos datos nos garantizan, hasta el momento, que están ingresando a la Universidad estudiantes con puntajes altos de acuerdo a las ponderaciones propuestas para la admisión a las distintas carreras.

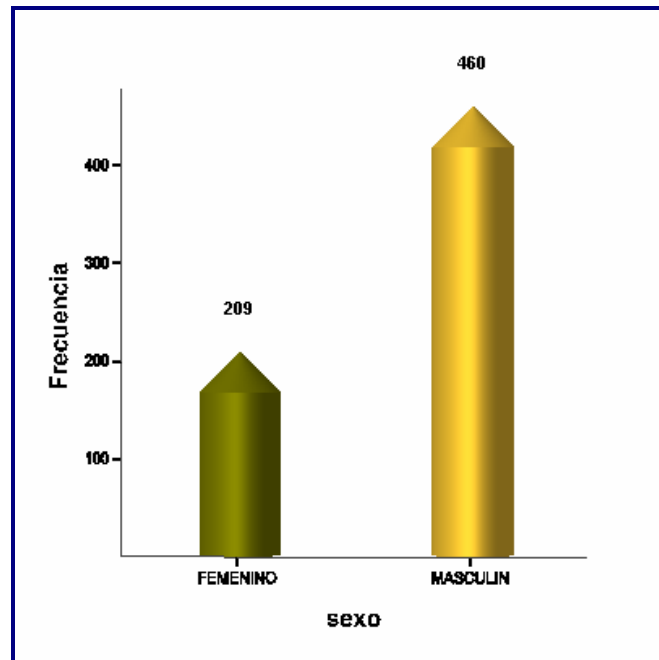


Figura 12. Frecuencias de los individuos en estudio por género

El estudio se elaboró principalmente con estudiantes de las carreras de Ingeniería Química (99 estudiantes), Ingeniería Industrial (97 estudiantes) e Ingeniería Mecánica (83 estudiantes). Ver figura 11.

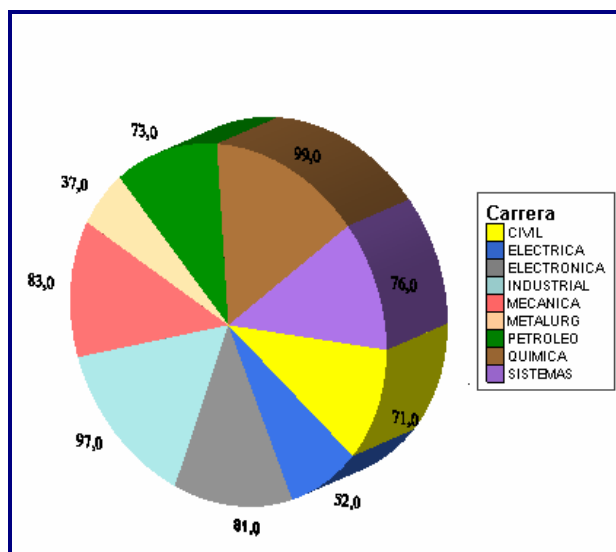


Figura 13. Frecuencias de los individuos en estudio por carreras

Observamos en la figura 12, que el 80% de estudiantes que ingresan al primer semestre, lo hacen con edades de 17 y 18 años, y tan solo el 20% restante, lo hacen en edades diferentes. Hay estudiantes que ingresan desde la edad de 15 años como la edad mínima hasta 45 años que es la edad más alta.

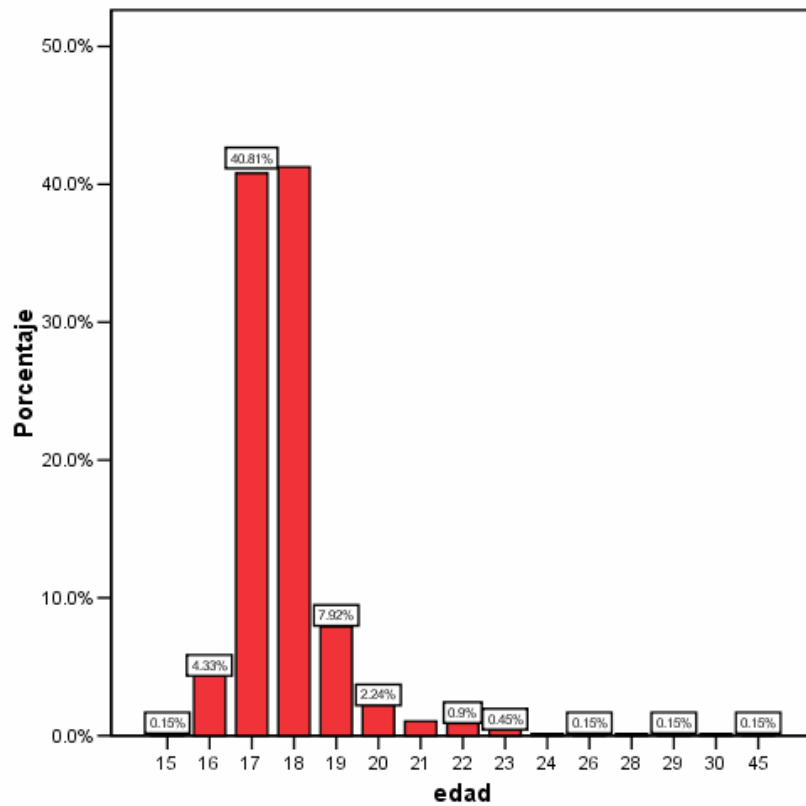


Figura 14. Frecuencias de los individuos en estudio por edades

Se observa en la tabla 21, que el estadístico KMO (Kaiser Meyer Olkin) vale 0.755, valor muy cercano a la unidad que es lo más aconsejado, lo que indica una adecuación asertiva de nuestros datos a un modelo del análisis factorial<sup>2</sup>.

El contraste de Bartlett se aplica con la formulación del contraste de hipótesis siguiente

$$H_0 : |R| = 1$$

$$H_1 : |R| \neq 1$$

Este test se basa en la distribución Chi-cuadrado donde valores altos llevan a rechazar  $H_0$ , así, la prueba de esfericidad de Bartlett contrasta si la matriz de correlaciones es una matriz identidad, que indicaría que el modelo factorial es inadecuado a los datos.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,755
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1718,565
	gl	15
	Sig.	,000

Tabla 21. Prueba de Bartlett y KMO

En este caso se rechaza la Hipótesis Nula con un nivel de significación de 0,05, dado que el valor de p-value es igual 0; lo que indica que el determinante de la matriz de correlación generalizada entre todas las variables es distinto de 1. Haciendo posible concluir que existe correlación significativa entre las variables. Por lo tanto tiene sentido aplicar el análisis factorial a estos datos.

Los coeficientes de la matriz de correlaciones deben ser grandes en valor absoluto fuera de la diagonal principal de la matriz de correlaciones, asociado al hecho de

<sup>2</sup> A partir de 0,5 se considera que es buena la adecuación muestral para un análisis factorial

que, el determinante es menor que 1 (el máximo valor del determinante es 1 si las variables están incorrelacionadas).

		notacalc	notaalg	notaquim	icfesmat	icfesfis	icfesleng
Correlación	notacalc	1,000	,805	,792	,124	,101	,156
	notaalg	,805	1,000	,820	,171	,138	,150
	notaquim	,792	,820	1,000	,167	,129	,172
	icfesmat	,124	,171	,167	1,000	,335	,287
	icfesfis	,101	,138	,129	,335	1,000	,127
	icfesleng	,156	,150	,172	,287	,127	1,000

a Determinante = ,075

Tabla 22. Matriz de correlaciones

En este caso, el determinante de la matriz es 0.075 que está relativamente cercano a 0 e indica algún tipo de correlación entre las variables, sin identificar por el momento cuales son las variables que se relacionan entre si.

Observamos la tabla 22, en la que se muestra que las variables que tienen mayor correlación son nota de álgebra lineal y nota de cálculo I (0.805), nota de química y nota de cálculo I (0.792), nota de Química y nota de Álgebra lineal (0.820). De la misma manera, las variables que tienen menor correlación entre ellas está nota de cálculo con Icfes en Matemáticas (0.124), Icfes Física con nota de Álgebra Lineal (0.138), nota de Química con Icfes Física (0.129), Icfes Lenguaje con Nota de Química (0.127), nota de cálculo con Icfes Física (0.101). Si observamos detenidamente esta matriz es simétrica, con unos en su diagonal principal.

Luego de chequear el ajuste de la técnica a los datos y la existencia de correlación entre las variables, se procede a la aplicación de la técnica Componentes Principales.

	Inicial	Extracción
notacalc	1,000	,863
notaalg	1,000	,878
notaquim	1,000	,870
icfesmat	1,000	,653
icfesfis	1,000	,489
icfesleng	1,000	,371

Tabla 23. Matriz de Comunalidades

En la tabla 23, se presentan las comunalidades, que representan el porcentaje de variabilidad de cada variable que es explicada por cada componente. Sería igual a 0 si los factores comunes no explicaran nada la variabilidad de una variable, y sería igual a 1 si quedase totalmente explicada.

