Capítulo 4

4. ANÁLISIS ESTADÍSTICO MULTIVARIADO DE LA POBLACIÓN INVESTIGADA

4.1 Introducción

En este capítulo se realiza un análisis estadístico multivariado de los datos, el cual consiste en estudiar 2 o más variables a la vez con la finalidad de determinar si existe alguna relación entre las variables. Las técnicas multivariadas que se utilizarán en el estudio son: tablas de contingencias, análisis de componentes principales, correlación lineal, análisis de varianza y correlación canónica.

4.2. Matriz de Datos o Tablas de Datos

La matriz de datos es una tabla rectangular que contiene toda la información que dispone la muestra, determinada previamente; es una matriz de n filas que corresponden al número de individuos(532 estudiantes) y p columnas que son las 42 variables estudiadas de los 532 individuos investigados como se muestra en la tabla VIII.

CUADRO 4.1: FORMATO DE UNA MATRIZ DE DATOS

$$X = \begin{bmatrix} X_{11} & X_{12} & . & . & . & X_{1p} \\ X_{21} & X_{22} & . & . & . & X_{2p} \\ X_{31} & X_{32} & . & . & . & X_{3p} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ X_{n1} & X_{n2} & & & & X_{np} \end{bmatrix}$$

Donde : X_{ij}: Es el valor de la j-ésima variable efectuado al i-ésimo individuo.

TABLA XVII

LISTADO DE VARIABLES DE ESTUDIO

COD_COLE	X_0	MEDIA	X_{22}
ESPECIALZ	X_1	PROBABIL	X_{23}
EDAD	X_2	NOT_MAT	X_{24}
SEXO	X_3	LECT_COM	X_{25}
ACT_CURR	X_4	PAL_ORAC	X_{26}
NOT_CIEN	X_5	SUJ_NUCL	X_{27}
PROBL_1	X_6	PRED_NUC	X_{28}
PROBL_2	X_7	ORAC_SYC	X_{29}
PROBL_3	X_8	CORR_PAL	X_{30}
CONJUNTO	X_9	HOM_PALA	X_{31}
DES_CONJ	X_{10}	ORT_DIP	X_{32}
OPE_POL1	X_{11}	ORT_TRIP	X_{33}
OPE_POL2	X_{12}	ORT_HIAT	X_{34}
IDENT_GR	X_{13}	IDENT_PA	X_{35}
GRAF_FUN	X_{14}	VOC_SIN	X_{36}
ECUA_REC	X_{15}	VOC_ANT	X_{37}
ECUA_LIN	X_{16}	GEN_LITE	X_{38}
ECUA_CIR	X_{17}	OBRAS_LI	X_{39}
TRIGON_1	X_{18}	GEN_ORAT	X_{40}
TRIGON_2	X_{19}	NOT_LENG	X_{41}
SUPERFIC	X_{20}	NOT_GEN	X_{42}
VOLUMEN	X_{21}		

4.3 Análisis de Correlación Lineal

En la Matriz de Correlación lineal mostrada en el Anexo 5, se aprecian los valores de las correlaciones de cada par de variables, y se procede a realizar el análisis de la relación lineal a partir de los resultados, obteniendo lo siguiente:

- Podemos observar que existe una fuerte relación lineal entre la variable X_{27} (Sujeto y núcleo) y X_{28}(Predicado y núcleo), donde el coeficiente de correlación entre ambas variables se estima de 0.792; indica una relación positiva entre las variables sujeto y predicado con sus correspondientes núcleos, y expresa que si saben sujeto, saben predicado y viceversa.

	X_{27}	X_{28}
X_{27}	1	0.792
X_{28}	0.792	1

- Existe dependencia lineal de la variable diptongo(X_{32}), la cual esta directamente relacionada con las variables con el triptongo(X_{33}), y el coeficiente de correlación entre ambas variables se estima de 0.632.

	X_{32}	X_{33}
X_{32}	1	0.632
X_{33}	0.632	1

- También entre la variable diptongo(X_{32}) y la variable hiato(X_{34}), existe una fuerte relación lineal, donde el coeficiente de correlación es de 0.708, indica que los estudiantes que responden correctamente el diptongo, responden hiato.

	X_{32}	X_{34}
X_{32}	1	0.708
X_{34}	0.708	1

- La variable de triptongo X_{33} y la variable hiato X_{34}, se encuentran fuertemente correlacionadas en forma lineal con un coeficiente de correlación de 0.589.

	X_{33}	X_{34}
X_{33}	1	0.589
X_{34}	0.589	1

- La variable nota de lenguaje (X_{41}) posee una fuerte relación lineal con la variable corrección de palabras X_{30} , donde el coeficiente de correlación que se estima es de 0.535.

	X_{41}	X_{30}
X_{41}	1	0.535
X_{30}	0.535	1

- Existe una fuerte relación lineal entre la variable nota de lenguaje y diptongo , donde el coeficiente de correlación de ambas variables es de 0.661.

	X_{41}	X_{32}
X_{41}	1	0.661
X_{32}	0.661	1

- La variable nota de lenguaje también está fuertemente relacionada con las variables " triptongo", "hiato", "identificar palabras", "sinónimos", y "obras literarias", donde los coeficientes de correlación son superiores a 0.5, como se ilustra en el Anexo 6, es decir que estas variables son las que más relacionadas han estado con la nota de lenguaje para influir en el promedio del mismo.

- De la prueba de matemáticas, las variables con una fuerte relación lineal son, la variable “operaciones algebraicas”(X_{11}) con la variable “operaciones con polinomios 2” (X_{12}), donde el coeficiente de correlación entre ambas variables es de 0.561, es decir que los estudiantes que responden una de las dos preguntas también responde la otra.

	X_{11}	X_{12}
X_{11}	1	0.561
X_{12}	0.561	1

- La variable grafica de funciones (X_{14}) esta relacionada positivamente con la variable ecuación de la recta (X_{15}), el coeficiente de correlación se estima de 0.563.

	X_{14}	X_{15}
X_{14}	1	0.563
X_{15}	0.563	1

- La fuerte relación lineal entre la variable grafica de funciones (X_{14}) y la variable identidades trigonométricas (X_{19}),tiene un coeficiente de correlación estimado de 0.562.

	X_{14}	X_{19}
X_{14}	1	0.562
X_{19}	0.562	1

- Existen una fuerte relación lineal entre las variables ecuación de la recta (X_{15}) y la variable identidades trigonométricas (X_{19}), con un coeficiente de correlación entre ambas variables estimado por 0.609, es decir que la mayor parte de estudiante que saben ecuación de la recta contestan correctamente la segunda pregunta de trigonometría.

	X_{15}	X_{19}
X_{15}	1	0.609
X_{19}	0.609	1

- La variable Superficie(X_{20}) y la variable identidades trigonométricas (X_{19}), tienen una fuerte relación lineal, donde el coeficiente de correlación entre ambas variables se estima de 0.592.

	X_{19}	X_{20}
X_{19}	1	0.592
X_{20}	0.592	1

- Las variables que se encuentran fuertemente relacionadas con la variable nota de matemáticas son: “desigualdad de conjuntos”, “operaciones algebraicas”, “operaciones con polinomios 2”, “identificar gráfica”, “ecuación de la recta”, “ecuación lineal”, “ecuación de la circunferencia”, “trigonometría”, “la variable identidades trigonométricas”, “superficie”, y “volumen”, donde le coeficiente de correlación entre la variable nota de matematicas y las demás son superiores a 0.54, esto significa que depende de las respuesta de cada una de las variables para obtener la nota del estudiante; cabe recalcar que a pesar de ser la nota una combinación lineal de todas las demás variables de la prueba no existe tanta relación con las variables que no están correlacionadas.

- Por último, la variable actividad curricular (X_4) posee una fuerte relación lineal con la variable nota de matemáticas(X_{24}), donde el coeficiente de correlación entre ambas variables se estima de 0.555, indica que depende de la nota de matemáticas que obtenga el estudiante se puede concluir que realiza o no alguna actividad.

Entre las variables que se espero que existiera una fuerte relación lineal, se tienen los siguiente resultados:

Media X_{22} y Probabilidad X_{23},	0.034
Identificar gráfica X_{13} y Graficar funciones X_{14}	0.312
Palabras en la oración X_{26} y Sujeto, núcleo X_{27}	0.266
Sinónimo X_{36} y Antónimo X_{37}	0.159
Géneros literarios X_{38} y Género oratoria X_{40}	-0.035

4.4 Tablas de Contingencias

Generalmente no se puede estudiar a la población en total, esto quiere decir que lo que comúnmente tenemos es un subconjunto de ella, por lo tanto cuando se analizan datos contenidos de una muestra queremos saber si dos variables son independientes o no, para ello utilizamos la técnica estadística denominada tabla de contingencia, que es una técnica bivariada, compuesta por una tabla de r filas y c columnas. Supóngase que la primera variable tiene r niveles del factor 1 o de la variable X_i ,y la segunda tienen c niveles del factor 2 o de la variable X_j .

	Factor 1				
Factor 2	**Nivel 1**	**Nivel 2**		**Nivel c**	
Nivel 1	X_{11}	X_{12}	...	X_{1c}	$X_{1.}$
Nivel 2	X_{21}	X_{22}	. . .	X_{2c}	$X_{2.}$
Nivel r	X_{r1}	X_{r2}	. . .	X_{rc}	$X_{r.}$
	$X_{.1}$	$X_{.2}$		$X_{.c}$	

Para determinar si existe independencia entre las variables o factores formulamos las siguientes hipótesis:

H_o : Los dos factores son independientes el uno de otro

vs.

H_1 :no es verdad H_o

El estadístico de prueba esta definido por:

$$c^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{\left(FO_{ij} - VE_{ij}\right)^2}{VE_{ij}} \sim c^2_{(r-1)(c-1)}$$

Donde:

$$VE_{ij} = \frac{X_{i.}}{n}(X_{.j})$$

Se rechaza H_o, en favor de H_1, con un nivel de confianza de (1-α)100%, es decir, los factores no son independientes el uno del otro, sí: $c^2 > c^2_a(r-1)(c-1)$.

Obteniendo el mínimo nivel de significancia p, se puede aceptar o rechazar H_o.

Cálculos de la Independencia entre las variables

En esta sección lo que haremos es analizar la dependencia e independencia entre las variables analizadas en el capítulo anterior, a través del uso de tablas de contingencia, con la ayuda del software estadístico SPSS 7.0. En las tablas se muestran dos simbologías que son:

FO, es la frecuencia observada, es decir el número de casos en cada opción, y; VE es el valor esperado de la frecuencia absoluta .

- **Variables: Sexo y Edad**

Factor 1:

Antes de realizar los cálculos para la técnica a utilizarse, nos vimos en la necesidad de agrupar las edades para resolver el problema, el cual se ha dividido en :

F1: los estudiantes que se encuentran en edades entre [16, 18)

F2: los estudiantes que se encuentran en edades entre [18, 20)

F3: son los estudiantes que se encuentran en edad igual o superior a 20

Factor 2:

0: Masculino

1: Femenino

Para determinar si existe independencia entre las variables se formula la siguiente hipótesis:

H_o : La variable sexo es independiente de la variable edad

vs.

H_1 : no es verdad H_o

TABLA XVIII

TABLA DE CONTINGENCIA

SEXO VS EDAD

		Sexo		
Edad		0	1	**Total**
F1	FO	266	73	339
	VE	232.585	106.415	339
F2	FO	77	71	148
	VE	101.541	46.459	148
F3	FO	22	23	45
	VE	30.874	14.126	45
Total	FO	365	167	532
	VE	365	167	532

El resultado de la prueba χ^2 es de 42.314 con un valor p de 0,000, en la que podemos concluir que existe suficiente evidencia estadística para rechazar la independencia de las variables , según los resultados existe dependencia entre el sexo y la edad. Se esperaba que no existiera dependencia, es decir que no importa la edad que tenga el estudiante para ser de algún sexo específico, pero lo que se muestra es que existen más estudiantes varones en edades acordes al curso(16-18) y mayor número de mujeres con edades mayores a 20 años.

- Variables Edad y Actividad extra-educativa

Factor 1:

F1: los estudiantes que se encuentran en edades entre [16,18)

F2: los estudiantes que se encuentran en edades entre [18, 20)

F3: son los estudiantes que se encuentran en edad igual o superior a 20

Factor 2:

0: Si realiza actividad

1: No realiza actividad

El contraste de hipótesis planteado es:

H_o : La variable edad es independiente de la actividad no académica que realizan los estudiantes

vs.

H_1 : no es verdad H_o

TABLA XIX

TABLA DE CONTINGENCIA

EDAD VS ACTIVIDAD CURRICULAR

		Actividad extra-educativa		
Edad		0	1	**Total**
F1	FO	242	97	339
	VE	247.241	91.759	339
F2	FO	114	34	148
	VE	107.940	40.060	148
F3	FO	32	13	45
	VE	32.820	12.180	45
Total	FO	388	144	532
	VE	388	144	532

De acuerdo a los resultados mostrados a partir de la tabla XIX, se puede indicar que el valor de la χ^2 es de 1.74298956, con un valor p de 0.41, lo cual nos muestra que existe evidencia estadística para asegurar que la variable edad es independiente de la variable actividad extra-educativa; es decir que sin importar la edad que el estudiante posea puede o no realizar otra actividad.

- Variables Edad y Nota de Matemáticas

Factor 1:

F1: los estudiantes que se encuentran en edades entre [16, 18)

F2: son los estudiantes que se encuentran en edad igual o superior a 18

Factor 2:

a: Nota de matemáticas con promedios menores a 15.789

b: Nota de matemáticas con promedios mayores o iguales a 15.789

El contraste de hipótesis planteado es:

H_o : La variable edad es independiente de la variable nota de matemáticas

vs.