La variabilidad de la nota de cálculo I está explicada en un 86.3% por las componentes extraídas en el procedimiento y así para el resto de las variables. Como vemos en la gran mayoría de la variables son explicadas con un porcentaje alto por las componentes, excepto Icfes en lenguaje que tiene un valor poco significativo (37.1%) a diferencia de las demás variables.

El gráfico de varianza total explicada se utiliza para determinar cuantos factores deben retenerse.

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2,764	46,066	46,066	2,764	46,066	46,066
2	1,360	22,672	68,738	1,360	22,672	68,738
3	,873	14,550	83,288			
4	,617	10,279	93,567			
5	,209	3,477	97,044			
6	,177	2,956	100,000			

Tabla 24. Varianza total explicada

Para determinar cuantos componentes retener, miramos en la Tabla 4.1.5 la varianza total explicada. Notamos que el primer componente explica un 46.06% de la varianza total y la siguiente componente explica un 22.67% de la varianza total; Se observa que las dos primeras componentes, explican el 69% de la varianza total. El resto de los valores propios son muy pequeños al compararlos con las dos primeras. Obsérvese que en este caso la suma de los valores propios es seis, el número total de variables consideradas. Por lo que se decide retener sólo las dos primeras componentes.

Esto es fácilmente verificable en la figura 4.1.5, la cual muestra los valores propios asociados a las componentes. Obsérvese que el primer cambio del comportamiento del gráfico ocurre en el segundo valor propio.

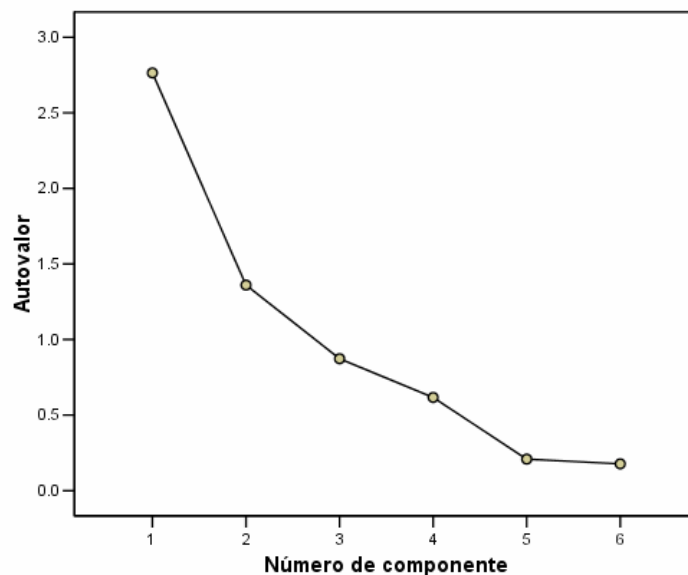


Figura 15. Diagrama de Sedimentación

Según la información de la matriz de componentes (Ver tabla 25) tenemos ya las variables iniciales definidas en función de las componentes (factores). Para ver qué variables se agrupan en torno a cada Componente y cuya combinación lineal define la componente. Para ello representamos cada una de las 6 variables por

medio de los dos factores extraídos utilizando la matriz de componentes. Podemos escribirlo de esta manera:

$$\text{nota calculo} : 0.89C_1 - 0.24C_2$$

$$\text{IcfesMatemáticas} : 0.342C_1 - 0.732C_2$$

$$\text{nota Química} : 0.91C_1 - 0.207C_2$$

$$\text{IcfesFísica} : 0.27C_1 - 0.64C_2$$

$$\text{nota Álgebra} : 0.91C_1 - 0.20C_2$$

$$\text{IcfesLenguaje} : 0.32C_1 - 0.51C_2$$

	Componente	
	1	2
notacalc	,895	-,249
notaalg	,914	-,207
notaquim	,911	-,201
icfesmat	,342	,732
icfesfis	,278	,642
icfesleng	,322	,518

Tabla 25. Matriz de Componentes

Para ver qué variables se agrupan en cada componente (factor) hay que observar las variables cuyas cargas sean altas en un factor y bajas en los otros (valores menores que 0,25 suelen considerarse bajos).

En la primera componente está representada claramente las notas en la universidad (nota Química, nota de Cálculo I, nota de Álgebra Lineal) y en la segunda (Icfes Física, Icfes Lenguaje, Icfes Matemáticas). Se ve claramente marcados los grupos de variables notas en la universidad y puntajes de Icfes, las cuales son componentes incorrelacionadas y cada una tiene un peso sobre su Componente Principal.

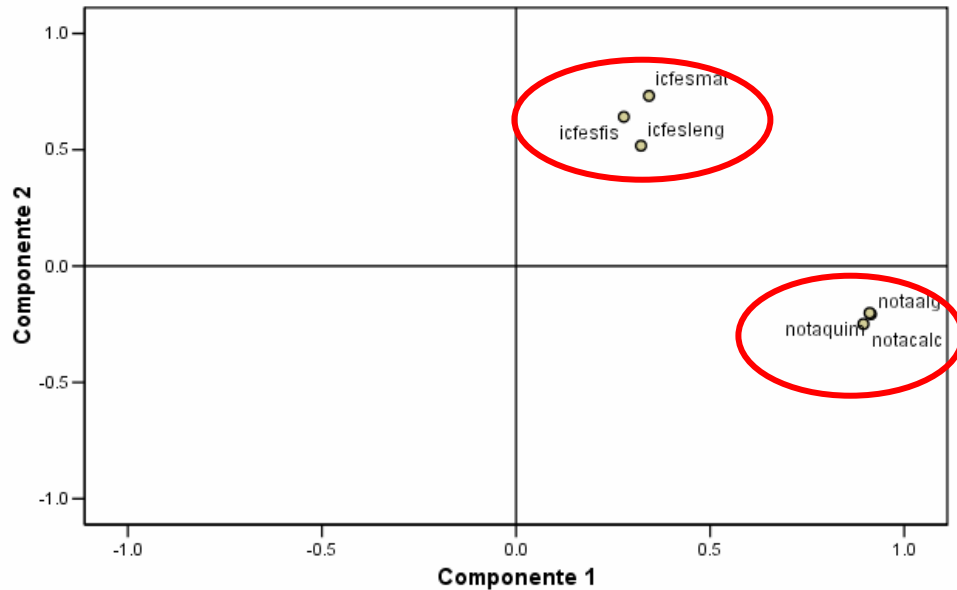


Figura 16. Gráfico de dos Componentes

Es importante recordar que las variables que “cargan alto” (es decir, existe alta correlación entre la variable y el factor) están altamente correlacionadas entre si y a su vez tienen poca correlación con las variables que carguen alto en otro factor, por la ortogonalidad presente entre los componentes.

El Análisis de Componentes Principales entrega como resultado las variables originales resumidas en dos Componentes o Factores (FACTOR NOTAS EN LA UIS, y FACTOR ICFES). Además ratifica que entre el puntaje del Icfes y el rendimiento en la universidad no existe relación alguna para este conjunto de datos.

	Componente	
	1	2
notacalc	,367	-,061
notaalg	,363	-,029
notaquim	,360	-,026
icfesmat	-,069	,548
icfesfis	-,068	,477
icfesleng	-,021	,397

Tabla 26. Matriz de componentes para el cálculo de las puntuaciones en las componentes

De la matriz de coeficientes de las puntuaciones en las componentes anteriores, son los coeficientes de cada Componente Principal, que son estimados con el valor propio y el vector propio de cada una. De esta manera, podemos deducir la siguiente relación entre componentes y variables:

#### FACTOR 1

$$NOTAS EN LA UIS = 0.367 * notacalc + 0.363 * notaalg + 0.360 * notaquim - 0.069 * icfesmat - 0.068 * icfesfis - 0.021 * icfesleng$$

#### FACTOR 2

$$PUNTAJE DE ICFES = -0.061 * notacalc - 0.029 * notaalg - 0.026 * notaquim + 0.548 * icfesmat + 0.477 * icfesfis + 0.397 * icfesleng$$

En cuanto al gráfico de dispersión en dos dimensiones de CP1 y CP2, observamos la gran variabilidad de las observaciones, teniendo un comportamiento de no correlación entre las variables, acumulándose la gran mayoría de los puntos en dos nubes donde la nube de la izquierda agrupa los estudiantes con notas no superiores a 2.0. Para el resto de los casos, la nube de la derecha

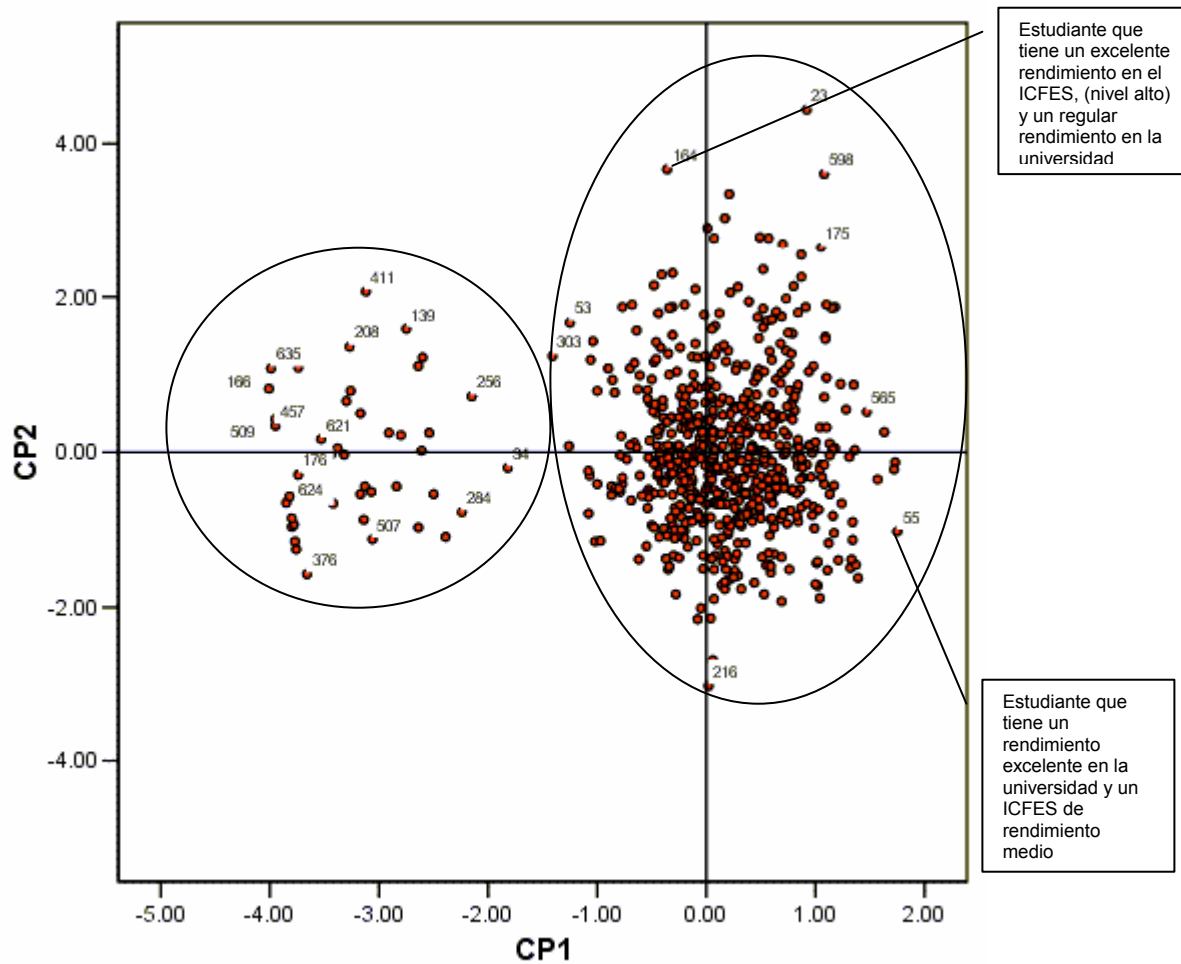


Figura 17. Diagrama Dispersión. Factor 1 vs. Factor 2

## **3.2 ANÁLISIS CON TABLAS DE CONTINGENCIA**

En la anterior sección trabajamos el análisis de Componentes Principales, con variables numéricas. En esta sección involucramos en el estudio variables que son importantes y muy valiosas en los resultados. Variables de tipo categóricas como sexo, colegio de procedencia, tipo de colegio, carrera, asociadas a las variables numéricas trabajadas anteriormente, arrojarán resultados muy interesantes para nuestro estudio. Cuando trabajamos con variables categóricas, los datos suelen organizarse en tablas de doble entrada, en las que cada entrada representa un criterio de clasificación (una variable categórica). Como resultado de esta clasificación, las frecuencias (el número o porcentaje de casos) aparecen organizadas en casillas que contienen información sobre la relación existente entre ambos criterios. A estas tablas de frecuencias se les llama *Tablas de Contingencia*.

A continuación presentaremos una serie de tablas de contingencia para contrastar las variables categóricas que involucran nuestro estudio, que permitan obtener información concluyente y decisiva en los procesos de admisiones y la situación en los primeros semestres en esta universidad. Es importante resaltar que los números que aparecen en las tablas no son puntuaciones, sino frecuencias absolutas (número de casos).

### **3.2.1 Estudio de la asociación entre género y resultados en las notas de cálculo 1 con los niveles de “aprobó” y “no aprobó”.**

A continuación presentaremos las observaciones encontradas en el análisis estadístico realizado entre los resultados obtenidos en las notas de cálculo 1, categorizando esta variable en estudiantes aprobados y no aprobados, para establecer la existencia o no de asociación con la variable género (modalidades de la variable: masculino y femenino). Vale aclarar que los aprobados son aquellos

estudiantes que obtuvieron su nota definitiva mayor o igual que 3.0, viéndola por primera vez y aprobándola directamente en su rendimiento durante el semestre o con habilitación. De la misma manera para los no aprobados, aquellos que obtuvieron su nota definitiva menor que 3.0.

Además si existe relación, sería interesante conocer en qué relación está el riesgo de no aprobación según el género del individuo.

### **Presentación de los datos**

POBLACIÓN	Estudiantes primer semestre de ingenierías 2004-2006
MUESTRA	1260
FUENTE	Oficina de registro académico UIS.

Tabla 27. Ficha Técnica

### **3.2.1.1 Resultados**

En la búsqueda de la existencia o no de relaciones entre el género y el resultado de cálculo 1, se obtuvieron los siguientes resultados:

		Resultados Cálculo I			
		Aprobó	No Aprobó	Total	
Género	Femenino	Frecuencia	361	75	436
		Frecuencia Esperada	341,9	94,1	436,0
		Género	82,8%	17,2%	100,0%
		Resultado Cálculo I	36,5%	27,6%	34,6%
		Total	28,7%	6,0%	34,6%
Masculino		Frecuencia	627	197	824
		Frecuencia Esperada	646,1	177,9	824,0
		Género	76,1%	23,9%	100,0%
		Resultado Cálculo I	63,5%	72,4%	65,4%
		Total	49,8%	15,6%	65,4%
Total		Frecuencia	988	272	1260
		Frecuencia Esperada	988,0	272,0	1260,0
		Género	78,4%	21,6%	100,0%
		Resultado Cálculo I	100,0%	100,0%	100,0%
		Total	78,4%	21,6%	100,0%

Tabla 28. Asociación entre género y notas de cálculo 1. (Fuente: Oficina de Admisión y Registro académico UIS)

Del total de estudiantes del género femenino, el 82.8% aprobaron cálculo I y el restante 17.2% no lo aprobaron. En igual forma, del total de estudiantes del género masculino, el 76.1% aprobaron cálculo I y el restante 23.9% no lo aprobaron. En la Tabla 28, notemos que el porcentaje de no aprobación del género masculino es de 23.9% siendo este mayor que el porcentaje para el género femenino 17.2% (ver figura 16); no obstante, debemos tener en cuenta los tamaños de las poblaciones entre los géneros y realizar un análisis que compare los dos grupos de estudiantes no aprobados (masculino-femenino) en donde se tenga en cuenta las proporciones con sus respectivas poblaciones.

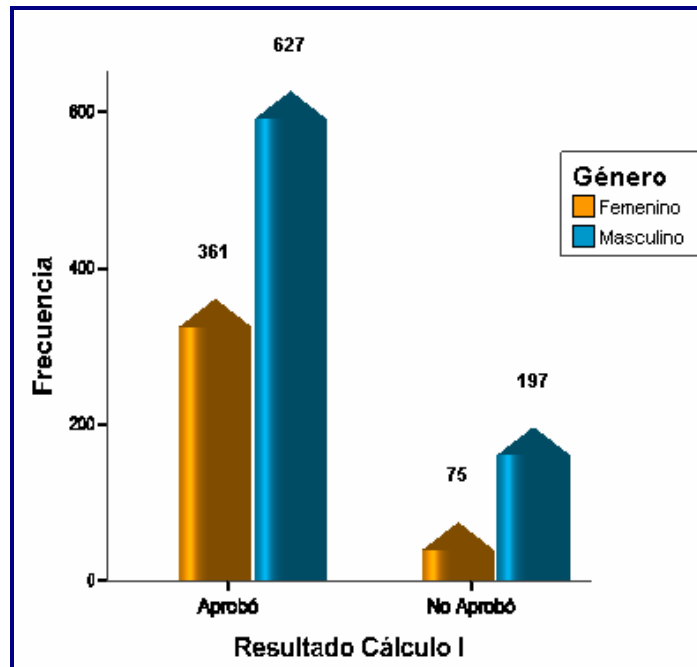


Figura 18. Diagrama de frecuencias de género y resultados cálculo 1

Del 100% (1260) de estudiantes, el 34.6% (436) son del género femenino y el 65.4% (824) son del género masculino. Observe la figura 16.

Probaremos la hipótesis de que no existen diferencias entre los géneros con sus resultados en cálculo 1.

Paso 1: Formulación de la Hipótesis

$H_0$  = No existe relación entre el género con sus resultados en cálculo 1.

$H_1$  = Existe relación entre el género con sus resultados en cálculo 1.