H_1 : no es verdad H_o

TABLA XX

TABLA DE CONTINGENCIA

EDAD VS NOTA DE MATEMATICAS

Edad		**Nota de matemáticas**		**Total**
		a	b	
F1	Fo	140	199	339
	Ve	175.872	163.1278	339
F2	Fo	136	57	193
	Ve	100.1278	92.8722	193
	Fo	276	256	532
Total	Ve	276	256	532

El valor de χ^2 es 41.9125, con un valor p de 0,000 indica que existe evidencia estadística de dependencia entre la variable edad y nota de matemáticas, nos podemos dar cuenta en la tabla XX que el 52% de estudiantes poseen una nota de matemáticas inferior al valor de la mediana, es decir que más de la mitad de los estudiantes tienen un rendimiento bajo en la prueba que realizaron, pero con respecto a la edad los estudiantes con edades entre 16 a 18 son los que obtienen un promedio mayor en dicha nota.

- Variables Edad y Nota de Lenguaje

Factor 1:

F1: los estudiantes que se encuentran en edades entre [16, 18)

F2: son los estudiantes que se encuentran en edad igual o superior a 18

Factor 2:

a: Nota de lenguaje con promedios menores a 66.875

b: Nota de lenguaje con promedios mayores o iguales a 66.875

El contraste de hipótesis planteado es:

H_o : La variable edad es independiente de la variable nota de lenguaje

vs.

H_1 : no es verdad H_o

TABLA XXI

TABLA DE CONTINGENCIA

EDAD VS NOTA DE LENGUAJE

Edad		**Nota de lenguaje**		**Total**
		< 66.875	> 66.875	
F1	FO	154	185	339
	VE	173.323	165.676	339
F2	FO	118	75	193
	VE	98.677	94.324	193
	FO	272	260	532
Total	VE	272	260	532

Con el valor p de 0,000 obtenido del estadístico de prueba, podemos indicar que existe evidencia estadística para resaltar que la edad depende de la nota de lenguaje y viceversa; existe una fuerte dependencia entre dicha nota y la edad, en conclusión lo que se podría indicar que la buena o mala resolución de la prueba del estudiante influye en la edad que este posee; y además que la mayor parte de los estudiantes con edades acordes al curso (16-18) son los que obtienen un mayor promedio.

- Variables Sexo y Actividad Extra-educativa

Factor 1:

0: Masculino

1: Femenino

Factor 2:

0: Realiza alguna actividad

1: No realiza actividad alguna

El contraste de hipótesis planteado es:

H_o : La variable sexo es independiente de la variable actividad extra-educativa

vs.

H_1 : no es verdad H_o

TABLA XXII

TABLA DE CONTINGENCIA

SEXO VS ACTIVIDAD EXTRA-EDUCATIVA

Sexo		Actividad extra-educativa		Total
		0	1	
0	FO	245	120	365
	VE	266.203	98.797	365
1	FO	143	24	167
	VE	121.797	45.203	167
Total	FO	388	144	532
	VE	388	144	532

El valor p es de 0,000 con el valor del estadístico de prueba de 19.876, el cual nos permite concluir que existe evidencia estadística de dependencia entre la variable sexo y actividad extra-educativa, los varones que realizan o no otra actividad representan el 68.6% de los cuales el 32.8% realizan otra actividad; las mujeres que realizan otra actividad representan el 14.37% y el resto solo estudia en el colegio.

- Variables Edad y Lectura Comprensiva

Factor 1:

F1: los estudiantes que se encuentran en edades entre [16, 18)

F2: son los estudiantes que se encuentran en edad igual o superior a 18

Factor 2:

0-2: que respondan de cero a dos literales de preguntas correspondientes a la lectura

3-4: que respondan de tres a cuatro literales de preguntas correspondientes a la lectura

El contraste de hipótesis planteado es:

H_o : La variable edad es independiente de la variable lectura comprensiva

vs.

H_1 : no es verdad H_o

TABLA XXIII

TABLA DE CONTINGENCIA

EDAD(X_2) VS LECTURA COMPRENSIVA(X_{25})

		X_{25}		Total
X_2		0-2	3-4	
F0	FO	123	216	339
	VE	133.1785	205.821	339
F1	FO	86	107	193
	VE	75.8215	117.179	193
	FO	209	323	532
	VE	209	323	532

En el análisis la variable X_2 que representa la edad del estudiante, con la variable Lectura comprensiva, según la prueba χ^2 indica que su valor p es de 0.06, concluyendo que no existe evidencia estadística para rechazar o aceptar la independencia entre las variables, es decir puede o no existir dependencia entre la edad que posee el estudiante para comprender de una manera clara la lectura y responder sus preguntas. La tabla XXIII muestra que, los estudiantes que responden sin dificultad la pregunta, el 66.87% está entre las edades de [16, 18) y el resto de estudiantes superiores a dicha edad no lo hacen.

- Variables Notación científica y Media aritmética

Factor 1:

0: No contestan la pregunta de media aritmética

1-2: Reconocen la media aritmética y responden correctamente

Factor 2:

0: No responden la pregunta de notación científica

1: Plantean el problema pero no lo resuelven

2: Plantean el problema y lo resuelven

El contraste de hipótesis planteado es:

H_o : La variable notación científica es independiente de la variable media

vs.

H_1 : no es verdad H_o

TABLA XXIV

TABLA DE CONTINGENCIA

NOTACION CIENTIFICA(X_5) VS MEDIA ARITMETICA (X_{22})

X_5		**X_{22}**		**Total**
		0	1-2	
0	FO	183	131	314
	VE	169.985	144.015	314
1	FO	46	51	97
	VE	52.511	44.489	97
2	FO	59	62	121
	VE	65.504	55.496	121
Total	FO	288	244	532
	VE	288	244	532

Como en el caso anterior el valor p de la prueba es de 0.069, lo cual indica que no existe evidencia estadística para rechazar o aceptar la independencia entre las variables notación científica y media aritmética, es decir que los estudiantes que responden ambas preguntas representan el 25.4%, mientras que la mayor parte no responden las dos, o responde solo una de las dos preguntas como se muestra en la tabla XXIV.

- Variables Probabilidad y Media aritmética

Factor 1:

0: No contestan la pregunta de probabilidad

1: Resuelven correctamente el problema de probabilidad

Factor 2:

0: No responden la pregunta de media aritmética

1: Reconocen la media pero no la resuelven

2: Plantean el problema y lo resuelven

El contraste de hipótesis planteado es:

H_o : La variable probabilidad es independiente de la variable media aritmética

vs.

H_1 : no es verdad H_o

TABLA XXV

MEDIA ARITMETICA(X_{22}) VS PROBABILIDAD(X_{23})

X_{23}		X_{22}			Total
		0	1	2	
0	FO	269	37	185	491
	VE	265.80	38.76	186.43	491
1	FO	19.000	5.000	17.000	41
	VE	22.20	3.24	15.57	41
	FO	288	42	202	532
Total	VE	288	42	202	532

En este análisis quisimos verificar si los estudiantes que resuelven el problema de la media también podrían resolver el de probabilidad, para lo cual podemos concluir que el valor p es de 0.431, lo cual indica que existe evidencia estadística para aceptar la independencia de las variables, es decir que el estudiante que sabe media no necesariamente sabe probabilidad y resuelve ambas preguntas.

- Variables Sistemas de ecuaciones lineales y determinar la ecuación de la recta

Factor 1:

0: No resuelven el problema de sistemas de ecuaciones lineales

1: Responden correctamente la pregunta

Factor 2:

0: No resuelven el problema determinación de la ecuación de la recta

1: Plantean el problema pero no lo resuelven

2: Plantean el problema y lo resuelven

El contraste de hipótesis planteado es:

H_o : La variable determinar la ecuación de la recta es independiente de la variable sistemas de ecuaciones lineales

vs.

H_1 : no es verdad H_o

TABLA XXVI

SISTEMAS DE ECUACIONES LINEALES(X_{16}) VS DETERMINAR LA ECUACION DE LA RECTA (X_{16})

X_{16}		X_{15}			Total
		0	1	2	
0	FO	390	10	12	412
	VE	357.015	14.714	40	412
1	FO	71.00	9.00	40	120
	VE	103.985	4.285714	11.72932	120
Total	FO	461	19	52	532
	VE	461	19	52	532

El valor p de la prueba es de 0,000 el cual indica que existe suficiente evidencia estadística para rechazar la independencia entre las variables, es decir que la variable determinar la ecuación de la recta depende de la variable sistemas de ecuaciones lineales, se puede indicar que si los estudiantes no resuelven el problema de determinar la ecuación de la recta podríamos concluir que tampoco podrán realizar el problema de sistemas de ecuaciones lineales, y viceversa. Los estudiantes que no responden ninguna de las dos preguntas tienen un porcentaje 73.3% y el resto si contestan las dos preguntas, pero los que contestan una pregunta de las dos el porcentaje es relativamente bajo en ambos casos, que si contestaran ambas preguntas a la vez.

- Variables Sexo y Lectura Comprensiva

Factor 1:

0: Masculino

1: Femenino

Factor 2:

0: No responden la pregunta

1: Responde correctamente al menos un literal

2: Responde correctamente dos literales

3: Responde correctamente tres literales

4: Responde correctamente todos los literales

El contraste de hipótesis planteado es:

H_o : La variable sexo es independiente de la variable lectura comprensiva

vs.

H_1 : no es verdad H_o

TABLA XXVII

TABLA DE CONTINGENCIA

SEXO(X_3) VS LECTURA COMPRENSIVA(X_{25})

X_3		X_{25}					Total
		0	1	2	3	4	
0	Fo	37	16	86	112	114	365
	Ve	37.049	15.780	90.564	120.066	101.541	365
1	Fo	17	7	46	63	34	167
	Ve	16.951	7.220	41.436	54.934	46.459	167
Total	Fo	54	23	132	175	148	532
	Ve	54	23	132	175	148	532

De acuerdo a la prueba que se aplicó para realizar el análisis el valor p obtenido fue de 0.119 el cual indica que existe evidencia estadística para aceptar la independencia de las variables. Muestra que el sexo del estudiante no influye en lo que tiene que si responde o no a la pregunta de lectura comprensiva.

- Variables Actividad extra-educativa y Nota de Matemáticas

Factor 1:

0: No realiza actividad alguna

1: Si realiza alguna actividad

Factor 2:

a: Nota de matemáticas con promedios menores a 15.789

b: Nota de matemáticas con promedios mayores o iguales a 15.789

El contraste de hipótesis planteado es:

H_o : La variable actividad extra-educativa es independiente de la variable nota de matemáticas

vs.

H_1 : no es verdad H_o

TABLA XXVIII

TABLA DE CONTINGENCIA

ACTIVIDAD EXTRA-EDUCATIVA VS NOTA MATEMATICAS

Act ext-educ.		**Nota de matemáticas**		**Total**
		a	b	
0	FO	217	171	388
	VE	201.293	186.706	388
1	FO	59	85	144
	VE	74.707	69.294	144
Total	FO	276	256	532
	VE	276	256	532

En la prueba realizada se obtuvo el valor p de 0,000 indica que existe evidencia estadística para aceptar la dependencia entre la variable Actividad extra-educativa y la nota de matemáticas; podemos observar en la tabla XXVIII que el 78.6% de los estudiantes que no realizan otra actividad y tienen un promedio menor o igual a 15.8 y el resto si realiza alguna actividad ; así mismo quienes no realizan otra actividad y tiene el promedio superior a 15.8 son un total de 66.79% y el resto es decir quienes realizan otra actividad tienen bajo rendimiento en dicha materia, por ese motivo es que existe dependencia entre quienes realizan otra actividad y no lo hacen; con respecto a la nota de matemáticas.

- Variables Actividad Extra-educativa y Nota de Lenguaje

Factor 1:

0: No realiza actividad alguna

1: Si realiza actividad

Factor 2:

a: Nota de lenguaje con promedios menores a 66.875

b: Nota de lenguaje con promedios mayores o iguales a 66.875

El contraste de hipótesis planteado es:

H_o : La variable actividad extra-educativa es independiente de la variable nota de lenguaje

vs.

H_1 : no es verdad H_o

TABLA XXIX

TABLA DE CONTINGENCIA

ACTIVIDAD EXTRA-EDUCATIVA VS NOTA LENGUAJE

Act extra-educ.		**Nota de lenguaje**		**Total**
		a	b	
0	FO	214	174	388
	VE	199.105	73.895	388
1	FO	59	85	144
	VE	73.895	70.105	144
Total	FO	273	259	532
	VE	273	259	532

El valor p para esta prueba es de 0.003 muestra que existe suficiente evidencia estadística para rechazar la independencia entre las variables de Actividad extra-educativa y la nota de lenguaje, es decir que si influye el hecho de que un estudiante realice actividades curriculares con el rendimiento de la materia; a pesar, que si podemos observar en la tabla XXIX, de los estudiantes que no realizan actividad alguna, el 55.15% tienen un promedio inferior a la mediana, y el resto tienen promedio mayor ; en cambio los estudiantes que realizan otra actividad, el 59% de ellos poseen un promedio superior a la mediana.

- Variables Operaciones algebraicas y Operaciones con polinomios 2

Factor 1:

0: No responde la pregunta de operaciones algebraicas

1: Realiza correctamente algunas operaciones algebraicas

2: Realiza correctamente todas las operaciones algebraicas

Factor 2:

0: No responde la pregunta

1: Realiza correctamente algunas operaciones

2: Realiza correctamente todas las operaciones

El contraste de hipótesis planteado es:

H_o : La variable operaciones algebraicas es independiente de la variable Operaciones con polinomios 2

vs.