Paso 2: Nivel de significación

$\alpha = 0.05$

### Paso 3: Estadístico de Prueba

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-Cuadrado de Pearson	7,575	1	,006		
Corrección por Continuidad	7,184	1	,007		
Razón de Verosimilitud	7,785	1	,005		
Estadístico Exacto de Fisher				,006	,003
Número de Casos Válidos	1260				

Tabla 29. Prueba de Chi cuadrado. (Fuente: Oficina de Admisión y Registro académico UIS)

### Paso 4: Regla de Decisión

Para determinar la existencia de asociación entre el género y las notas de cálculo 1, observamos que el chi cuadrado calculado (7.57) para un nivel de confianza del 95% (0.05 nivel de significación) es mayor que el valor del chi cuadrado tabulado (3.85) (ver tabla chi cuadrado, significación: 0.05,  $gl=1$ ), se rechaza la hipótesis nula de independencia entre los factores, no rechazando que existe diferencias significativas entre el género con los resultados de cálculo1.

De acuerdo a los resultados, determinamos la existencia de una relación entre estas dos variables categóricas. Pasemos a identificar el riesgo según el género para la no aprobación de cálculo1.

### 3.2.1.2. Estimación de riesgo entre el género y los resultados de cálculo 1

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para sexo (Femenino / Masculino)	1,512	1,125	2,033
Para la cohorte Resultado Cálculo I = Aprobó	1,088	1,027	1,152
Para la cohorte Resultado Cálculo I = No Aprobó	,720	,566	,914
Número de Casos Válidos	1260		

Tabla 30. Estimación de riesgo de perder calculo entre los dos géneros. (Fuente: Oficina de Admisión y Registro académico UIS)

De acuerdo con los datos de la tabla 30, el riesgo relativo de *no aprobación* de las mujeres frente a los hombres es:

$$RR = \frac{75/436}{197/824} = \frac{0.172}{0.239} = 0.719$$

Donde (75) son los estudiantes que no aprobaron del género femenino entre el total de estudiantes del género femenino (436), y (197) son estudiante que no aprobaron del género masculino entre el total de estudiantes que aprobaron y no aprobaron del género masculino. De esta manera, al dividir estos riesgos relativos, arroja el riesgo relativo de no aprobación de las mujeres frente a los hombres.

De la misma forma, el riesgo relativo de *aprobación* de las mujeres frente a los hombres es el siguiente:

$$RR = \frac{361/436}{627/824} = \frac{0.827}{0.760} = 1.088$$

Donde (361) son los estudiantes del género femenino que aprobaron Cálculo I entre 436 que es el total de estudiantes de este género. De la misma forma 627 indica los estudiantes que aprobaron la asignatura del género masculino frente a 827 el total de estudiantes que aprobaron y no aprobaron.

Para cuantificar la importancia de la asociación, la forma es calcular el cociente de esos dos riesgos relativos calculados anteriormente, dando como resultado 1.512, y nos indica cuánto más probable es que ocurra que las mujeres aprueben cálculo 1 respecto a los hombres; es decir que en grupo de cinco estudiantes aprobados, se espera que dos sean hombres y tres sean mujeres, notando la ventaja del género femenino en cálculo 1.

Si miramos los resultados desde los estudiantes no aprobados, (resultacal1 = NA), el riesgo de encontrar tal desenlace entre las mujeres es menor que entre los hombres: por cada hombre no aprobado, podemos esperar 0.72 mujeres no aprobadas.

Entre los estudiantes que aprueban cálculo1, (resultacal1 = A), en proporción el número de mujeres que aprueban es 1.08 mayor que los hombres.

### **3.2.2 Estudio de la asociación entre puntaje de Icfes en Matemáticas y resultados en Cálculo 1 con los niveles de “aprobó” y “no aprobó”.**

A continuación presentaremos las observaciones encontradas en el análisis estadístico realizado entre los resultados obtenidos en cálculo 1, haciendo dentro de estos una distinción para los aprobados y no aprobados, para establecer si

existe asociación con la categoría Icfes en matemáticas (*nivel medio*: 40-70, *nivel alto*: mayores o iguales a 70). Además, si existe relación sería interesante conocer en qué relación está el riesgo de no aprobación según el rango del puntaje del Icfes en Matemáticas del individuo.

### Presentación de los datos

POBLACIÓN	Estudiantes primer semestre de ingenierías 2004-2006
MUESTRA	1244
FUENTE	Oficina de registro académico UIS.

Tabla 31. Ficha Técnica

#### 3.2.1.1 Resultados

En la búsqueda de la existencia o no de relaciones entre el puntaje de Icfes en Matemáticas y el resultado de cálculo 1, se obtuvo:

		Resultados Cálculo I		Total
		Aprobó	No Aprobó	
Icfes Matemáticas entre 40-70	Frecuencia	698	222	920
	Frecuencia esperada	718,8	201,2	920,0
	% de Icfes Matemáticas	75,9%	24,1%	100,0%
	% de Nota Cálculo	71,8%	81,6%	74,0%
	% del total	56,1%	17,8%	74,0%
mayor 70	Frecuencia	274	50	324
	Frecuencia esperada	253,2	70,8	324,0
	% de Icfes Matemáticas	84,6%	15,4%	100,0%
	% de Nota Cálculo	28,2%	18,4%	26,0%
	% del total	22,0%	4,0%	26,0%
Total	Frecuencia	972	272	1244
	Frecuencia esperada	972,0	272,0	1244,0
	% de Icfes Matemáticas	78,1%	21,9%	100,0%
	% de Nota Cálculo	100,0%	100,0%	100,0%
	% del total	78,1%	21,9%	100,0%

Tabla 32. Clasificación de los estudiantes según el puntaje del Icfes en Matemáticas y la nota de Cálculo 1. (Fuente: Oficina de Admisión y Registro académico UIS)

Del total de estudiantes que se están en el rango de 40-70 puntos en el Icfes de Matemáticas, el 75.9% aprobaron cálculo I y el restante 24.1% no aprobaron cálculo I. En este mismo orden de ideas, del total de estudiantes que obtuvieron puntajes mayores de 70 puntos en el Icfes en Matemáticas, el 84.6% aprobaron cálculo I y el restante 15.4% no aprobaron cálculo I. En la tabla 32, notemos que el porcentaje de no aprobación de estudiantes que tienen puntajes entre 40-70 puntos en el Icfes de matemáticas es de 24.1% siendo este mayor que el porcentaje para los que tienen puntaje mayor de 70 en el Icfes de matemáticas 15.4% (ver figura 17); a pesar de esto, debemos tener en cuenta los tamaños de las poblaciones entre los géneros y realizar un análisis que compare los dos grupos de estudiantes no aprobados (puntaje entre 40-70 y mayores de 70) en donde se tenga en cuenta las proporciones con sus respectivas poblaciones.

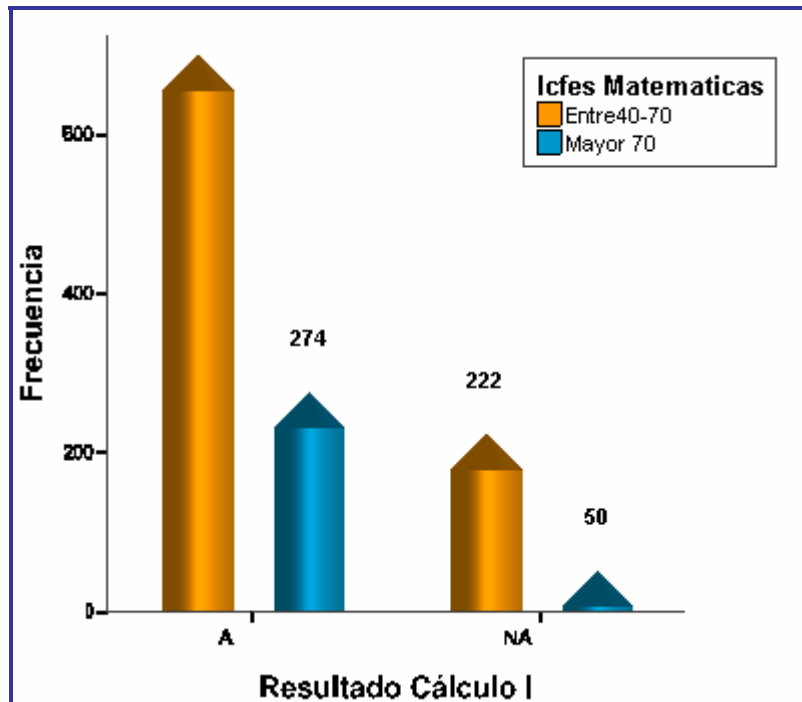


Figura 19. Diagrama de frecuencias de Icfes en Matemáticas y Resultados cálculo 1

Del 100% (1244) de estudiantes, el 74% (920) tienen puntajes de 40-70 en el Icfes de Matemáticas y el 26% (324) pertenecen al grupo de mayor de 70 puntos en el Icfes de Matemáticas. Observe la figura 17.

Probaremos la hipótesis de que no existen diferencias entre los puntajes de Icfes con los resultados en cálculo 1.

Paso 1: Formulación de la Hipótesis

$H_0$  = No existe relación entre los puntajes de Icfes en Matemáticas con los resultados en Cálculo I.