H_1 : no es verdad H_o

TABLA XXX

TABLA DE CONTINGENCIA

OPERACIONES ALGEBRAICAS(X_{11}) VS OPERACIÓN POLINOMIOS2(X_{12})

X_{11}		X_{12}			Total
		0	1	2	
0	FO	250	22	12	284
	VE	169.759	50.180	64.060	284
1	FO	14	18	28	60
	VE	35.865	10.602	13.534	60
2	FO	54	54	80	188
	VE	112.376	33.218	42.406	188
Total	FO	318	94	120	532
	VE	318	94	120	532

El valor p de la prueba según los datos de la tabla XXX, es de 0,000, en el que se puede concluir que existe suficiente evidencia estadística para aceptar la dependencia entre la variable operaciones algebraicas y ope_pol 2, las cuales son operaciones con polinomios e indica que si el estudiante está en la capacidad de resolver un problema tiene facilidad de resolver el otro también, aunque al 29.5 % de los estudiantes que responden correctamente la pregunta operaciones algebraicas se les hace más fácil resolverlo en comparación con los que responden solo la pregunta Ope_pol2. Así mismo el total de estudiantes que responden correctamente ambas preguntas es 14.4%.

- Variables Sujeto, núcleo y Predicado, núcleo

Factor 1:

0: No responde la pregunta de sujeto, núcleo

1-2: Identifica correctamente el sujeto y/o su núcleo

Factor 2:

0: No responde la pregunta predicado, núcleo

1-2: Identifica correctamente el sujeto y/o su núcleo

El contraste de hipótesis planteado es:

H_o : La variable Sujeto, núcleo es independiente de la variable Predicado, núcleo

vs.

H_1 : no es verdad H_o

TABLA XXXI

TABLA DE CONTINGENCIA

SUJETO Y NUCLEO (X_{27}) VS PREDICADO Y NUCLEO (X_{28})

X_{27}		X_{28}		Total
		0	1-2	
0	FO	64	6	70
	VE	9.342	60.657	70
1-2	FO	7	455	462
	VE	61.658	400.343	462
Total	FO	71	461	532
	VE	71	461	532

Se desea verificar la dependencia o independencia entre la variable Sujeto, núcleo y Predicado,núcleo; y de acuerdo a los resultados se obtiene que el valor p es de 0,000 y se puede concluir que existe suficiente evidencia estadística para verificar la dependencia entre las dos variables, es decir que de acuerdo a los datos ilustrados en la tabla XXXI, se muestra que existe un porcentaje mínimo de los estudiantes que identifican solo el sujeto o solo el predicado a diferencia de los que identifican correctamente las dos partes, con un porcentaje del 85%. En conclusión los estudiantes que identifican correctamente el sujeto pueden identificar también el predicado.

- Variable Especialización y Nota de Matemáticas

Factor 1:

1: Fisico matemático

2:Quimico biológico

3-6-7. Ciencias Sociales, Secretariado y Técnico

4: Contabilidad

5: Informática

Factor 2:

a: Nota de matemáticas con promedios menores a 15.789

b: Nota de matemáticas con promedios mayores o iguales a 15.789

El contraste de hipótesis planteado es:

H_o : La variable especialización es independiente de la variable nota de matemáticas

vs.

H_1 : no es verdad H_o

TABLA XXXII

TABLA DE CONTINGENCIA

ESPECIALZACION (X_1) VS NOTA MATEMATICAS (X_{24})

X_1		X_{24}		Total
		a	b	
1	FO	8	93	101
	VE	52.398	48.603	101
2	FO	14	41	55
	VE	28.534	26.466	55
3-6-7	FO	59	7	66
	VE	34.241	31.759	66
4	FO	99	44	143
	VE	74.188	68.812	143
5	FO	96	71	167
	VE	86.639	80.36	167
Total	FO	276	256	532
	VE	276	256	532

En la tabla XXXII se va a calcular si existe independencia entre la especialización del alumno y la nota que este adquiere en matemáticas, la variable especialización se agrupa en cinco factores; la codificación de la variable se detalla en el capítulo 2, para lo cual en este análisis se vio en la necesidad de agrupar algunas especializaciones que corresponden a la codificación 3,6 y 7 ilustrados en la tabla con sus correspondientes valores. Donde le valor del estadístico de prueba es de 150.1155 y el valor p es 0,000 podemos concluir que existe suficiente evidencia estadística para aceptar la dependencia entre la variable especialización y la nota de matemáticas, observando que los estudiantes que mayor promedio poseen en la prueba son los de la especialización 1 que representa a físico matemático, como era de esperarse, con 92.07% de ellos que tienen un valor superior a la mediana.

- Variable Especialización y Nota de Lenguaje

Factor 1:

1: Físico matemático

2:Quimico biológico

3: Ciencias sociales,

4: Contabilidad

5: Informática

6-7: Secretariado y Técnico

Factor 2:

a: Nota de lenguaje con promedios menores a 66.875

b: Nota de lenguaje con promedios iguales o mayores a 66.875

El contraste de hipótesis planteado es:

H_o : La variable especialización es independiente de la variable nota de lenguaje

vs.

H_1 : no es verdad H_o

TABLA XXXIII

TABLA DE CONTINGENCIA

ESPECIALZACION (X_1) VS NOTA LENGUAJE (X_{41})

X_1		X_{41}		Total
		a	b	
1	FO	36	65	101
	VE	49.93	51.07	101
2	FO	19	36	55
	VE	27.19	27.81	55
3	FO	17	24	41
	VE	20.269	20.731	41
4	FO	76	67	143
	VE	70.694	72.306	143
5	FO	99	68	167
	VE	82.558	84.44	167
6-7	FO	16	9	25
	VE	12.359	12.64	25
Total	FO	263	269	532
	VE	263	269	532

La tabla de contingencia que se va a analizar es la verificación del supuesto que la nota de lenguaje es independiente de la especialización, donde se observa que existen 6 factores, y se a considerado para este análisis agrupar los que tienen codificación 6 y 7, para realizar los cálculos; y el valor p obtenido en la tabla es de 0,000 el cual muestra que existe suficiente evidencia estadística para aceptar que la variable especialización depende de la variable nota de lenguaje, como se ilustra en la tabla XXXIII, los estudiantes que poseen promedio mayor a la mediana en la nota de lenguaje son los de la especialización 5 que es informática.

En la tabla XXXIV se muestran las variables con su correspondiente valor p; se acepta la hipótesis nula cuando el valor p es igual o mayor a 0.1, es decir existe independencia entre las variables, y no existe evidencia estadística para aceptar o rechazar la hipótesis nula si el valor p está entre 0.05 y 0.1.

TABLA XXXIV

PARES DE VARIABLES PARA EL ANALISIS DE TABLAS DE CONTINGENCIA

Factor 1	Factor 2	Valor p
Planteamiento y resolución de problemas	Sistemas de ecuaciones lineales	0.03
Regla de tres compuesta	Sistemas de ecuaciones lineales	0.105
Sucesión	Operaciones con polinomios	0.218
Conjunto	Probabilidad	0.867
Operaciones algebraicas	Lectura comprensiva	0.221
Identificar gráfico	Elementos de la oración	0.07
Género oratoria	Sistemas de ecuaciones lineales	0.405
Lectura comprensiva	Determinar la ecuación de la recta	0.145
Género literario	Lectura comprensiva	0.378
Probabilidad	Gráfica de funciones	0.076
Palabras homófonas	Identificar gráfico	0.33
Género de oratoria	Palabras homófonas	0.07
Género de oratoria	Predicado y núcleo	0.533

4.2.4 Componentes Principales

El análisis de componentes principales, explica las varianzas y covarianzas de un conjunto de datos a través de unas pocas combinaciones lineales de los mismos. En forma algebraica se pude indicar que las componentes principales son una combinación de p variables observadas o vectores aleatorios $X_1, X_2, \dots, X_p$.

Las componentes principales dependen de la matriz de varianzas y covarianza Σ estimada por **S,** o de la matriz de correlación ρ estimado por **R** ; de $X_1, X_2, \dots, X_p$.

Sea $\mathbf{X}=(x_1, x_2,\dots,x_p)$ un vector observable p-variado con media μ estimado por $\overline{X}$ y matriz de varianzas y covarianzas Σ estimada por **S,** cuyos valores propios son: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ y se definen k<p variables no observadas $Y_1, Y_2, \dots Y_k$ como una combinación lineal de $X_1, X_2, \dots, X_p$, esto es:

$$\mathbf{Y_1} = \mathbf{a'_1 X} = a'_{11} X_1 + a'_{21} X_2 + \ .... + a'_{p1} X_p$$

$$\mathbf{Y_2} = \mathbf{a'_2 X} = a'_{12} X_1 + a'_{22} X_2 + ... + a'_{p2} X_p$$

..

$$\mathbf{Y_p} = \mathbf{a'_p X} = a'_{1p} X_1 + a'_{2p} X_2 + ... + a'_{pp} X_p$$

Se puede probar que:

$$Var(Y_i) = \mathbf{a_i}^T S \, \mathbf{a_i} = \lambda_i$$

$$Cov(Y_i \, , \, Y_j) = \mathbf{a_i}^T S \, \mathbf{a_j} \ = 0 \ ;$$

$$i \neq j; \quad para \ \ i=1,2,...p; \ j=1,2,3,...p$$

Las componentes principales del vector **X** son las combinaciones lineales Y_1, Y_2,...,Y_p que se encuentran ordenadas de tal manera que entre menor sea el subíndice de la componente, la varianza es lo más grande posible.

$$Var(Y_i) > Var(Y_j) \text{ para todo } i<k$$

La matriz de varianza - covarianza muestral es **S** asociada con el vector p-variado **X** $=(X_1, X_2, \ldots, X_p)$; y sean los valores propios y vectores propios asociados a **S**: $(\lambda_1,\mathbf{a_1})$, $(\lambda_2,\mathbf{a_2})$, $(\lambda_3,\mathbf{a_3})$, ..., $(\lambda_p,\mathbf{a_p})$

Donde: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$

La proporción del total de la variación explicada por la k-ésima componente principal es:

$$\frac{l_k}{l_1 + l_2 + l_3 + \ldots + l_p}, \text{ k = 1, 2, ..., p}$$

El número de componentes principales que se escojan para el estudio depende del porcentaje de varianza que se desee explicar.

Cálculos de Componentes Principales

Realizando los cálculos, con la ayuda del software estadístico SPSS 7.0, se presenta en la tabla XXXV los resultados obtenidos de la matriz de datos originales. De 42 variables que se utilizaron para el estudio, la técnica logró reducir 2 componentes principales, para los valores originales que explican el 96.23% de la varianza total; en la misma tabla se muestran los valores propios de cada componente, y el porcentaje acumulado de la varianza, en el que podemos comprobar su explicación con las dos primeras componentes.

TABLA XXXV

ANALISIS DE COMPONENTES PRINCIPALES

DATOS ORIGINALES

Comp.	Valor Propio	% de Varianza	% Acum.
1	412.21	58.62	58.62
2	264.43	37.61	96.23
3	6.77	0.96	97.19
4	3.34	0.47	97.67
5	1.76	0.25	97.92
6	1.50	0.21	98.13
7	1.29	0.18	98.31
8	1.10	0.16	98.47
9	0.93	0.13	98.60
10	0.80	0.11	98.72
11	0.76	0.11	98.83
12	0.72	0.10	98.93
13	0.66	0.09	99.02
14	0.60	0.09	99.11
15	0.58	0.08	99.19
16	0.50	0.07	99.26
17	0.48	0.07	99.33
18	0.38	0.05	99.38
19	0.37	0.05	99.43
20	0.34	0.05	99.48
21	0.31	0.04	99.53

Comp.	Valor Propio	% de Varianza	% Acum.
22	0.30	0.04	99.57
23	0.30	0.04	99.61
24	0.26	0.04	99.65
25	0.25	0.04	99.68
26	0.23	0.03	99.72
27	0.21	0.03	99.75
28	0.20	0.03	99.77
29	0.19	0.03	99.80
30	0.18	0.03	99.83
31	0.16	0.02	99.85
32	0.16	0.02	99.87
33	0.15	0.02	99.90
34	0.13	0.02	99.91
35	0.12	0.02	99.93
36	0.12	0.02	99.95
37	0.11	0.02	99.97
38	0.09	0.01	99.98
39	0.08	0.01	99.99
40	0.06	0.01	100.00
41	0.01	0.00	100.00
42	0.00	0.00	100.00

En la tabla XXXVI, se ilustran las dos componentes principales con su correspondiente peso para cada variable.