$H_1$  = Existe relación entre los puntajes de Icfes en Matemáticas con los resultados en Cálculo I

## Paso 2: Nivel de significación

$$\alpha = 0.05$$

## Paso 3: Estadístico de Prueba

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	10,612	1	,001		
Corrección por continuidad	10,109	1	,001		
Razón de verosimilitud	11,211	1	,001		
Estadístico exacto de Fisher				,001	,001
Asociación lineal por lineal					
N de casos válidos	1244				

Tabla 33. Prueba Chi cuadrado. (Fuente: Oficina de Admisión y Registro académico UIS)

## Paso 4: Regla de Decisión

Para determinar la existencia de asociación entre el puntaje de Icfes en matemáticas y las notas de cálculo 1, observamos que el chi cuadrado calculado (10.61) para un nivel de confianza del 95% (5% nivel de significación) es mayor que el valor del chi cuadrado tabulado (3.85) (ver tabla Chi cuadrado, significación: 5%,  $gl=1$ ), se rechaza la hipótesis nula de independencia entre los factores, aceptando por tanto que existe diferencias significativas entre los puntajes de Icfes en matemáticas con los resultados de cálculo1, con una probabilidad de cometer el error tipo I (rechazar  $H_0$  de forma errónea) 0.001.

Con la seguridad de la existencia de una relación entre estas dos variables categóricas, pasemos a identificar el riesgo según el rango de puntaje en el Icfes de Matemáticas para la no aprobación de cálculo1.

### 3.2.1.2 Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para icfesmatemat (entre 40-70 / mayor 70 )	,574	,410	,804
Para la cohorte Resultado de Cálculo I = Aprobó	,897	,846	,952
Para la cohorte Resultado de Cálculo I = No Aprobó	1,564	1,182	2,068
N de casos válidos	1244		

Tabla 34. Estimaciones de riesgo entre Puntaje Icfes en Matemáticas y Resultados de Cálculo I.  
(Fuente: Oficina de Admisión y Registro académico UIS)

De acuerdo con los datos de la tabla anterior, el riesgo relativo de *no aprobación* del rango Icfes en matemáticas (40-70) sobre el rango (mayor de 70) es:

$$RR = \frac{222/920}{50/324} = \frac{0.24}{0.15} = 1.56$$

De la misma forma, el riesgo relativo de *aprobación* del rango Icfes en matemáticas (40-70) sobre el rango (mayor de 70) es:

$$RR = \frac{698/920}{274/324} = \frac{0.75}{0.84} = 0.89$$

Para cuantificar la importancia de la asociación, la forma es calcular el cociente de esos dos riesgos relativos calculados anteriormente, dando como resultado 0.574,

y nos indica cuanto más probable es que ocurra que los estudiantes que sacan entre 40-70 puntos en matemáticas en el Icfes aprueben sobre los que sacan mas de 70 puntos en el Icfes en matemáticas.

Si miramos los resultados desde los estudiantes no aprobados, (notacalc = NA), el riesgo de encontrar tal desenlace entre los que sacan entre 40-70 puntos en matemáticas en el Icfes es mayor que el estudiante que saca mas de 70 puntos en matemáticas: por cada estudiante que esté en el rango de mas de 70 puntos en el Icfes matemáticas no aprobado, podemos esperar 1.56 estudiantes que están el rango de 40-65 puntos en el Icfes en matemáticas no aprobados.

Claramente les va mejor a los estudiantes que estén en el rango de más de 70 puntos en el Icfes en Matemáticas, ya que la probabilidad de que ellos pierdan con relación al otro grupo en estudio (rango 40-70), es menor, y mayor probabilidad que aprueben.

### **3.2.3 Estudio de la asociación entre puntaje de Icfes en Lenguaje y resultados en Cálculo 1 con los niveles de “aprobó” y “no aprobó”.**

Analizaremos si existe alguna relación entre el puntaje del Icfes en Lenguaje y la nota de Cálculo 1 en el primer semestre en la Universidad Industrial de Santander. Este puntaje lo tomamos en dos grupos de clasificación como lo realiza el ICSES en su valoración de escalas: Medio (40-70), alto (mayor o igual a 70).

## Presentación de los datos

UNIVERSO	Estudiantes primer semestre de ingenierías
MUESTRA	126
FUENTE	Oficina de registro académico UIS.

Tabla 35. Ficha Técnica

### 3.2.3.1 Resultados

			Resultado Cálculo 1		Total
			Aprobó	No Aprobó	
Icfes Lenguaje	Entre 40-70	Recuento	806	229	1035
		Frecuencia esperada	810,8	224,3	1035,0
		% de icfeslenguaje	77,9%	22,1%	100,0%
		% de notacalc	81,7%	83,9%	82,1%
		% del total	64,0%	18,2%	82,1%
	Mayor 70	Recuento	181	44	225
		Frecuencia esperada	176,3	48,8	225,0
		% de icfeslenguaje	80,4%	19,6%	100,0%
		% de notacalc	18,3%	16,1%	17,9%
		% del total	14,4%	3,5%	17,9%
Total		Recuento	987	273	1260
		Frecuencia esperada	987,0	273,0	1260,0
		% de icfeslenguaje	78,3%	21,7%	100,0%
		% de notacalc	100,0%	100,0%	100,0%
		% del total	78,3%	21,7%	100,0%

Tabla 36. Clasificación de los estudiantes según el puntaje del Icfes en Lenguaje y la nota de Cálculo 1. (Fuente: Oficina de Admisión y Registro académico UIS)

Del total de estudiantes que se están en el rango de 40-70 puntos en el Icfes de Lenguaje, el 77.9% aprobaron cálculo I y el restante 22.1% no aprobaron cálculo I. En este mismo orden de ideas, del total de estudiantes que obtuvieron puntajes mayores de 70 puntos en el Icfes en Lenguaje, el 80.4% aprobaron cálculo I y el restante 19.6% no aprobaron cálculo I. En la tabla 36, notemos que el porcentaje de no aprobación de estudiantes que tienen puntajes entre 40-70 puntos en el Icfes de Lenguaje es de 22.1% siendo este mayor que el porcentaje para los que tienen puntaje mayor de 70 en el Icfes de matemáticas 19.6% (ver figura 18); comparando los porcentajes en los dos niveles del resultado en el Icfes en lenguaje medio y alto, con sus respectivos casos de no aprobados en cálculo 1, encontramos que es mayor para el caso de el nivel medio; a pesar de esto es necesario aplicar una prueba donde se considere si la diferencia entre estos porcentajes me permite argumentar la existencia de una relación entre el nivel de los resultados del ICFES en lenguaje con los resultados en cálculo 1.

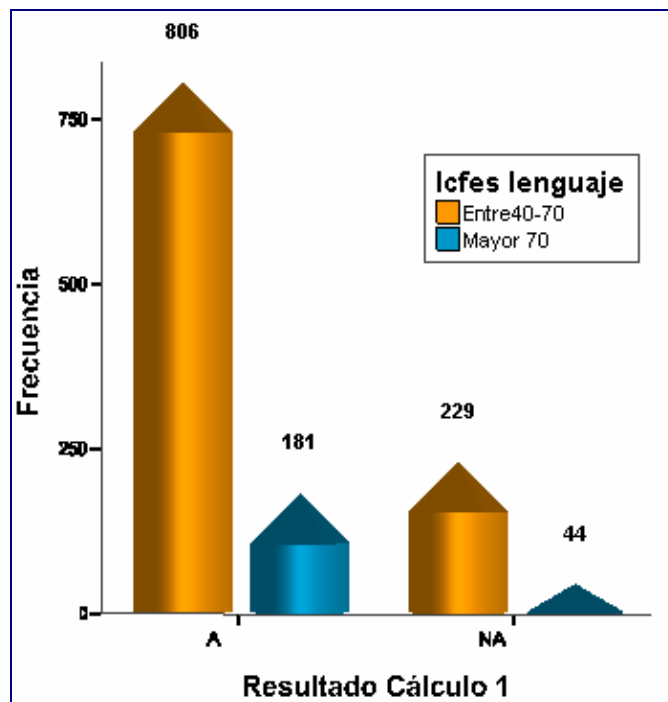


Figura 20. Diagrama de frecuencias de Icfes en Lenguaje y Resultados cálculo 1

Del 100% (1260) de estudiantes, el 82.1% (1035) tienen puntajes de 40-70 en el Icfes de Lenguaje y el 17.9 (225) pertenecen al grupo de mayor de 70 puntos en el Icfes de Lenguaje. Observe la figura 18.

Probaremos la hipótesis de que no existen diferencias entre los puntajes de Icfes en Lenguaje con los resultados en cálculo 1.