TABLA XXXVI

MATRIZ DE LAS DOS COMPONENTES PRINCIPALES

	Y_1	Y_2		Y_1	Y_2
X_0	-0.824	-0.150	X_{22}	0.280	-0.195
X_1	-0.893	0.403	X_{23}	0.407	-0.232
X_2	-0.673	-0.059	X_{24}	0.061	-0.043
X_3	-0.168	0.121	X_{25}	15.94	-9.888
X_4	0.059	-0.030	X_{26}	0.413	0.297
X_5	0.354	-0.219	X_{27}	0.166	0.158
X_6	0.103	-0.083	X_{28}	0.256	0.205
X_7	0.102	-0.076	X_{29}	0.299	0.215
X_8	0.072	-0.085	X_{30}	0.400	0.491
X_9	0.169	-0.111	X_{31}	0.484	0.295
X_{10}	0.322	-0.172	X_{32}	0.334	0.348
X_{11}	0.490	-0.301	X_{33}	0.383	0.388
X_{12}	0.464	-0.299	X_{34}	0.265	0.377
X_{13}	0.274	-0.081	X_{35}	0.476	0.386
X_{14}	0.597	-0.406	X_{36}	0.253	0.319
X_{15}	0.316	-0.243	X_{37}	0.304	0.267
X_{16}	0.208	-0.142	X_{38}	0.197	0.211
X_{17}	0.253	-0.165	X_{39}	0.213	0.238
X_{18}	0.395	-0.278	X_{40}	0.158	0.259
X_{19}	0.663	-0.457	X_{41}	0.022	0.101
X_{20}	0.470	-0.295	X_{42}	12.335	12.806

De la tabla XXXVI se obtienen las combinaciones lineales de cada componente principal, en este caso de las dos componentes, multiplicando el coeficiente o vector con las 41 variables originales; los coeficientes representan el peso que posee cada variable original como se presenta a continuación:

$$
\begin{aligned}
Y_1 = {} & -0.824\,X_1 + -0.893\,X_2 + -0.673\,X_3 + -0.168\,X_4 + 0.059\,X_5 + 0.354\,X_6 \\
& + 0.102\,X_7 + 0.072\,X_8 + 0.169\,X_9 + 0.322\,X_{10} + 0.490\,X_{11} + 0.464\,X_{12} \\
& + 0.274\,X_{13} + 0.597\,X_{14} + 0.316\,X_{15} + 0.208\,X_{16} + 0.253\,X_{17} + 0.395\,X_{18} \\
& + 0.663\,X_{19} + 0.470\,X_{20} + 0.280\,X_{21} + 0.407\,X_{22} + 0.061\,X_{23} + 15.94\,X_{24} \\
& + 0.413\,X_{25} + 0.166\,X_{26} + 0.256\,X_{27} + 0.299\,X_{28} + 0.400\,X_{29} + 0.484\,X_{30} \\
& + 0.334\,X_{30} + 0.334\,X_{31} + 0.383\,X_{32} + 0.265\,X_{33} + 0.476\,X_{34} + 0.253\,X_{35} \\
& + 0.304\,X_{36} + 0.197\,X_{37} + 0.213\,X_{38} + 0.158\,X_{40} + 0.022\,X_{41} + 12.335\,X_{42}
\end{aligned}
$$

$$
\begin{aligned}
Y_2 = {} & 0.403\,X_1 + -0.059\,X_2 + 0.121\,X_3 + -0.030\,X_4 + -0.219\,X_5 + -0.083\,X_6 \\
& + -0.076\,X_7 + -0.085\,X_8 + -0.111\,X_9 + -0.172\,X_{10} + -0.301\,X_{11} + -0.299\,X_{12} \\
& + -0.081\,X_{13} + -0.406\,X_{14} + -0.243\,X_{15} + -0.142\,X_{16} + -0.165\,X_{17} + -0.278\,X_{18} \\
& + -0.278\,X_{19} + -0.295\,X_{20} + -0.195\,X_{21} + -0.232\,X_{22} + -0.043\,X_{23} + -9.888\,X_{24} \\
& + 0.297\,X_{25} + 0.158\,X_{26} + 0.205\,X_{27} + 0.215\,X_{28} + 0.491\,X_{29} + 0.295\,X_{30} \\
& + 0.348\,X_{30} + 0.348\,X_{31} + 0.388\,X_{32} + 0.377\,X_{33} + 0.386\,X_{34} + 0.319\,X_{35} \\
& + 0.267\,X_{36} + 0.211\,X_{37} + 0.238\,X_{38} + 0.259\,X_{40} + 0.101\,X_{41} + 12.886\,X_{42}
\end{aligned}
$$

Se puede apreciar que la primera componente contiene mayor peso de variables y las variables que más aportan son la nota de matemáticas y la de lenguaje.

Trabajando con la matriz de datos estandarizadas se obtienen los resultados mostrados en la tabla XXXVII, ilustrando que de las 42 variables originales se reduce a 11 variables con una varianza total de explicación de 61.16%, el porcentaje de explicación de cada componente se muestran en la misma tabla.

Como en el estudio, se trabaja con variables en diversas escalas, pueden surgir inconvenientes al trabajar con la matriz de varianzas y covarianzas, por el motivo de que en el momento del análisis, las variables de mayores escalas absorben los pesos más significativos; y para evitar se utiliza la matriz de datos estandarizada la cual, lleva todas las variables a una misma escala estandarizando cada una de ellas de la siguiente manera: a la variable se le resta su media aritmética $\overline{x}_i$ y se la divide para la desviación estándar **S_i**, es decir:

$$Z_i = \left(\frac{x_i - \overline{x}_i}{s_i} \right); \qquad i = 1,2,\ldots,p$$

Donde Z_1, Z_2 , . . . , Z_p son los valores estandarizados de las variables X_1, $X_2,\ldots,X_p$

La tabla XXXVII muestra los valores propios y el porcentaje de explicación de las componentes principales, donde se ilustra que con 11 componentes principales se explica el 60.86% .

TABLA XXXVII

ANALISIS DE COMPONENTES PRINCIPALES

MATRIZ ESTANDARIZADA

Comp.	Valor propio	% de Varianza	% Acum.	Comp.	Valor propio	% de Varianza	% Acum.
1	8.438	20.091	20.091	22	0.673	1.602	82.240
2	4.805	11.439	31.531	23	0.637	1.518	83.758
3	1.835	4.370	35.901	24	0.586	1.396	85.154
4	1.731	4.122	40.023	25	0.578	1.376	86.529
5	1.476	3.514	43.537	26	0.559	1.332	87.861
6	1.424	3.391	46.928	27	0.513	1.221	89.082
7	1.372	3.268	50.195	28	0.492	1.172	90.254
8	1.168	2.781	52.976	29	0.483	1.149	91.404
9	1.152	2.744	55.720	30	0.454	1.081	92.485
10	1.109	2.640	58.360	31	0.421	1.003	93.488
11	1.051	2.503	60.863	32	0.380	0.906	94.394
12	0.957	2.279	63.142	33	0.378	0.900	95.294
13	0.934	2.225	65.366	34	0.332	0.791	96.085
14	0.889	2.118	67.484	35	0.329	0.784	96.870
15	0.865	2.059	69.543	36	0.318	0.758	97.627
16	0.851	2.025	71.569	37	0.306	0.729	98.356
17	0.827	1.969	73.538	38	0.271	0.646	99.002
18	0.778	1.853	75.391	39	0.233	0.554	99.557
19	0.763	1.817	77.208	40	0.182	0.433	99.990
20	0.727	1.731	78.938	41	0.003	0.008	99.998
21	0.714	1.700	80.638	42	0.001	0.002	100.000

Una vez obtenido el porcentaje de explicación de las componentes, el Anexo 6, muestra la matriz con el peso de cada componente. Se ilustra que la primera componente es la que tiene la mayor carga de todas. Para lo cual, nos vemos en la necesidad de efectuar una rotación de las componentes trabajando con la matriz de datos estandarizada, utilizando el método de VARIMAX y lograr distribuir la varianza a lo largo de las componentes obteniendo resultados simplificados y precisos como los que se muestran en la tabla XXXVIII.

TABLA XXXVIII

VALORES DEL ANALISIS DE COMPONENTES PRINCIPALES

APLICANDO VARIMAX

Comp.	Valores propios	% de Varianza	% Acumulado
1	6.85408519	16.3192505	16.3192505
2	3.42290616	8.14977657	24.469027
3	2.65073303	6.31126912	30.7802962
4	2.24564419	5.34677188	36.127068
5	1.96776688	4.68515924	40.8122273
6	1.78439089	4.24854974	45.060777
7	1.55252048	3.69647732	48.7572543
8	1.32936765	3.16516107	51.9224154
9	1.28591051	3.0616917	54.9841071
10	1.24976757	2.97563707	57.9597442
11	1.21939724	2.90332676	60.8630709

Aplicando VARIMAX se han obtenido nuevos valores en el porcentaje de varianza explicada por la 11 componentes y se mantiene el total de explicación en 61.16%, como se puede apreciar en la tabla XXXVIII.

La rotación ayuda a distribuir de manera más equitativa el porcentaje de explicación de cada componente principal, pero comparando los datos no se ha tenido una buena rotación y en el Anexo 7, se muestra los datos del peso de las componentes. A continuación se detallan las combinaciones lineales de las 11 componentes principales y luego se procede a realizar el análisis de cada una de ellas, verificando el mayor peso que posean los coeficientes (en valor absoluto) de las variable en cada componente.

$$
\begin{aligned}
Y_1 = \; & -0.12\,X_0 + -0.6699\,X_1 + -0.1387\,X_2 + -0.37968\,X_3 + 0.0718\,X_4 + 0.364\,X_5 \\
& + 0.07\,X_6 + 0.1558\,X_7 + 0.11678\,X_8 + 0.3865\,X_9 + 0.5916\,X_{10} + 0.5503\,X_{11} \\
& + 0.609\,X_{12} + 0.435\,X_{13} + \mathbf{0.76723}\,X_{14} + \mathbf{0.75997}\,X_{15} + 0.6281\,X_{16} + 0.557\,X_{17} \\
& + 0.652\,X_{18} + \mathbf{0.7763}\,X_{19} + 0.65996\,X_{20} + 0.38425\,X_{21} + 0.4781\,X_{22} + 0.1434\,X_{23} \\
& + \mathbf{0.922}\,X_{24} + 0.0034\,X_{25} + 0.00344\,X_{26} + 0.09477\,X_{27} + 0.1381\,X_{28} + 0.0267\,X_{29} \\
& + 0.112\,X_{30} + 0.0419\,X_{31} + 0.07847\,X_{32} + -0.02416\,X_{33} + 0.1333\,X_{34} + 0.0266\,X_{35} \\
& + 0.148\,X_{36} + 0.0425\,X_{37} + 0.02025\,X_{38} + -0.03456\,X_{40} + -0.0413\,X_{41} + 0.0902\,X_{42}
\end{aligned}
$$

$$
\begin{aligned}
Y_2 = \; & -0.15\,X_0 + -0.0635\,X_1 + -0.0799\,X_2 + 0.00914\,X_3 + -0.1109\,X_4 + 0.1177\,X_5 \\
& + 0.028\,X_6 + 0.0254\,X_7 + -0.1209\,X_8 + 0.10362\,X_9 + 0.0586\,X_{10} + -0.0764\,X_{11} \\
& + -0.05\,X_{12} + 0.1469\,X_{13} + 0.04745\,X_{14} + 0.02165\,X_{15} + -0.0545\,X_{16} + 0.0463\,X_{17} \\
& + 0.062\,X_{18} + 0.0381\,X_{19} + 0.07747\,X_{20} + 0.07029\,X_{21} + 0.0554\,X_{22} + -0.0197\,X_{23} \\
& + 0.066\,X_{24} + 0.1838\,X_{25} + 0.22644\,X_{26} + 0.09056\,X_{27} + 0.0849\,X_{28} + 0.1849\,X_{29} \\
& + 0.28\,X_{30} + 0.2245\,X_{31} + \mathbf{0.85504}\,X_{32} + \mathbf{0.83312}\,X_{33} + \mathbf{0.8338}\,X_{34} + 0.5825\,X_{35} \\
& + 0.311\,X_{36} + 0.0191\,X_{37} + 0.01739\,X_{38} + 0.17418\,X_{40} + 0.0447\,X_{41} + 0.6472\,X_{42}
\end{aligned}
$$

$Y_3 = -0.16\ X_0 + -0.2307\ X_1 + -0.0535\ X_2 + -0.05238\ X_3 + 0.0533\ X_4 + -0.0483\ X_5$
$+ 0.057\ X_6 + -0.0118\ X_7 + -0.014\ X_8 + -0.07312\ X_9 + 0.0758\ X_{10} + -0.0253\ X_{11}$
$+ -0.1\ X_{12} + 0.0997\ X_{13} + -0.0375\ X_{14} + 0.03492\ X_{15} + 0.0917\ X_{16} + 0.0799\ X_{17}$
$+ -0.06\ X_{18} + -0.0109\ X_{19} + 0.12033\ X_{20} + -0.13435\ X_{21} + -0.094\ X_{22} + 0.0129\ X_{23}$
$+ 0.000\ X_{24} + 0.0804\ X_{25} + 0.16036\ X_{26} + 0.10655\ X_{27} + 0.1289\ X_{28} + 0.1603\ X_{29}$
$+ 0.307\ X_{30} + 0.3252\ X_{31} + 0.09792\ X_{32} + 0.02509\ X_{33} + 0.1044\ X_{34} + 0.4684\ X_{35}$
$+ \mathbf{0.629}\ X_{36} + \mathbf{0.7022}\ X_{37} + 0.57212\ X_{38} + \mathbf{0.66708}\ X_{40} + -0.0343\ X_{41} + 0.5261\ X_{42}$

$Y_4 = 0.009\ X_0 + 0.0226\ X_1 + -0.1276\ X_2 + 0.1673\ X_3 + 0.0697\ X_4 + 0.0914\ X_5$
$+ 0.024\ X_6 + 0.0034\ X_7 + 0.00104\ X_8 + 0.02789\ X_9 + 0.0473\ X_{10} + 0.0797\ X_{11}$
$+ 0.091\ X_{12} + -0.0438\ X_{13} + 0.06458\ X_{14} + 0.03074\ X_{15} + 0.0853\ X_{16} + 0.0571\ X_{17}$
$+ 0.063\ X_{18} + 0.0375\ X_{19} + -0.0125\ X_{20} + -0.05479\ X_{21} + -0.0717\ X_{22} + 0.041\ X_{23}$
$+ 0.056\ X_{24} + 0.2212\ X_{25} + 0.37316\ X_{26} + \mathbf{0.8847}\ X_{27} + \mathbf{0.8843}\ X_{28} + 0.3642\ X_{29}$
$+ 0.141\ X_{30} + 0.0635\ X_{31} + 0.05661\ X_{32} + 0.08428\ X_{33} + 0.0815\ X_{34} + -0.0673\ X_{35}$
$+ 0.06\ X_{36} + 0.0583\ X_{37} + 0.15975\ X_{38} + 0.03418\ X_{40} + -0.0447\ X_{41} + 0.4084\ X_{42}$

$Y_5 = 0.09\ X_0 + -0.0921\ X_1 + -0.0605\ X_2 + \mathbf{-0.51881}\ X_3 + \mathbf{0.5305}\ X_4 + 0.1975\ X_5$
$+ -0.07\ X_6 + 0.1012\ X_7 + 0.06609\ X_8 + 0.0325\ X_9 + 0.0041\ X_{10} + 0.3375\ X_{11}$
$+ 0.357\ X_{12} + 0.3225\ X_{13} + -0.0231\ X_{14} + -0.10685\ X_{15} + 0.0036\ X_{16} + -0.0975\ X_{17}$
$+ -0.02\ X_{18} + 0.1675\ X_{19} + 0.0339\ X_{20} + 0.33004\ X_{21} + 0.0538\ X_{22} + 0.1237\ X_{23}$
$+ 0.207\ X_{24} + 0.4558\ X_{25} + 0.0887\ X_{26} + 0.02693\ X_{27} + 0.0589\ X_{28} + 0.1353\ X_{29}$
$+ 0.416\ X_{30} + 0.4156\ X_{31} + 0.02461\ X_{32} + -0.09653\ X_{33} + 0.1029\ X_{34} + 0.2244\ X_{35}$
$+ 0.151\ X_{36} + 0.0914\ X_{37} + 0.0205\ X_{38} + -0.15401\ X_{40} + 0.0754\ X_{41} + 0.237\ X_{42}$

$Y_6 = \mathbf{-0.64}\ X_0 + 0.0643\ X_1 + -0.7261\ X_2 + -0.02196\ X_3 + -0.2507\ X_4 + 0.1126\ X_5$
$+ -0.13\ X_6 + 0.1552\ X_7 + 0.07851\ X_8 + 0.04499\ X_9 + -0.0233\ X_{10} + 0.2681\ X_{11}$
$+ 0.218\ X_{12} + 0.2594\ X_{13} + -0.0103\ X_{14} + -0.10622\ X_{15} + 0.0934\ X_{16} + -0.0096\ X_{17}$
$+ 0.029\ X_{18} + 0.0521\ X_{19} + 0.05007\ X_{20} + 0.09828\ X_{21} + 0.2857\ X_{22} + 0.1209\ X_{23}$
$+ 0.183\ X_{24} + 0.1317\ X_{25} + 0.32768\ X_{26} + 0.03825\ X_{27} + 0.0476\ X_{28} + -0.0254\ X_{29}$
$+ 0.206\ X_{30} + -0.1265\ X_{31} + 0.05317\ X_{32} + 0.03523\ X_{33} + 0.0852\ X_{34} + 0.0935\ X_{35}$
$+ 6E\text{-}04\ X_{36} + 0.0503\ X_{37} + -0.0562\ X_{38} + 0.35625\ X_{40} + 0.0884\ X_{41} + 0.147\ X_{42}$

$$\begin{aligned}
Y_7= & -0.17\ X_0 + -0.1221\ X_1 + -0.0389\ X_2 + 0.01496\ X_3 + 0.0036\ X_4 + -0.0209\ X_5 \\
+ & 0.162\ X_6 + 0.0215\ X_7 + 0.70812\ X_8 + \mathbf{0.60624}\ X_9 + -0.0528\ X_{10} + -0.1951\ X_{11} \\
+ & -0.16\ X_{12} + -0.1589\ X_{13} + 0.06121\ X_{14} + 0.02375\ X_{15} + -0.034\ X_{16} + 0.3182\ X_{17} \\
+ & 0.261\ X_{18} + 0.1446\ X_{19} + 0.24984\ X_{20} + 0.20507\ X_{21} + -0.0609\ X_{22} + -0.0954\ X_{23} \\
+ & 0.105\ X_{24} + 0.1811\ X_{25} + 0.11142\ X_{26} + 0.00589\ X_{27} + 0.0103\ X_{28} + -0.0789\ X_{29} \\
+ & 0.243\ X_{30} + 0.0007\ X_{31} + 0.00771\ X_{32} + -0.00326\ X_{33} + -0.003\ X_{34} + -0.1237\ X_{35} \\
+ & -0.19\ X_{36} + -0.0462\ X_{37} + 0.15002\ X_{38} + -0.05336\ X_{40} + 0.0444\ X_{41} + 0.0265\ X_{42}
\end{aligned}$$

$$\begin{aligned}
Y_8= & -0.32\ X_0 + -0.0832\ X_1 + 0.15926\ X_2 + 0.12631\ X_3 + 0.2915\ X_4 + 0.1693\ X_5 \\
+ & \mathbf{0.607}\ X_6 + \mathbf{0.6552}\ X_7 + 0.07444\ X_8 + 0.1305\ X_9 + 0.1317\ X_{10} + 0.0972\ X_{11} \\
+ & 0.061\ X_{12} + 0.0619\ X_{13} + 0.01059\ X_{14} + -0.00627\ X_{15} + -0.1274\ X_{16} + 0.0258\ X_{17} \\
+ & -0.15\ X_{18} + 0.0917\ X_{19} + 0.12522\ X_{20} + 0.14415\ X_{21} + -0.0147\ X_{22} + 0.0472\ X_{23} \\
+ & 0.141\ X_{24} + -0.0138\ X_{25} + -0.0601\ X_{26} + 0.04119\ X_{27} + -0.006\ X_{28} + 0.0321\ X_{29} \\
+ & -0.16\ X_{30} + -0.1896\ X_{31} + 0.06411\ X_{32} + -0.01349\ X_{33} + 0.063\ X_{34} + -0.1381\ X_{35} \\
+ & 0.027\ X_{36} + 0.0215\ X_{37} + 0.0768\ X_{38} + -0.03085\ X_{40} + -0.0277\ X_{41} + -0.0268\ X_{42}
\end{aligned}$$

$$\begin{aligned}
Y_9= & -0.02\ X_0 + 0.1523\ X_1 + 0.07843\ X_2 + 0.3007\ X_3 + -0.1502\ X_4 + \mathbf{-0.5924}\ X_5 \\
+ & 0.005\ X_6 + -0.0701\ X_7 + 0.0185\ X_8 + -0.09115\ X_9 + 0.1551\ X_{10} + -0.0026\ X_{11} \\
+ & 0.018\ X_{12} + 0.0037\ X_{13} + 0.13616\ X_{14} + -0.02238\ X_{15} + -0.2415\ X_{16} + 0.0854\ X_{17} \\
+ & 0.003\ X_{18} + -0.1319\ X_{19} + -0.1503\ X_{20} + 0.27512\ X_{21} + 0.3981\ X_{22} + -0.0267\ X_{23} \\
+ & -0.02\ X_{24} + 0.2124\ X_{25} + 0.03859\ X_{26} + -0.00143\ X_{27} + 0.0354\ X_{28} + 0.524\ X_{29} \\
+ & -0.05\ X_{30} + 0.1138\ X_{31} + 0.06581\ X_{32} + -0.07474\ X_{33} + 0.0386\ X_{34} + 0.0218\ X_{35} \\
+ & 0.051\ X_{36} + 0.0092\ X_{37} + -0.0531\ X_{38} + 0.12186\ X_{40} + 0.0521\ X_{41} + 0.1463\ X_{42}
\end{aligned}$$

$$\begin{aligned}
Y_{10}= & 0.179\ X_0 + 0.0501\ X_1 + -0.0095\ X_2 + -0.11588\ X_3 + -0.1359\ X_4 + 0.0264\ X_5 \\
+ & -0.11\ X_6 + 0.2761\ X_7 + 0.04283\ X_8 + -0.12029\ X_9 + -0.1594\ X_{10} + 0.0009\ X_{11} \\
+ & 0.033\ X_{12} + 0.0215\ X_{13} + -0.0213\ X_{14} + 0.05452\ X_{15} + -0.0388\ X_{16} + 0.2352\ X_{17} \\
+ & 0.043\ X_{18} + 0.0629\ X_{19} + 0.21072\ X_{20} + \mathbf{0.43219}\ X_{21} + -0.0436\ X_{22} + 0.0569\ X_{23} \\
+ & 0.083\ X_{24} + 0.0195\ X_{25} + 0.16432\ X_{26} + 0.00253\ X_{27} + 0.0138\ X_{28} + -0.0008\ X_{29} \\
+ & 0.076\ X_{30} + -0.1607\ X_{31} + -0.0113\ X_{32} + -0.00907\ X_{33} + 0.0221\ X_{34} + -0.0319\ X_{35} \\
+ & 0.047\ X_{36} + -1E\text{-}04\ X_{37} + 0.02267\ X_{38} + -0.0157\ X_{40} + -0.8133\ X_{41} + -0.1257\ X_{42}
\end{aligned}$$

$$
\begin{array}{lrlrlrlrlrlrl}
Y_{11}= & -0.02\ X_{0} & + & 0.0648\ X_{1} & + & -0.1403\ X_{2} & + & -0.19949\ X_{3} & + & 0.0464\ X_{4} & + & 0.0607\ X_{5} \\
+ & 0.259\ X_{6} & + & -0.1856\ X_{7} & + & -0.0993\ X_{8} & + & -0.00854\ X_{9} & + & 0.0837\ X_{10} & + & -0.1249\ X_{11} \\
+ & -0.09\ X_{12} & + & 0.0803\ X_{13} & + & 0.0279\ X_{14} & + & -0.0519\ X_{15} & + & 0.0372\ X_{16} & + & 0.2377\ X_{17} \\
+ & 0.091\ X_{18} & + & 0.0665\ X_{19} & + & 0.0369\ X_{20} & + & 0.12975\ X_{21} & + & -0.1202\ X_{22} & + & \mathbf{0.8021}\ X_{23} \\
+ & 0.1\ X_{24} & + & 0.0209\ X_{25} & + & -0.294\ X_{26} & + & 0.02981\ X_{27} & + & 0.0327\ X_{28} & + & 0.0631\ X_{29} \\
+ & -0.1\ X_{30} & + & 0.1001\ X_{31} & + & -0.0801\ X_{32} & + & 0.06522\ X_{33} & + & 0.0536\ X_{34} & + & -0.1971\ X_{35} \\
+ & -0.01\ X_{36} & + & 0.2209\ X_{37} & + & -0.0638\ X_{38} & + & -0.14931\ X_{40} & + & -0.0223\ X_{41} & + & -0.0276\ X_{42}
\end{array}
$$

Los coeficientes de las componentes principales de la matriz de datos estandarizadas aplicando rotación, que poseen mayor peso son:

Primera componente principal.- Nota de matemáticas e identificación de funciones y trigonometría

- Graficar funciones: lineal y cuadrática (X_{14})
- Pendiente y ecuación de la recta (X_{15})
- Identidades Trigonométricas (X_{19})
- Nota de Matemáticas (X_{24})

Segunda componente principal.- Teoría gramatical

- Diptongo (X_{32})
- Triptongo (X_{33})
- Hiato (X_{34})

Tercera componente principal.- Reconocimiento del vocabulario y oratoria

- Sinónimo (X_{36})
- Antónimo (X_{37})
- Genero de oratoria (X_{38})

Cuarta componente principal.- Funciones de la oración

- Sujeto y su respectivo núcleo
- Predicado y su núcleo

Al observar que las componentes principales no se redujeron considerablemente, se procede a verificar si el método de componentes principales es apropiado para realizarlo en este estudio, y se utiliza la prueba de Barlett para asegurar que la matriz de varianza y covarianza sea factoriazable y; la hipótesis que se plantea es la siguiente:

H_o: La matriz de varianza y covarianza es factorizable

Vs

H_1: no es verdad H_o

Los valores obtenidos aplicando la prueba de Barlett fueron:

Prueba de Bartlett	χ^2 (ji-cuadrado)	13192.9045
	grados de libertad	861
	valor p	0

El valor p es cero, lo que indica que existe evidencia estadística para rechazar la hipótesis nula, donde la matriz de varianza y covarianza no es factorizable y esto sugiere que el método de componentes principales no es apropiado en esta situación.

4.2.5 Análisis de correlación canónica

El análisis de correlación canónica es un método en el que se desea conocer la fuerza de asociación entre dos grupos de variables. El primer grupo de variables es representadas por un vector aleatorio p variado $\mathbf{X}^{(1)}$ y el segundo grupo, de q variables es representado por un vector aleatorio q variado $\mathbf{X}^{(2)}$.

El primer vector tiene un menor número de componentes que el segundo vector es decir $p \leq q$. Cada uno de los vectores $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ tiene vector de medias, y una matriz de varianzas y covarianzas lo cual se expresa como:

$$E(\mathbf{X}^{(1)}) = m^{(1)} \text{ estimada por } \overline{\mathrm{X}}^{(1)}$$
$$E(\mathbf{X}^{(2)}) = m^{(2)} \text{ estimada por} \overline{\mathrm{X}}^{(2)}$$
$$\mathrm{Cov}(\mathbf{X}^{(1)}) = \Sigma_{11} \text{ estimada por } S_{11}$$
$$\mathrm{Cov}(\mathbf{X}^{(2)}) = \Sigma_{22} \text{ estimada por } S_{22}$$
$$Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \Sigma_{12} \text{ estimada por} S_{12} = \mathrm{S}_{21}{}^{t}$$

Cuando p y q son relativamente grandes la interpretación de los elementos en la matriz Σ_{12} (estimada por $\mathbf{S}_{12}$) es muy rutinario por ese motivo se prefiere utilizar el método de correlación canónica, lo que realiza es disminuir las asociaciones entre los conjuntos de variables de $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ en unas pocas covarianzas escogidas cuidadosamente en lugar de las pq covarianzas contenidas en $\mathbf{S}_{12}$.