### Paso 1: Formulación de la Hipótesis

$H_0$  = No existe relación entre los puntajes de Icfes en Lenguaje con los resultados en Cálculo I

$H_1$  = Existe relación entre los puntajes de Icfes en Lenguaje con los resultados en Cálculo I

### Paso 2: Nivel de significación

$$\alpha = 0.05$$

### Paso 3: Estadístico de Prueba

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,719	1	,396		
Corrección por continuidad	,576	1	,448		
Razón de verosimilitud	,733	1	,392		
Estadístico exacto de Fisher				,423	,225
Asociación lineal por lineal					
N de casos válidos	1260				

Tabla 37. Prueba Chi-Cuadrado. (Fuente: Oficina de Admisión y Registro académico UIS)

#### Paso 4: Regla de Decisión

Para determinar la existencia de asociación entre el puntaje de Icfes en lenguaje y las notas de cálculo 1, observamos que el Chi cuadrado calculado (0.719) para un nivel de confianza del 95% (5% nivel de significación) es menor que el valor del Chi-cuadrado tabulado (3.84) (ver tabla Chi cuadrado, significación: 5%,  $gl=1$ ), se acepta la hipótesis nula de independencia entre los factores, aceptando por tanto que no existe diferencias significativas entre los puntajes de Icfes en lenguaje con los resultados de cálculo 1. Este resultado indica que realmente el Icfes en lenguaje no es tan determinante a la hora de tomar decisiones de selección de aspirantes, ya que no garantiza que se vaya a tener un buen rendimiento en las materias de cálculo; pero no se quiere decir que podemos dar porcentado este resultado, ya que debemos sustentarlo buscando relaciones con otras materias (Química ó Álgebra)

#### **3.2.4 Estudio de la asociación entre puntaje de Icfes en Lenguaje y resultados en Química Básica con los niveles de “aprobó” y “no aprobó”.**

La conclusión que surgió del anterior estudio (4.2.3) sobre la no existencia de la relación entre el rendimiento de los estudiantes en el primer semestre en Cálculo y el resultado del ICFES en el área de lenguaje, nos motivó a observar si esta independencia se conserva para la asignatura de Química Básica con la misma prueba en el examen de estado o si por el contrario aparece relación.

## Presentación de los datos

POBLACIÓN	Estudiantes primer semestre de ingenierías 2004-2006
MUESTRA	685
FUENTE	Oficina de registro académico UIS.

Tabla 38. Ficha Técnica

### 3.2.4.1 Resultados

		NOTA QUIMICA		Total	
		Aprobó	No Aprobó		
ICFES LENGUAJE	ENTRE40-70	Frecuencia	394	172	566
		Frecuencia esperada	399,9	166,1	566,0
		% de ICFESLENGUAJE	89,6%	30,4%	100,0%
		% de NOTAQUIMICA	81,4%	85,6%	82,6%
		% del total	57,5%	25,1%	82,6%
MAYORES 70		Frecuencia	90	29	119
		Frecuencia esperada	84,1	34,9	119,0
		% de ICFESLENGUAJE	75,6%	24,4%	100,0%
		% de NOTAQUIMICA	18,6%	14,4%	17,4%
		% del total	13,1%	4,2%	17,4%
Total		Frecuencia	484	201	685
		Frecuencia esperada	484,0	201,0	685,0
		% de ICFESLENGUAJE	70,7%	29,3%	100,0%
		% de NOTAQUIMICA	100,0%	100,0%	100,0%
		% del total	70,7%	29,3%	100,0%

Tabla 39. Clasificación de los estudiantes según el puntaje del Icfes en Lenguaje y la nota de Química. (Fuente: Oficina de Admisión y Registro académico UIS)

En la Tabla 39, se representa la frecuencia de los puntajes de Icfes en Lenguaje en los mismos niveles utilizados en el anterior estudio 4.2.3 con la nota en Química Básica categorizadas entre “aprobó” o “no aprobó”.

Del total de estudiantes que se están en el rango de 40-70 puntos en el Icfes de Lenguaje, el 69.6% aprobaron Química I y el restante 30.4% no aprobaron Química I. En este mismo orden de ideas, del total de estudiantes que obtuvieron puntajes mayores de 70 puntos en el Icfes en Lenguaje, el 75.6% aprobaron Química I y el restante 24.4% no aprobaron Química I. En la tabla 39, notemos que el porcentaje de no aprobación de estudiantes que tienen puntajes entre 40-70 puntos en el Icfes de Lenguaje es de 30.4% siendo este mayor que el porcentaje para los que tienen puntaje mayor de 70 en el Icfes de matemáticas 24.4% (ver figura 19); a pesar de esto, comparando los porcentajes en los dos niveles del resultado en el Icfes en lenguaje medio y alto, con sus respectivos casos de no aprobados en Química, encontramos que es mayor para el caso de el nivel medio; nuevamente se hace necesario aplicar una prueba donde se considere si la diferencia entre estos porcentajes me permite argumentar la existencia de una relación entre el nivel de los resultados del ICFES en lenguaje con los resultados en Química Básica.

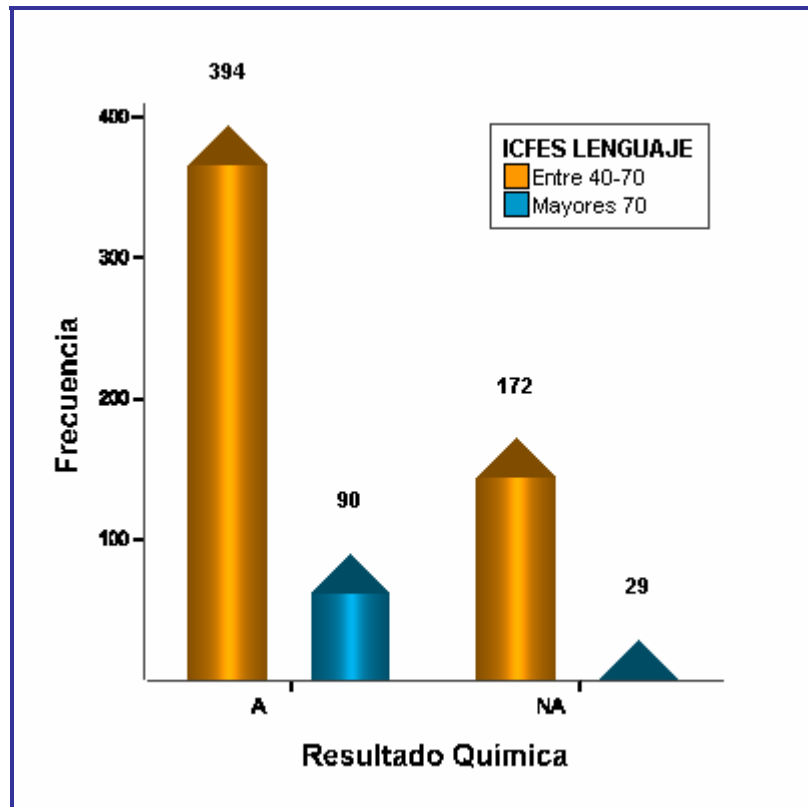


Figura 21. Diagrama de frecuencias de Icfes en Lenguaje y Resultados en Química

Probaremos la hipótesis de que no existen diferencias entre los puntajes del Icfes en Lenguaje a un nivel de significación del 5% con las notas de Química Básica.

Paso 1: Formulación de la Hipótesis

$H_0$  = No existe relación entre los puntajes Icfes en Lenguaje con los resultados en Química Básica.

$H_1$  = Existe relación entre los puntajes Icfes en Lenguaje con los resultados en Química Básica.

Paso 2: Nivel de significación

$$\alpha = 0.05$$

### Paso 3: Estadístico de Prueba

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	1,718	1	,190		
Corrección por continuidad	1,440	1	,230		
Razón de verosimilitud	1,769	1	,184		
Estadístico exacto de Fisher				,223	,114
Asociación lineal por lineal					
N de casos válidos	685				

Tabla 40. Prueba Chi-cuadrado (Fuente: Oficina de Admisión y Registro académico UIS)

El valor obtenido por el Chi cuadrado es de 1.718, siendo este menor que el valor tabulado en Chi cuadrado con un grado de libertad y un nivel de significación del 5%, que es de 3.84, lo cual nos permite concluir que no existe evidencia suficiente para rechazar la hipótesis nula planteada.

#### **3.2.5 Estudio de la asociación entre la categoría del colegio de procedencia es decir “privado” o “público” y resultados en Calculo 1 con las categorías de “aprobó” y “no aprobó”.**

Realizamos esta asociación, ya que nos pareció interesante sacar resultados de los estudiantes que proceden de diferentes establecimientos públicos o privados y mirar si es determinante la categoría del colegio del cual son egresados con el rendimiento en la universidad en diferentes materias, en este caso cálculo 1. En los datos encontramos que hay 178 colegios públicos en estudio y 293 instituciones privadas.

Hay que tener en cuenta los diferentes niveles de calidad académica existentes dentro de cada colegio sea público o privado; es por esto que los resultados que se obtendrán no hablan de que algún colegio tendrá el mismo nivel académico que otro a pesar de tener la misma categoría, pero si notaremos un comportamiento general de los resultados a la tendencia de el rendimiento de los estudiantes en la universidad.

Llámesese colegio público, aquel que es subsidiado por el gobierno municipal y al colegio privado donde los costos educativos son pagos por los representantes legales del estudiante.

### **Presentación de los datos**

POBLACIÓN	Estudiantes primer semestre de ingenierías 2004-2006
MUESTRA	1258
FUENTE	Oficina de registro académico UIS.