Al considerar a $\mathbf{X}^{(1)}$ y a $\mathbf{X}^{(2)}$ como un solo vector se tiene que:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \cdots \\ X_{q+1} \\ \vdots \\ X_p \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} q \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} p-q \end{matrix} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \cdots \\ \mathbf{X}^{(2)} \end{bmatrix}$$

El vector de medias se expresaría como:

$$\overline{X} = E[\mathbf{X}] = \begin{bmatrix} m_1 \\ \vdots \\ m_q \\ \cdots \\ m_{q+1} \\ \vdots \\ m_p \end{bmatrix} = \begin{bmatrix} \mu^{(1)} \\ \cdots \\ \mu^{(2)} \end{bmatrix} estimado - por \begin{bmatrix} \overline{c}^{(1)} \\ \cdots \\ \overline{c}^{(2)} \end{bmatrix}$$

Y la matriz de varianzas y covarianzas Σ estimada por:

$$S = \left[\begin{array}{c|c} \underset{pxp}{S_{11}} & \underset{pxq}{S_{12}} \\ \hline \underset{qxp}{S_{21}} & \underset{qxq}{S_{22}} \end{array}\right]$$

Las covarianzas entre pares de variables de diferentes conjuntos está contenida en Σ_{12} .

Considerando las siguientes combinaciones lineales: $\begin{array}{l} \hat{U} = \hat{\mathbf{a}}^t \mathbf{X}^{(1)} \\ \hat{V} = \hat{\mathbf{b}}^t \mathbf{X}^{(2)} \end{array}$, se tiene que:

$$Var(\hat{U}) = \hat{\mathbf{a}}^t \mathrm{S}_{12} \hat{\mathbf{a}}$$
$$Var(\hat{V}) = \hat{\mathbf{b}}^t \mathrm{s}_{12} \hat{\mathbf{b}}$$
$$Cov(\hat{U}, \hat{V}) = \hat{\mathbf{a}}^t \mathrm{s}_{12} \hat{\mathbf{b}}$$

$$Corr(\hat{U}, \hat{V}) = \frac{\hat{\mathbf{a}}^t \mathrm{S}_{12} \hat{\mathbf{b}}}{\sqrt{\hat{\mathbf{a}}^t \mathrm{S}_{11} \hat{\mathbf{a}}} \sqrt{\hat{\mathbf{b}}^t \mathrm{S}_{22} \hat{\mathbf{b}}}}$$

A partir de estas combinaciones, se forman las variables canónicas de la siguiente manera:

El primer par de variables canónicas, es el par de combinaciones lineales $\hat{U}_1, \hat{V}_1$ que tiene varianza unitaria y que maximiza la correlación entre ambas.

$$\underset{a,b}{Max}\, Corr(\hat{U}, \hat{V}) = r^*$$

El segundo par de variables canónicas, es el par de combinaciones lineales $\hat{U}_2, \hat{V}_2$ que tiene varianza unitaria y que maximiza la correlación entre ambas, y además en todos los casos no esta correlacionada con el primer par de variables canónicas.

En general podemos definir el k-ésimo par de variables canónicas, es el par de combinaciones lineales $\hat{U}_k, \hat{V}_k$ que tiene varianza unitaria y que maximiza la correlación entre ambas, y además en todos los casos no esta correlacionada con las k -1 pares de variables canónicas.

El k-ésimo par de variables canónicas se forma como:

$$\hat{U}_k = \mathbf{e}_k^t S_{11}^{-1/2} \mathbf{X}^{(1)}$$
$$\hat{V}_k = \mathbf{f}_k^t S_{22}^{-1/2} \mathbf{X}^{(2)}$$

A partir de la matriz resultado de la multiplicación de: $S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2}$, se obtienen los valores propios

$$\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$$

y, los vectores propios normalizados

$$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$$

$\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_p$ son los vectores propios normalizados de la matriz obtenida de la multiplicación de $S_{22}^{-1/2} S_{21} S_{11}^{-1} S_{12} S_{22}^{-1/2}$.

Con $Corr(\hat{U}_k, \hat{V}_k) = r_k^*$

Por último se puede probar que:

$$
\begin{aligned}
&Var(\hat{U}_k) = Var(\hat{V}_k) = 1 \\
&Cov(\hat{U}_k, \hat{U}_l) = Cov(\hat{U}_l, \hat{U}_k) = 0 \qquad k \neq l \\
&Cov(\hat{V}_k, \hat{V}_l) = Cov(\hat{V}_l, \hat{V}_k) = 0 \qquad k \neq l \\
&Cov(\hat{U}_k, \hat{V}_l) = Cov(\hat{U}_l, \hat{V}_k) = 0 \qquad k \neq l \\
&para \quad k, l = 1,2,\ldots, p
\end{aligned}
$$

Una vez definida toda la teoría necesaria, procedemos a analizar las variables con cada uno de los métodos descritos.

Cálculos de Correlación Canónica

A continuación vamos a calcular las variables canónicas, utilizando el software estadístico SPSS 7.0. Se tienen dos conjuntos de variables, en el que el primer vector, sería el conjunto de variables(17 variables) de la prueba de lenguaje y el segundo conjunto es el vector que corresponden a las variables de la prueba de matemáticas(20 variables). Para el estudio tenemos las combinaciones lineales de las variables de matemáticas que son un total de 20 variables y las variables de lenguaje con un total de 17.

Las variables U_k y V_k corresponden a la combinación lineal de las dos pruebas. En la tabla XIL se muestran los coeficientes de las correlaciones canónicas de cada par de variables(en total 17 variables) , para verificar que tan correlacionadas están entre ellas. Vamos a considerar las variables canónicas más significativas que poseen un coeficiente de correlación canónico mayores de 0.34 y de acuerdo a este criterio se escogen las 3 primeras variables para obtener los vectores correspondientes a ellas.

TABLA XIL

CORRELACION CANONICA

Correlación Canónica	
1	0,511
2	0,370
3	0,344
4	0,320
5	0,313
6	0,278
7	0,245
8	0,232
9	0,216
10	0,195
11	0,175
12	0,146
13	0,118
14	0,084
15	0,073
16	0,055
17	0,033

La tabla XIL muestra los coeficientes de los tres primeros pares de variables canónicas, en las cuales se puede formar sus combinaciones lineales correspondientes.

TABLA XL

COEFICIENTES DE LAS VARIABLES CANONICAS

Variables de Lenguaje	Coef. U_1	Coef. U_2	Coef. U_3
X_{25}	-0,292	-0,241	-0,059
X_{26}	-,0187	-0,042	-0,215
X_{27}	-0,198	0,073	-0,101
X_{28}	-0,275	0,113	-0,133
X_{29}	-0,096	-0,054	-0,273
X_{30}	-0,532	0,030	0,004
X_{31}	-0,101	0,136	-0,249
X_{32}	-0,391	0,071	-0,341
X_{33}	-0,053	-0,047	-0,032
X_{34}	-0,541	0,120	-0,105
X_{35}	-0,261	0,537	-0,239
X_{36}	-0,462	0,402	0,151
X_{37}	-0,180	0,180	0,269
X_{38}	-0,256	0,414	0,282
X_{39}	0,115	0,662	-0,102
X_{40}	0,181	-0,154	-0,040
X_{41}	-0,425	0,251	-0,129

Variables De Matemáticas	Coef. V_1	Coef. V_2	Coef. V_3
X_5	-0,456	-0,163	0,060
X_6	-0,186	-0,216	0,225
X_7	-0,359	0,108	-0,165
X_8	-0,173	-0,294	0,174
X_9	-0,346	-0,379	-0,155
X_{10}	-0,404	0,121	-0,400
X_{11}	-0,303	0,041	-0,219
X_{12}	-0,435	-0,115	-0,226
X_{13}	-0,677	0,190	0,072
X_{14}	-0,431	-0,147	-0,274
X_{15}	-0,536	-0,093	-0,002
X_{16}	-0,281	0,174	0,231
X_{17}	-0,379	0,135	-0,119
X_{18}	-0,441	-0,101	-0,137
X_{19}	-0,676	-0,212	-0,158
X_{20}	-0,700	0,162	0,096
X_{21}	-0,546	-0,305	-0,171
X_{22}	-0,166	0,029	-0,410
X_{23}	-0,081	-0,232	0,378
X_{24}	-0,759	-0,069	-0,143

Sabiendo que :

$$Var(\hat{U}_k) = Var(\hat{V}_k) = 1$$
$$Cov(\hat{U}_k, \hat{U}_l) = Cov(\hat{U}_l, \hat{U}_k) = 0 \qquad k \neq l$$
$$Cov(\hat{V}_k, \hat{V}_l) = Cov(\hat{V}_l, \hat{V}_k) = 0 \qquad k \neq l$$
$$Cov(\hat{U}_k, \hat{V}_l) = Cov(\hat{U}_l, \hat{V}_k) = 0 \qquad k \neq l$$
$$para \quad k, l = 1, 2, \ldots, p$$

Las combinaciones lineales formadas por cada par de variable canónica se detalla a continuación:

El primer par de variables canónicas, es el par de combinaciones lineales (U_1, V_1).

$$
\begin{aligned}
U_1 = {} & -{,}292\ X_{25} + -{,}187\ X_{26} + -{,}198\ X_{27} + -{,}275\ X_{28} + -{,}096\ X_{29} + -{,}532\ X_{30} \\
& + -{,}101\ X_{31} + -{,}391\ X_{32} + -{,}053\ X_{33} + -{,}541\ X_{34} + -{,}261\ X_{35} + -{,}462\ X_{36} \\
& + -{,}180\ X_{37} + -{,}256\ X_{38} + {,}115\ X_{39} + {,}181\ X_{40} + -{,}425\ X_{41}
\end{aligned}
$$

$$
\begin{aligned}
V_1 = {} & -{,}456\ X_{5} + -{,}186\ X_{6} + -{,}359\ X_{7} + -{,}173\ X_{8} + -{,}346\ X_{9} + -{,}404\ X_{10} \\
& + -{,}303\ X_{11} + -{,}435\ X_{12} + -{,}677\ X_{13} + -{,}431\ X_{14} + -{,}536\ X_{15} + -{,}281\ X_{16} \\
& + -{,}379\ X_{17} + -{,}441\ X_{18} + -{,}676\ X_{19} + -{,}700\ X_{20} + -{,}546\ X_{21} + -{,}166\ X_{22} \\
& + -{,}081\ X_{23} + -{,}759\ X_{24}
\end{aligned}
$$

$Corr(U_1, V_1) = 0.511$

Las variables que tienen mayor peso en el primer conjunto (la prueba de lenguaje) representado por U_1 son:

- La variable de hiato (X_{34}), el peso de la variable es de –0.541
- Corrección de palabras (X_{30}), el peso de la variable es de –0.532
- La variable sinónimo (X_{36}), con un coeficiente de carga –0.462

Del segundo conjunto de variables representado con V_1 se ha escogido las variables cuyo coeficiente es mayor a 0.6 y menor a –0.6:

- La variable X_{20} es el factor de caga con mayor peso y representa a la pregunta de superficie con un coeficiente de –0.7,
- La segunda variable con mayor carga es la de identificar el gráfico (X_{13}) con un coeficiente de -0.677
- La variable X_{19} que representa a Identidades trigonométricas con un coeficiente de –0.676

Las variables que representan mayor peso en cada par de combinaciones lineales son las que están fuertemente correlacionadas, es decir maximiza la correlación entre ambos vectores con un coeficiente de 0.511.

El segundo par de variables canónicas, es el par de combinaciones lineales formadas de la siguiente manera:

$$U_2 = -,241\ X_{25} + -,042\ X_{26} + ,073\ X_{27} + ,113\ X_{28} + -,054\ X_{29} + ,030\ X_{30}$$
$$+ ,136\ X_{31} + ,071\ X_{32} + -,047\ X_{33} + ,120\ X_{34} + ,537\ X_{35} + ,402\ X_{36}$$
$$+ ,180\ X_{37} + ,414\ X_{38} + ,662\ X_{39} + -,154\ X_{40} + ,251\ X_{41}$$

$$V_2 = -,163\ X_5 + -,216\ X_6 + ,108\ X_7 + -,294\ X_8 + -,379\ X_9 + ,121\ X_{10}$$
$$+ ,041\ X_{11} + -,115\ X_{12} + ,190\ X_{13} + -,147\ X_{14} + -,093\ X_{15} + ,174\ X_{16}$$
$$+ ,135\ X_{17} + -,101\ X_{18} + -,212\ X_{19} + ,162\ X_{20} + -,305\ X_{21} + ,029\ X_{22}$$
$$+ -,232\ X_{23} + -,069\ X_{24}$$

Este par de variables canónicas tiene varianza unitaria y en todos los casos no está correlacionada con el primer par de variables canónicas.

En el primer vector representado por U_2 las variables con un coeficiente superior a 0.5 son escogidos, porque aportan mayor carga:

- La variable X_{39} Obras literarias es la que aporta mayor carga con un coeficiente de 0.662, es la variable de la pregunta obras literarias.
- Y otra variable que está dentro del intervalo de mayor carga es X_{35} que representa a la pregunta de identificar la palabra a partir del contexto, con un coeficiente de 0.537

Para el segundo vector de combinaciones lineales V_2 los coeficientes son mayores a 0.3 y menores a –0.3:

- X_9 que corresponde a la pregunta de conjuntos con un coeficiente de -0.379 y la variable X_{21} que representa a la pregunta de volumen de un cubo con un coeficiente de –0.305.

Las variables de cada conjunto se encuentran fuertemente correlacionadas entre ellas, con un coeficiente de correlación de 0.37 que maximiza a ambas.

En el tercer par de variables canónicas que se ilustra a continuación , forman el conjunto de combinaciones lineales, y muestran las variables que aportan con mayor peso detallándose de la siguiente manera:

$$\begin{aligned} U_3 = {} & -,059\ X_{25} + -,215\ X_{26} + -,101\ X_{27} + -,133\ X_{28} + -,273\ X_{29} + ,004\ X_{30} \\ & + -,249\ X_{31} + -,341\ X_{32} + -,032\ X_{33} + -,105\ X_{34} + -,239\ X_{35} + ,151\ X_{36} \\ & + ,269\ X_{37} + ,282\ X_{38} + -,102\ X_{39} + -,040\ X_{40} + -,129\ X_{41} \end{aligned}$$

$$\begin{aligned} V_3 = {} & ,060\ X_{5} + ,225\ X_{6} + -,165\ X_{7} + ,174\ X_{8} + -,155\ X_{9} + -,400\ X_{10} \\ & + -,219\ X_{11} + -,226\ X_{12} + ,072\ X_{13} + -,274\ X_{14} + -,002\ X_{15} + ,231\ X_{16} \\ & + -,119\ X_{17} + -,137\ X_{18} + -,158\ X_{19} + ,096\ X_{20} + -,171\ X_{21} + -,410\ X_{22} \\ & + -,143\ X_{23} + ,378\ X_{24} \end{aligned}$$

En el primer vector del tercer par de variables canónicas los coeficientes con un valor superior a 0.28 e inferior a –0.28, son las variables de X_{32} representando al diptongo y X_{38} que representa los géneros literarios.

En el segundo conjunto de combinaciones lineales, las variables con coeficientes mayores a 0.35 y menores a –0.35, son X_{22} que corresponde a la media aritmética, X_{10} a la desigualdad de conjuntos y X_{24} que representa la nota de matemáticas. Las variables de ambos conjuntos, es decir el tercer par de combinaciones lineales U_3 ,V_3 están correlacionadas fuertemente entre ellas con un coeficiente de 0.344.

4.2.5 Análisis de Varianza

El modelo que se analiza, es un diseño factorial con dos factores; se investigan todas las posibles combinaciones de los niveles de factores en cada réplica del estudio.

El modelo de análisis de varianza de dos factores se expresa:

$$y_{ijk} = m + t_i + b_j + (tb)_{ij} + e_{ijk} \begin{cases} i = 1,2,\ldots, a \\ j = 1,2,\ldots, b \\ k = 1,2,\ldots, n \end{cases}$$

$$\varepsilon_{ijk} \sim N(0,\sigma^2)$$

Donde μ es el efecto medio general, τ_i es el efecto del i-ésimo nivel del factor A, β_j es el efecto del j-ésimo nivel del factor B, $(\tau\beta)_{ij}$ es el efecto de la iteración entre τ_i y β_j y $\in_{ijk}$ es el componente del error aleatorio, con distribución normal, media cero y varianza σ^2. Los valores de a,b,n corresponden a los niveles de cada factor y hay un total de abn observaciones porque se realizan n réplicas.

El interés consiste en probar hipótesis acerca de la igualdad de los efectos de los tratamientos y de las interacciones, las cuales se ilustran a continuación:

La prueba de hipótesis de cada efecto de tratamiento es:

$$H_o : t_1 = t_2 = \ldots = t_a$$

$$H_1 : t_i \neq 0 \quad \text{Para al menos un } \tau_i$$

$$H_o : b_1 = b_2 = \ldots = b_b$$

$$H_1 : b_j \neq 0 \quad \text{Para al menos un } \beta_i$$

La prueba de hipótesis del efecto de la interacción entre τ_i y β_j :

$$H_o : (tb)_{ij} = 0$$

$$H_1 : (tb)_{ij} \neq 0 \quad \text{Para al menos un } (\tau\beta)_{ij}$$

La suma total corregida muestra la variabilidad total de los datos, medida por la suma de los cuadrados de las diferencias entre los promedios de los tratamientos y el promedio general(denominada suma de cuadrados de tratamientos); y en la suma de cuadrados de las diferencias entre las observaciones dentro del tratamiento y el promedio del mismo (denominada suma cuadrados del error).

Es decir:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij}-\overline{Y}_{..})^2 = n\sum_{i=1}^{a}(\overline{Y}_{i.}-\overline{Y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij}-\overline{Y}_{i.})^2$$

Suma de cuadrados total= S. Cuadrados tratamientos + S Cuadrados total

Y_{ij} es el total de observaciones del i-ésimo tratamiento

$\overline{Y}_{i.}$ el promedio de las observaciones bajo el i-ésimo tratamiento

$\overline{Y}_{..}$ es la media general de las observaciones

Se trabaja con un modelo de efectos fijos, porque han sido considerados todos los niveles de cada factor en el análisis. La tabla de análisis de varianza para el modelo de efectos fijos con dos factores es la siguientes:

TABLA ANOVA

DISEÑO BIFACTORIAL

Fuente de variación	**Suma de cuadrados**	**Grados de libertad**	**Media cuadrática**	**F_0**
Tratamiento A	SC_A	a-1	MC_A	MC_A /SC_A
Tratamiento B	SC_B	b-1	MC_B	MC_B /SC_B
Iteración	SC_{AB}	(a-1)(b-1)	MC_{AB}	MC_{AB} /SC_{AB}
Error	SC_E	Ab(n-1)	MC_E	
Total	SC_T	Abn-1		

A continuación se realiza el cálculo correspondiente al modelo planteado. Para el primer análisis; los factores o tratamientos que se utilizan son la especialización(X_1) y la actividad extra curricular(X_4), donde la variable dependiente Y_{ijk} es la nota de matemáticas(X_{24}), la cual es una variable cuantitativa. Los resultados del análisis de varianza se muestran en la tabla XXXVIII, verificando cada factor e iteracción , la variable o factor X_4 tiene un valor p de 0.17 y se concluye que existe suficiente evidencia estadística para aceptar la hipótesis nula es decir que los efectos del tratamiento son cero, y no influye en el análisis dicho factor.

TABLA XLI

TABLA ANOVA PARA EL PRIMER MODELO (BIFACTORIAL)

Y_{ijk}: Variable Dependiente: NOTA DE MATEMATICAS					
Fuente de Variación	**Suma cuadrática**	**Grados de libertad**	**Media cuadrática**	**F**	**Valor p**
Tratamiento A	70345.78	6	11724.30	67.92	0.00
Tratamiento B	331.67	1	331.67	1.92	0.17
Interacción A *B	870.25	6	145.04	0.84	0.54
Error	89417.71	518	172.62		
Total	401137.43	532			

A: Especialización

B: Actividad extra-educativa

Para lo cual podemos concluir que la variable especialización (X_1) si influye en la nota de matemáticas, pero no sucede lo mismo con la otra variable; posteriormente se realizará el análisis de la mínima diferencia significativa (LSD) del factor que si influye en el modelo.

Aplicamos un modelo bifactorial para el análisis de varianza con la variable dependiente Y_{ijk} nota de lenguaje (X_{41}), utilizando los mismos factores descritos anteriormente, la tabla XLII muestra los resultados.

TABLA XLII

TABLA ANOVA PARA EL SEGUNDO MODELO (BIFACTORIAL)

Y_{ijk}: Variable Dependiente: NOTA DE LENGUAJE					
Fuente de Variación	**Suma cuadrática**	**Grados de libertad**	**Media cuadrática**	**F**	**Valor p**
Tratamiento A	9344.51	6	1557.42	5.22	0.00
Tratamiento B	37.76	1	37.76	0.13	0.72
A*B	4659.46	6	776.58	2.60	0.02
Error	154480.96	518	298.23		
Total	2319554.98	532			

A: Especialización

B: Actividad extra-educativa

Con la variable dependiente X_{41}, el modelo bifactorial al igual que el anterior, la variable actividad extra-educativa no influye en el modelo, y la variable especialización si , esto se lo puede observar en la tabla XLII la cual indica que la variable X_4 tiene un valor p 0.72 y posee suficiente evidencia estadística para aceptar la hipótesis nula donde indica que los efectos de los tratamientos son iguales a cero, a diferencia con la variable especialización que si influye con la variable nota de lenguaje.

El último modelo que se plantea es con respecto a la variable dependiente nota general (X_{42}); los factores que se analizan son la especialización y la actividad extra-educativa, los cálculos se muestran en la tabla XLIII, al igual que las anteriores indica que el efecto del tratamiento X_4 no afecta el modelo con un valor p de 0.604, y muestra que existe suficiente evidencia estadística para aceptar la hipótesis nula, y se concluye que los efectos del tratamiento son cero, pero el efecto del tratamiento X_1, que constituye a la variable especialización si influye en el modelo.

TABLA XLIII

TABLA ANOVA PARA EL TERCER MODELO (BIFACTORIAL)

Y_{ijk}: Variable Dependiente: NOTA GENERAL					
Fuente de Variación	**Suma cuadrática**	**Grados de libertad**	**Media cuadrática**	**F**	**Valor p**
Tratamiento A	25812.05	6	4302.01	31.78974	0.00000
Tratamiento B	36.40	1	36.40	0.26899	0.60423
A*B	1705.28	6	284.21	2.10020	0.05173
Error	70099.34	518	135.33		
Total	1038279.23	532			

A: Especialización

B: Actividad extra-educativa

A continuación procedemos a realizar el método de mínima diferencia significativa con las variables que influyen en los modelos.

4.2.5.1 Método de mínima diferencia significativa LSD

Una vez analizados los modelos, se procede a realizar la comparación de parejas de medias de los tratamientos, la cual inicia una vez rechazada la hipótesis nula en el análisis de varianza, y se desea probar la siguiente hipótesis:

$$
\begin{aligned}
&H_0 : m_i = m_j \\
&vs \\
&H_1 : m_i \neq m_j \\
&i \neq j
\end{aligned}
$$

El estadístico de prueba que se emplea es el siguiente:

$$t_0 = \frac{\overline{y_{i.}} - \overline{y_{j.}}}{\sqrt{MC_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Para utilizar el procedimiento de LSD, se comparan las diferencias observadas entre cada par de promedios con el valor de la LSD. Donde:

$$LSD = t_{a/2, N-a} \sqrt{Mc_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Y si

$$\left| \overline{y_{i.}} - \overline{y_{j.}} \right| > LSD$$

Se concluye que las medias muestrales son diferentes.

En el estudio que se está realizando se procede a analizar el LSD verificando la hipótesis. El factor que se va a analizar es de la variable X_1, que corresponde a la especialización. Este factor tiene a niveles de tratamientos que representan a las 7 especializaciones detalladas en el capítulo 2, la tabla XLIV muestra los parámetros de cada tratamiento y la media cuadrática del error que se utilizará de acuerdo al modelo que corresponda.

TABLA XLIV

ESTIMADORES PARA EL ANALISIS LSD DEL PRIMER MODELO

Nivel del Tratamiento	$\overline{Y}_{i.}$	n_i	
1	46.1250	101	
2	25.8515	55	**MS_E**
3	8.1632	41	**172.621**
4	11.6259	143	
5	14.9476	167	
6	5.2306	16	
7	7.31	9	

Para el caso en estudio, reemplazando los valores para obtener el LSD se tiene:

$$LSD = t_{0.025,14-7}\sqrt{172.621\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

La tabla XLV muestra los valores correspondiente a la diferencia de medias entre parejas, y su valor correspondiente de LSD, para cada una de las parejas con diseño no balanceado. La columna de resultado muestra un asterisco(*) a las parejas de medias que son significativamente diferentes y, las parejas que no lo posean son las que no difieren en forma significativa.

Como se realizo el análisis de varianza del primer modelo, el primer factor "especialización" si influía en él; ahora determinamos cuanto difieren las parejas de medias en dicho factor, como se ilustra en la tabla. Las parejas de medias que no difieren significativamente son:

$\overline{Y}_{3.}$, $\overline{Y}_{4.}$ representan a ciencias sociales y contabilidad

$\overline{Y}_{3.}$, $\overline{Y}_{6.}$ representa a ciencias sociales y secretariado

$\overline{Y}_{3.}$, $\overline{Y}_{7.}$ representa a ciencias sociales y técnico

$\overline{Y}_{4.}$, $\overline{Y}_{5.}$ representa a contabilidad e informática

$\overline{Y}_{4.}$, $\overline{Y}_{6.}$ representa a contabilidad y secretariado

$\overline{Y}_{4.}$, $\overline{Y}_{7.}$ representa a contabilidad y técnico

$\overline{Y}_{5.}$, $\overline{Y}_{7.}$ representa a informática y técnico

$\overline{Y}_{6.}$, $\overline{Y}_{7.}$ representa a secretariado y técnico

TABLA XLV

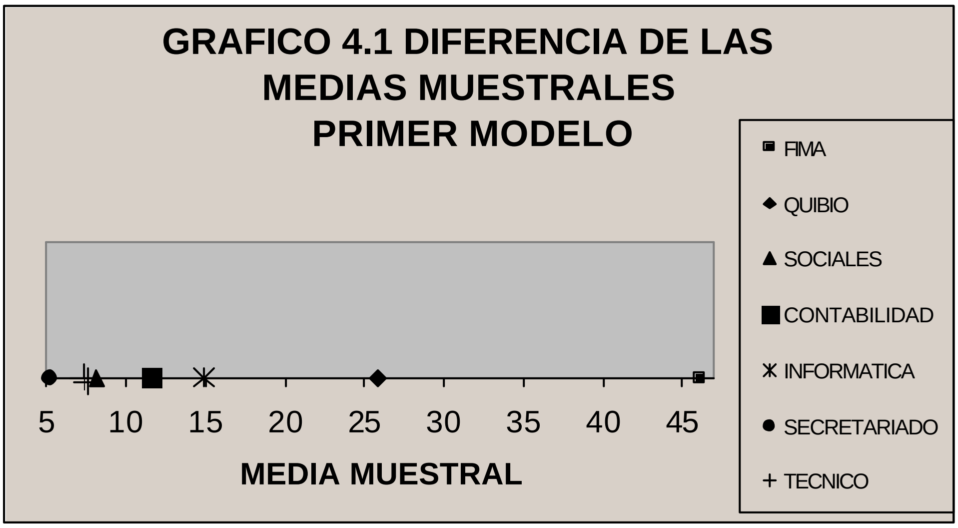
METODO LSD PARA EL PRIMER MODELO

(Variable Dependiente NOTA DE MATEMATICAS)