Tabla 41. Ficha Técnica

### 3.2.5.1 Resultados

		Resultado Cálculo I			
		Aprobó	No Aprobó	Total	
Tipo de colegio	Privado	Frecuencia	462	113	575
		Frecuencia esperada	450,7	124,3	575,0
		% de colegio	80,3%	19,7%	100,0%
		% de notacalc	46,9%	41,5%	45,7%
		% del total	36,7%	9,0%	45,7%
Público		Frecuencia	524	159	683
		Frecuencia esperada	535,3	147,7	683,0
		% de colegio	76,7%	23,3%	100,0%
		% de notacalc	53,1%	58,5%	54,3%
		% del total	41,7%	12,6%	54,3%
Total		Frecuencia	986	272	1258
		Frecuencia esperada	986,0	272,0	1258,0
		% de colegio	78,4%	21,6%	100,0%
		% de notacalc	100,0%	100,0%	100,0%
		% del total	78,4%	21,6%	100,0%

Tabla 42. Clasificación de los estudiantes según colegio de procedencia con resultados en cálculo 1. (Fuente: Oficina de Admisión y Registro académico UIS)

Analizaremos en forma general, los porcentajes no aprobación de los estudiantes que vienen de los colegios categorizados públicos o privados. De los 1258 estudiantes bajo estudio, hay 19.7% que no aprobaron cálculo 1 que proceden de colegios privados contra un 23.3% que no aprobaron procedentes de colegios públicos. La diferencia al parecer no es muy grande, lo cual presume la no relación entre las variables de estudio, pero se hace necesario ratificar con una prueba de hipótesis para dar soporte a esta conjetura.

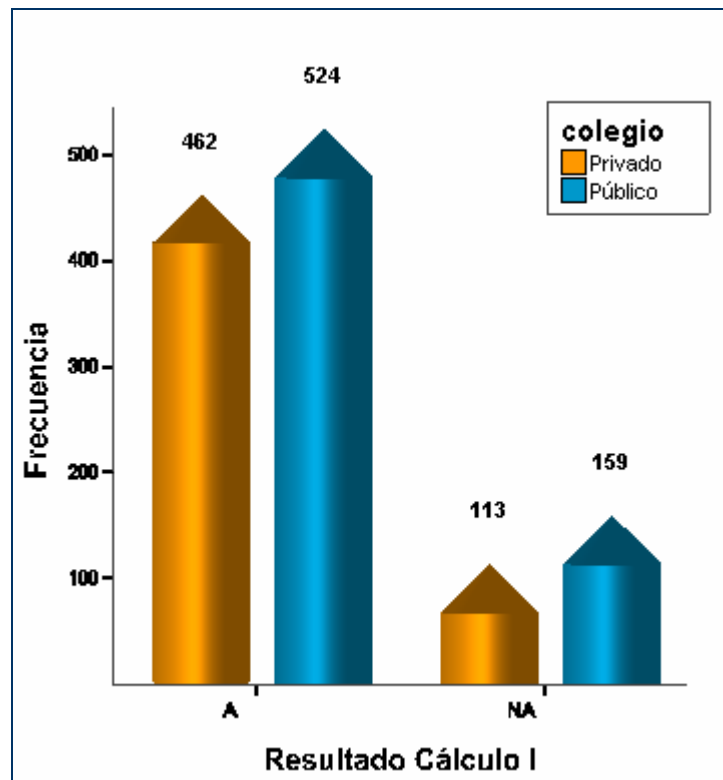


Figura 22. Diagrama de frecuencias de Tipo de Colegio de procedencia con el Resultados en Cálculo I

Del 100% (1258) de estudiantes, el 45.7% (575) provienen de colegios privados y el 54.3 (683) provienen de colegios públicos. Observe la figura 20.

Probaremos la siguiente hipótesis para corroborar la asociación entre las variables:

$H_0$  = No existen diferencias de proceder de un colegio público o privado a un nivel de significación del 5% con el rendimiento en cálculo 1 en la UIS.

$H_1$  = Existen diferencias de proceder de un colegio público o privado a un nivel de significación del 5% con el rendimiento en cálculo 1 en la UIS.

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	2,424	1	,119		
Corrección por continuidad	2,215	1	,137		
Razón de verosimilitud	2,435	1	,119		
Estadístico exacto de Fisher				,130	,068
Asociación lineal por lineal					
N de casos válidos	1258				

Tabla 43. Estadístico Chi cuadrado. (Fuente: Oficina de Admisión y Registro académico UIS)

El valor obtenido por el Chi cuadrado es de 2.42, siendo este menor que el valor tabulado en Chi cuadrado con un grado de libertad y un nivel de significación del 5%, que es de 3.84, lo cual nos permite concluir que no existe evidencia suficiente para rechazar la hipótesis nula planteada.

Es interesante este resultado, pues nos da a entender que no existe relación alguna de proceder de un colegio público o privado con el rendimiento del estudiante en el primer semestre en la asignatura de cálculo 1. Es de clarificar que cada colegio da pautas y bases de estudio para que el estudiante tenga un buen rendimiento en la Universidad, pues para la aprobación de cálculo 1, no es factor determinante el colegio de procedencia.

### **3.2.6 Estudio de la asociación entre la categoría del colegio de procedencia es decir “privado” o “público” y resultados en Álgebra Lineal con las categorías de “aprobó” y “no aprobó”.**

## Presentación de los datos

POBLACIÓN	Estudiantes primer semestre de ingenierías 2004-2006
MUESTRA	1026
FUENTE	Oficina de Registro Académico UIS.

Tabla 44. Ficha Técnica

### 3.2.6.1 Resultados

En la búsqueda de la existencia o no de relaciones entre el tipo de colegio y el resultado en Álgebra Lineal, se obtuvo:

		Resultado de Álgebra			
		Aprobó	No Aprobó	Total	
Tipo de colegio	Privado	Frecuencia	349	110	459
		Frecuencia esperada	358,3	100,7	459,0
		% de tipodecolegio	76,0%	24,0%	100,0%
		% de rendimientoalgebra	43,6%	48,9%	44,7%
		% del total	34,0%	10,7%	44,7%
Público		Frecuencia	452	115	567
		Frecuencia esperada	442,7	124,3	567,0
		% de tipodecolegio	79,7%	20,3%	100,0%
		% de rendimientoalgebra	56,4%	51,1%	55,3%
		% del total	44,1%	11,2%	55,3%
Total		Frecuencia	801	225	1026
		Frecuencia esperada	801,0	225,0	1026,0
		% de tipodecolegio	78,1%	21,9%	100,0%
		% de rendimientoalgebra	100,0%	100,0%	100,0%
		% del total	78,1%	21,9%	100,0%

Tabla 45. Clasificación de los estudiantes según colegio de procedencia con resultados en Álgebra Lineal. (Fuente: Oficina de Admisión y Registro académico UIS)

Del total de estudiantes que proceden de colegios privados, el 76% aprobaron Álgebra Lineal y el restante 24% no aprobaron Álgebra Lineal. En este mismo orden de ideas, del total de estudiantes que proceden de colegios públicos, 79.7% aprobaron Álgebra Lineal I y el restante 20.3% no aprobaron Álgebra Lineal. En la tabla 45, notemos que el porcentaje de no aprobación de estudiantes que proceden de colegios privados es de 24% siendo este mayor que el porcentaje para los que proceden de instituciones públicas 20.3% (ver figura 21); a pesar de esto, debemos tener en cuenta los tamaños de las poblaciones entre los géneros y realizar un análisis que compare los dos grupos de procedencia de planteles educativos donde se tenga en cuenta las proporciones con sus respectivas poblaciones.

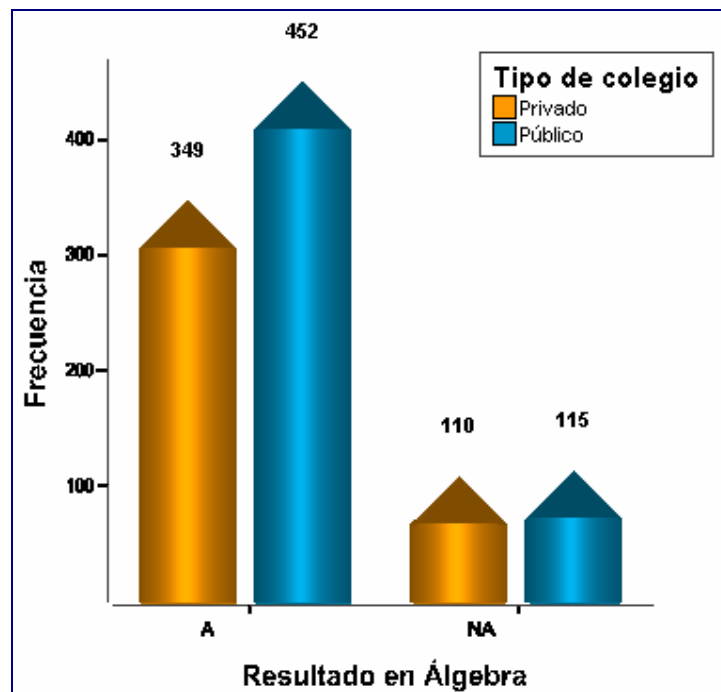


Figura 23. Diagrama de frecuencias de Tipo de Colegio de procedencia con el Resultados en Álgebra Lineal

Del 100% (1026) de estudiantes, el 44.7% (459) provienen de colegios privados y el 55.3% (567) provienen de colegios públicos. Observe la figura 21.