	Diferencia de promedios	LSD	Valor p	Resultado
$\overline{Y}_{1.} - \overline{Y}_{2.}$	21.1121404	5.20712768	0	*
$\overline{Y}_{1.} - \overline{Y}_{3.}$	37.9618788	5.75399548	0	*
$\overline{Y}_{1.} - \overline{Y}_{4.}$	34.4991754	4.03872392	0	*
$\overline{Y}_{1.} - \overline{Y}_{5.}$	31.1774447	3.91675525	0	*
$\overline{Y}_{1.} - \overline{Y}_{6.}$	40.8944245	8.36084576	0	*
$\overline{Y}_{1.} - \overline{Y}_{7.}$	38.8150495	10.8091703	0	*
$\overline{Y}_{2.} - \overline{Y}_{3.}$	16.8497384	6.41121443	0	*
$\overline{Y}_{2.} - \overline{Y}_{4.}$	13.387035	4.93016201	0	*
$\overline{Y}_{2.} - \overline{Y}_{5.}$	10.0653043	4.83075338	0	*
$\overline{Y}_{2.} - \overline{Y}_{6.}$	19.7822841	8.82603809	0	*
$\overline{Y}_{2.} - \overline{Y}_{7.}$	17.7029091	11.1728855	0	*
$\overline{Y}_{3.} - \overline{Y}_{4.}$	3.46270339	5.50461467	0.103	
$\overline{Y}_{3.} - \overline{Y}_{5.}$	6.78443406	5.41576065	0.001	*
$\overline{Y}_{3.} - \overline{Y}_{6.}$	2.93254573	9.15932496	0.407	
$\overline{Y}_{3.} - \overline{Y}_{7.}$	0.85317073	11.4379917	0.847	
$\overline{Y}_{4.} - \overline{Y}_{5.}$	3.32173066	3.54023875	0.015	
$\overline{Y}_{4.} - \overline{Y}_{6.}$	6.39524913	8.19121851	0.043	
$\overline{Y}_{4.} - \overline{Y}_{7.}$	4.31587413	10.6785056	0.295	
$\overline{Y}_{5.} - \overline{Y}_{6.}$	9.71697979	8.13177357	0.002	*
$\overline{Y}_{5.} - \overline{Y}_{7.}$	7.63760479	10.6329752	0.063	
$\overline{Y}_{6.} - \overline{Y}_{7.}$	2.079375	12.9469275	0.677	

* Pares de medias del factor especialización que difieren significativamente

Las variables que difieren significativamente en el primer modelo son: físico matemático y químico biológico con el resto de especializaciones y de esa manera se observa la diferencia que existe con las especializaciones de acuerdo al promedio en lo que tiene que ver con la nota de matemáticas, como se muestra en el gráfico 4.1.



Para realizar el análisis de la comparación de medias entre tratamientos del segundo modelo en la tabla XLVI se muestra, el promedio de cada uno de los 7 tratamientos en el factor especialización, con la variable dependiente nota de lenguaje y el número de replicas en cada tratamiento, con su valor correspondiente de la media cuadrática del error obtenida anteriormente.

TABLA XLVI

ESTIMADORES PARA EL ANALISIS DEL LSD DEL SEGUNDO MODELO

Nivel del Tratamiento	$\overline{Y}_{i.}$	n_i	
1	68.18673	101	**MS_E** **298.226**
2	70.42794	55	
3	67.84415	41	
4	61.7386	143	
5	60.25587	167	
6	63.88875	16	
7	49.55111	9	

Los resultados de los pares de medias de los tratamientos se muestra en la tabla XLVII la cual expresa, los que tienen un asterisco en la columna de resultado, poseen diferencia significativa ; comparando con los resultados de la tabla XLV, indica que en este último modelo las comparaciones entre los pares de medias de los tratamientos no difieren mucho en relación al modelo anterior, en este último los promedios no

TABLA XLVII

METODOD LSD PARA EL SEGUNDO MODELO

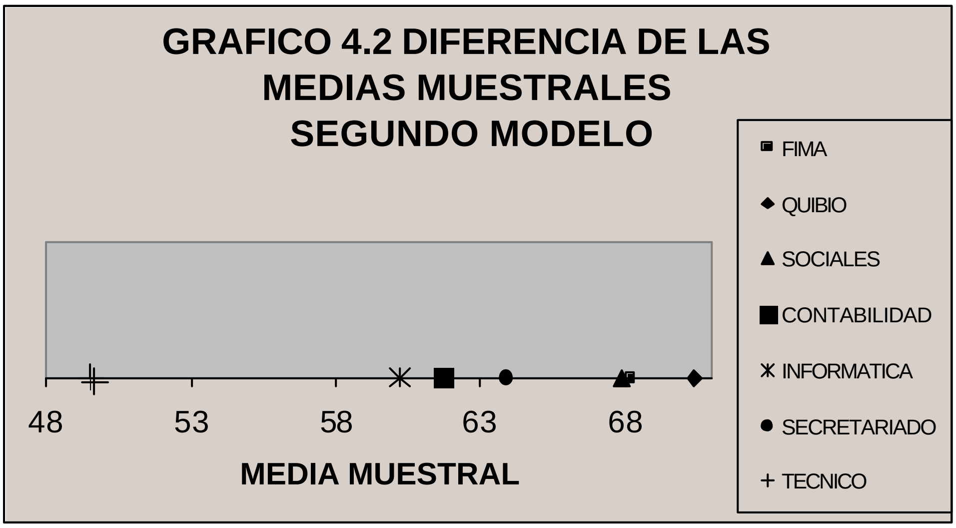
(Variable Dependiente NOTA DE LENGUAJE)

	Diferencia de promedios	LSD	Valor p	Resultado
$\overline{Y}_{1.} - \overline{Y}_{2.}$	2.241	6.844	0.727	
$\overline{Y}_{1.} - \overline{Y}_{3.}$	0.343	7.563	0.914	
$\overline{Y}_{1.} - \overline{Y}_{4.}$	6.448	5.308	0.004	*
$\overline{Y}_{1.} - \overline{Y}_{5.}$	7.931	5.148	0.000	*
$\overline{Y}_{1.} - \overline{Y}_{6.}$	4.298	10.989	0.352	
$\overline{Y}_{1.} - \overline{Y}_{7.}$	18.636	14.208	0.002	*
$\overline{Y}_{2.} - \overline{Y}_{3.}$	2.584	8.427	0.704	
$\overline{Y}_{2.} - \overline{Y}_{4.}$	8.689	6.480	0.006	*
$\overline{Y}_{2.} - \overline{Y}_{5.}$	10.172	6.350	0.001	*
$\overline{Y}_{2.} - \overline{Y}_{6.}$	6.539	11.601	0.277	
$\overline{Y}_{2.} - \overline{Y}_{7.}$	20.877	14.686	0.002	*
$\overline{Y}_{3.} - \overline{Y}_{4.}$	6.106	7.235	0.045	
$\overline{Y}_{3.} - \overline{Y}_{5.}$	7.588	7.118	0.011	*
$\overline{Y}_{3.} - \overline{Y}_{6.}$	3.955	12.039	0.434	
$\overline{Y}_{3.} - \overline{Y}_{7.}$	18.293	15.034	0.004	*
$\overline{Y}_{4.} - \overline{Y}_{5.}$	1.483	4.653	0.448	
$\overline{Y}_{4.} - \overline{Y}_{6.}$	2.150	10.766	0.634	
$\overline{Y}_{4.} - \overline{Y}_{7.}$	12.187	14.036	0.039	
$\overline{Y}_{5.} - \overline{Y}_{6.}$	3.633	10.688	0.418	
$\overline{Y}_{5.} - \overline{Y}_{7.}$	10.705	13.976	0.068	
$\overline{Y}_{6.} - \overline{Y}_{7.}$	14.338	17.017	0.045	

* Pares de medias del factor especialización que difieren significativamente

son tan diferentes entre especializaciones, por ese motivo, los pares que poseen diferencia son 8 de 21 :

- Físico matemático y químico biológico difieren con contabilidad, informática y técnico
- Ciencias sociales difiere con informática y con técnico



La última diferencia significativa se realiza para el modelo con la variable dependiente Y_{ijk} que representa a la variable X_{42} nota general y como se observa en la tabla XLVIII los parámetros para el análisis del LSD que representan los niveles del tratamiento (X_1) con su correspondiente promedio, y el valor de la media cuadrática del error mostrada anteriormente en el modelo planteado.

TABLA XLVIII

ESTIMADORES PARA EL ANALISIS LSD DEL TERCER MODELO

Nivel del Tratamiento	$\overline{Y}_{i.}$	n_i	
1	57.15589	101	
2	47.10055	55	**MS_E**
3	38.00366	41	
4	36.68224	143	**135.33**
5	37.60174	167	
6	34.55969	16	
7	28.43056	9	

La conclusión que se puede recalcar de la tabla IL es que la mitad de los pares de medias de tratamientos tiene diferencia significativa y la otra mitad no la tiene, es decir el promedio de la nota general entre los pares que poseen diferencia significativa son:

TABLA IL

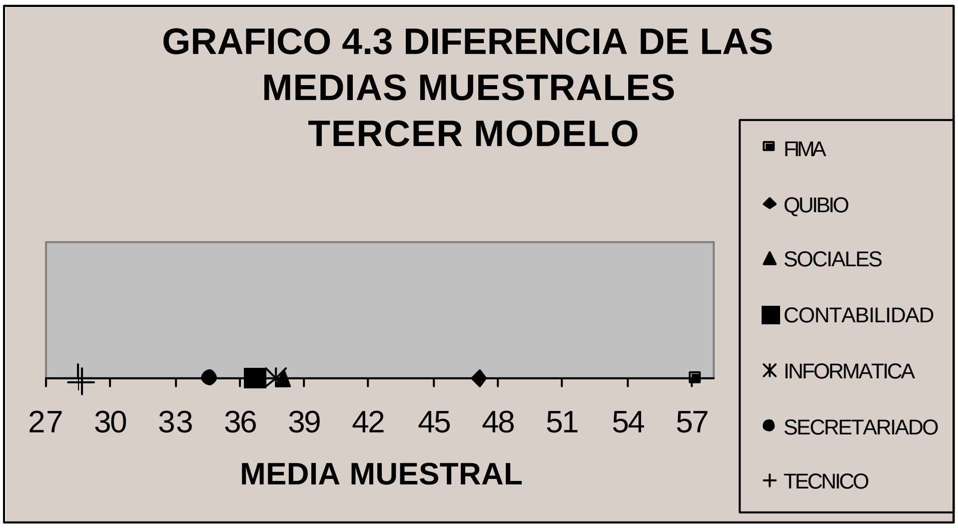
METODOD LSD PARA EL TERCER MODELO

(Variable Dependiente NOTA GENERAL)

	Diferencia de promedios	LSD	Valor p	Resultado
$\overline{Y}_{1.} - \overline{Y}_{2.}$	10.055	4.611	0	*
$\overline{Y}_{1.} - \overline{Y}_{3.}$	19.152	5.095	0	*
$\overline{Y}_{1.} - \overline{Y}_{4.}$	20.474	3.576	0	*
$\overline{Y}_{1.} - \overline{Y}_{5.}$	19.554	3.468	0	*
$\overline{Y}_{1.} - \overline{Y}_{6.}$	22.596	7.403	0	*
$\overline{Y}_{1.} - \overline{Y}_{7.}$	28.725	9.571	0	*
$\overline{Y}_{2.} - \overline{Y}_{3.}$	9.097	5.677	0	*
$\overline{Y}_{2.} - \overline{Y}_{4.}$	10.418	4.365	0	*
$\overline{Y}_{2.} - \overline{Y}_{5.}$	9.499	4.277	0	*
$\overline{Y}_{2.} - \overline{Y}_{6.}$	12.541	7.815	0	*
$\overline{Y}_{2.} - \overline{Y}_{7.}$	18.670	9.893	0	*
$\overline{Y}_{3.} - \overline{Y}_{4.}$	1.321	4.874	0.506	
$\overline{Y}_{3.} - \overline{Y}_{5.}$	0.402	4.795	0.837	
$\overline{Y}_{3.} - \overline{Y}_{6.}$	3.444	8.110	0.298	
$\overline{Y}_{3.} - \overline{Y}_{7.}$	9.573	10.127	0.021	
$\overline{Y}_{4.} - \overline{Y}_{5.}$	0.919	3.135	0.472	
$\overline{Y}_{4.} - \overline{Y}_{6.}$	2.123	7.253	0.473	
$\overline{Y}_{4.} - \overline{Y}_{7.}$	8.252	9.455	0.033	
$\overline{Y}_{5.} - \overline{Y}_{6.}$	3.042	7.200	0.300	
$\overline{Y}_{5.} - \overline{Y}_{7.}$	9.171	9.415	0.017	
$\overline{Y}_{6.} - \overline{Y}_{7.}$	6.129	11.463	0.190	

* Pares de medias del factor especialización que difieren significativamente

- La especialización de físico matemático tiene diferencia significativa con las demás especializaciones.
- Químico biológico tiene diferencia significativa con todas las especializaciones restantes.



En cambio las especializaciones de: ciencias sociales, contabilidad, informática, secretariado y técnico no tienen diferencia significativa , es decir que entre estos niveles de tratamientos no existe mayor variación en el promedio de la nota general como se muestra en el gráfico 4.3