Probaremos la hipótesis de que no existen diferencias entre el tipo de colegio de procedencia con los resultados en Álgebra Lineal.

### Paso 1: Formulación de la Hipótesis

$H_0$  = No existe relación entre el tipo de plantel de procedencia con los resultados en Álgebra Lineal.

$H_1$  = Existe relación entre el tipo de plantel de procedencia con los resultados en Álgebra Lineal

### Paso 2: Nivel de significación

$$\alpha = 0.05$$

### Paso 3: Estadístico de Prueba

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	2,010	1	,156		
Corrección por continuidad	1,800	1	,180		
Razón de verosimilitud	2,003	1	,157		
Estadístico exacto de Fisher				,172	,090
Asociación lineal por lineal					
N de casos válidos	1026				

Tabla 46. Estadístico Chi-cuadrado. (Fuente: Oficina de Admisión y Registro académico UIS)

#### Paso 4: Regla de Decisión

Para determinar la existencia de asociación entre el plantel de procedencia y el resultados en Álgebra Lineal, observamos que el Chi cuadrado calculado (2.010) para un nivel de confianza del 95% (0.05 nivel de significación) es mayor que el valor del Chi cuadrado tabulado (3.84) (ver tabla Chi cuadrado, significación: 0.05,  $gl=1$ ), se rechaza la hipótesis nula de independencia entre los factores, no rechazando que existe diferencias significativas entre el tipo de colegio que proceden los estudiantes con los resultados de Álgebra Lineal, con una probabilidad de cometer el error tipo I (rechazar  $H_0$  de forma errónea) 0.156.

#### 4. CONCLUSIONES Y RECOMENDACIONES

Con la técnica de Componentes Principales, claramente se destacó la absoluta independencia entre los resultados en el primer semestre de las materias del ciclo básico de los ingenieros de la UIS, con el puntaje del Icfes en matemáticas, Física y Lenguaje. Este resultado es sustentado por el hecho de encontrar que el conjunto de variables en estudio fue clasificado en dos componentes ortogonales, independientes entre sí. Las variables de Cálculo, Álgebra Lineal y Química, se reunieron en una componente que nombramos Factor UIS, y los puntajes de Matemáticas, Física y Lenguaje en una componente que denominamos Factor ICFES, quedando en la otra componente.

De hecho existe un inconveniente a la hora de la selección de los aspirantes a los diferentes programas académicos de ingenierías de la UIS, dada la poca relación que existe entre el criterio de el examen del ICFES con el rendimiento del estudiante en el primer semestre; este no debería ser el único referente de admisión, si pensamos en personal capacitado de admisión y tener un excelente desempeño del estudiante en nuestra alma mater. Esta afirmación se sustenta en las conclusiones resultantes de esta investigación. De seguir con este criterio, se genera como la idea que tal vez se está negando la posibilidad de ingreso a las Facultades de fisicoquímica y Físico mecánica de la UIS, algunos postulantes con bajos puntajes en el Icfes y que bien podría tener un buen rendimiento en materias del ciclo básico en la Universidad.

En general, se debería extender el estudio a otros programas académicos si existe relación entre el puntaje de Icfes y el rendimiento en el primer semestre. De esta manera se podría determinar, la conveniencia o no, de exigir las pruebas de ICFES como único criterio de ingreso a cualquier programa de la UIS. Además

debe averiguarse el efecto que otras variables, como por ejemplo, la metodología de la enseñanza, la forma de evaluación, el aspecto socioeconómico, el estrato, que puedan tener efecto en el buen rendimiento del estudiante.

Realizando el análisis por medio de las tablas de contingencia, encontramos la existencia de la relación entre el género del individuo y su rendimiento en el área de Cálculo I. Además, el factor de riesgo de no aprobación de Cálculo I es favorable para el género femenino en un 0.72. Es decir, encontramos que es más factible la no aprobación del género masculino en cálculo I.

De la relación entre Puntaje de Icfes en Matemáticas con el rendimiento en Cálculo I en la Universidad, encontramos una amplia relación; siendo esta favorable para aquellos estudiantes que ingresaron a la universidad con su Icfes en matemáticas dentro del considerado nivel alto (mayor de 70 puntos). Además, si hacemos la comparación de un estudiante en nivel medio en Puntaje Icfes en Matemáticas (40-69 puntos), con otro de nivel alto en la misma área, existe el doble de riesgo que el primero de estos individuos repruebe Cálculo I en su primer semestre en la universidad comparado con uno que este en nivel alto.

Resulta contradictoria la existencia de esta relación; ya que en el estudio de las componentes principales registramos la independencia entre los dos grupos de variables (Factor UIS) y (Factor ICFES). Más allá de esto, nótese que la variable Icfes en Matemáticas y Resultado de Cálculo I, son variables que están directamente relacionadas con el pensamiento matemático de cada individuo. Por tanto, el resultado obtenido es razonable a pesar del resultado del estudio multivariado con componentes principales.

Como se vio en los criterios de admisión, el área del Icfes en lenguaje, es tenida en cuenta en el primer criterio para ingresar a las dos Facultades de Ingenierías; al

realizar el estudio entre esta variable con dos materias Química Básica y Cálculo I, resultó que no existe relación entre el desempeño en el Icfes en lenguaje con el rendimiento en estas dos materias de la Universidad. Es decir, un individuo que ingrese a la universidad en un nivel alto en el área de lenguaje, no presenta ventaja sobre otro individuo que se ubica en el nivel medio de esta área en las materias descritas anteriormente. Como sugerencia, reevaluar el criterio 1 de admisión donde se encuentra el área de lenguaje en Icfes.

Analizando el plantel de procedencia de los estudiantes (público o privado) con el rendimiento de materias (cálculo I y Álgebra Lineal), encontramos que no existe relación alguna; a pesar de esto, no debemos omitir el hecho que este estudio considera a todos los colegios privados por igual sin distinguir diferencias que existen entre ellos como por ejemplo la exigencia académica, infraestructura, intensidad horaria, etc. Esta misma situación sucede con los colegios públicos.

## BIBLIOGRAFIA

Bueno, Y. (2004). *Una aplicación del análisis por componentes principales en los indicadores de gestión de las universidades públicas*. Trabajo de grado Licenciatura en Matemáticas. Universidad Industrial de Santander, Bucaramanga

Carvajal, P. (2004). *Búsqueda de la relación entre áreas ICFES en matemáticas, física, lenguaje y rendimiento en matemáticas I y matemáticas II a través del análisis de componentes principales*. *Revista de investigación Scientia et Técnica Año X, No. 26. p. 133-138*. Universidad Tecnológica de Pereira

Consejo Académico. Acuerdo No 207 de 2007. Universidad Industrial de Santander. Bucaramanga, Colombia.

Díaz, L. G. (2002). *Estadística Multivariada: Inferencia y métodos*. Universidad Nacional de Colombia. Bogotá, Colombia.

Dirección de Admisiones y Registro Académico (2000). *Sistema de admisión*. Universidad Industrial de Santander. Bucaramanga, Colombia.

Entralgo, E. (2003). *Apuntes de análisis Multivariado. Componentes Principales*. Universidad de los Andes, Mérida, Venezuela.

Gaona, M. (1998). *Examen de Estado. Admisión a la Educación Superior*. Ministerio de Educación Nacional. Bogotá, Colombia.

Manly, B. (2000). *Multivariate Statistical Methods, A primer*. Chapman an hall, New York, 2000.

Marín, J. (2000). *Análisis de Componentes Principales*. España. Recuperado el 13 de septiembre de 2007 de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema3am.pdf> -

Mendoza, H. (1991). *Predicción del rendimiento académico final de los estudiantes de pre-grado de la Facultad de Ciencias Físico-Mecánica de la Universidad Industrial de Santander*. Trabajo de grado Licenciatura en Matemáticas. Universidad Industrial de Santander, Bucaramanga.

Molinero, L. (2001). *Odds ratio, riesgo relativo y número necesario a tratar*. España. Recuperado el 10 de noviembre de 2007 de <http://www.seh-lelha.org/odssratio.htm>

Peña, D. (2002). *Análisis de datos multivariantes*. Mc Graw Hill , Madrid.

Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS*. Pearson Prentice Hall, Madrid.

Pita, S. (1997). *Determinación de factores de riesgo*. España. Recuperado el 10 de septiembre de 2007 de [http://www.fisterra.com/mbe/investiga/3f\\_de\\_riesgo/3f\\_de\\_riesgo.asp](http://www.fisterra.com/mbe/investiga/3f_de_riesgo/3f_de_riesgo.asp)

Universidad de Cádiz. (2002). *Análisis de variables categóricas. El procedimiento de tablas de contingencia*. España. Recuperado el 15 de octubre de 2007 de <http://www2.uca.es/serv/ai/formacion/spss/Pantalla/12contin.pdf>

Universidad Industrial de Santander. *Proceso de Admisión Primer periodo académico del año 2008*. Dirección de admisión y Registro académico. Recuperado el 12 de septiembre de 2007 de <http://www.uis.edu.co/admisiones>.

Vicéns y Medina. (2005). *Análisis de datos cualitativos*. Canadá. Recuperado el 24 de septiembre de 2007 de <http://www.er.uqam.ca/nobel/ieim/IMG/pdf/metho-2002-04-armony.pdf>