

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

TESINA QUE PRESENTA:

LIC. MIGUEL ZAMORA JUÁREZ

PARA OBTENER EL DIPLOMA DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

CON EL TEMA:

**“LA DEMANDA DE CARRERAS UNIVERSITARIAS Y SU RELACIÓN CON EL PROMEDIO,
SEXO Y NIVEL SOCIOECONÓMICO DE LOS ALUMNOS DE BACHILLERATO
INCORPORADO A LA UNAM”**

ASESOR: DR. IGNACIO MENDEZ RAMÍREZ



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS	3
INTRODUCCIÓN.....	4
MARCO CONCEPTUAL	5
La demanda de carreras universitarias por sexo y nivel socioeconómico.....	5
El promedio por área de estudio, sexo y nivel socioeconómico	9
Los criterios para elegir otra universidad en caso de no ser aceptado por la UNAM	9
MÉTODO.....	10
Selección de la muestra	10
Proceso de muestreo	12
Limitaciones al tomar la muestra	12
ESTADÍSTICAS DESCRIPTIVAS.....	12
Generalidades	12
Promedio de calificaciones	16
Preferencia de carreras.....	18
DEMANDA DE CARRERAS POR SEXO Y POR NIVEL SOCIOECONÓMICO	24
Demanda de carreras por sexo	24
Demanda de carreras por nivel socioeconómico.....	27
PROMEDIO DE CALIFICACIONES POR ÁREA DE ESTUDIO, SEXO Y NIVEL SOCIOECONÓMICO	30
Promedio por área de estudio.....	30
Promedio por sexo	32
Promedio por nivel socioeconómico	33
CRITERIOS PARA ELEGIR OTRA UNIVERSIDAD EN CASO DE NO SER ACEPTADO POR LA UNAM	35
CONCLUSIONES	46
ANEXO	48
Cuestionario sobre demanda de carreras universitarias en instituciones incorporadas a la UNAM.	49
ANÁLISIS DE CORRESPONDENCIA	52
BIBLIOGRAFÍA.....	63

AGRADECIMIENTOS

Deseo manifestar mi agradecimiento al IIMAS por brindarme la oportunidad de acrecentar mi conocimiento. Especialmente agradezco la paciencia y orientación de:

Dr. Ignacio Méndez Ramírez

M. en C. Patricia Isabel Romero Mares

M. en C. Leticia Eugenia Gracia Medrano Valdelamar

Dr. Juan González Hernández

Dr. Ramsés Humberto Mena Chavez

Que me apoyaron en la realización de este trabajo.

Agradezco a mi familia y amigos por su apoyo incondicional.

INTRODUCCIÓN

El presente trabajo tiene como finalidad mostrar un análisis estadístico acerca de las preferencias que tienen los alumnos de bachillerato sobre la oferta de carreras universitarias. El estudio se centra en las instituciones del Sistema Incorporado de la UNAM (SI) que se ubican en el Distrito Federal y se pretende conocer lo siguiente:

- La demanda de carreras universitarias por sexo y nivel socioeconómico;
- El promedio de calificaciones de los alumnos, por área de estudio, sexo y nivel socioeconómico; y
- Los criterios para elegir otra universidad en caso de no ser aceptado por la UNAM

Para identificar los planteles y el número de alumnos inscritos se consultó, durante el primer semestre del 2006, la información que presenta en su página web la Dirección General de Incorporación y Revalidación de Estudios de la UNAM (DGIRE)¹, que es la que se encarga de regular y supervisar a las instituciones incorporadas de nivel medio superior y superior. En esas fechas se identificaron 202 planteles de bachillerato en el Distrito Federal pertenecientes al SI, de ellos, 33 impartían el plan de estudios del CCH y 169 el plan de preparatoria, con una matrícula total de 16,042 alumnos. La encuesta se realizó durante el segundo semestre de 2006.

Por otra parte, se verificó que las carreras universitarias mencionadas en los resultados estuvieran registradas por la UNAM y la SEP².

¹ www.dgire.unam.mx

² www.unam.mx , <http://ses4.sep.gob.mx>

MARCO CONCEPTUAL

Para realizar la encuesta se aplicó un cuestionario de 11 preguntas, el cual se puede consultar en el anexo de este trabajo.

A continuación se definen los criterios utilizados para realizar el análisis estadístico.

La demanda de carreras universitarias por sexo y nivel socioeconómico

Las carreras universitarias que desean estudiar los alumnos (pregunta 7 del cuestionario) se clasificaron como sigue:

Área de la carrera deseada (área-carrera)	
0	No contestó
1	Área de Ciencias físico-matemáticas
2	Ciencias biológicas y de la salud
3	Ciencias sociales
4	Humanidades y artes
5	Otras
6	no sabe qué carrera estudiar

En la quinta clasificación se encuentran las carreras de gastronomía, turismo/hotelería, sobrecargo y piloto aviador, que formaron parte de las respuestas y que no las imparte la UNAM.

En este apartado interesa conocer la elección de carrera por sexo y también su posible relación con el nivel socioeconómico del alumno.

El nivel socioeconómico se estimó a partir de las características que manifestaron los estudiantes sobre su hogar. Para ello, en la pregunta 10 del cuestionario se les pidió que indicaran el número de recámaras, baños, televisores, computadoras y autos con que cuentan en su domicilio. Estas respuestas caen dentro de una escala de "0" a "4 o más", de tal forma que los alumnos con valores más cercanos a cero tendrían un nivel socioeconómico más bajo. Para tomar una referencia se

consultaron los niveles socioeconómicos que determina la Asociación Mexicana de Agencias de Investigación de Mercado y Opinión Pública, A. C. (AMAI). Estos se encuentran en el documento titulado "Nivel socioeconómico AMAI" de septiembre de 2009 y los clasifica de mayor a menor como sigue:

	Nivel
1 (más alto)	A/B
2	C+
3	C
4	D+
5	D
6 (más bajo)	E

Para comparar las características de los alumnos encuestados con los niveles de la AMAI, se tomaron los promedios por hogar que esta Asociación menciona en su documento, los cuales son:

nse	promedio AMAI				
	recámaras	baños	computadora	televisor	automóvil
A/B	2.89	3.09	1.32	3.27	1.46
C+	2.52	1.73	0.84	2.52	1.09
C	2.31	1.36	0.44	2.00	0.74
D+	1.96	1.13	1.15	1.47	0.34
D	1.53	0.94	0.03	1.07	0.08
E	1.25	0.81	0.01	0.61	0.01

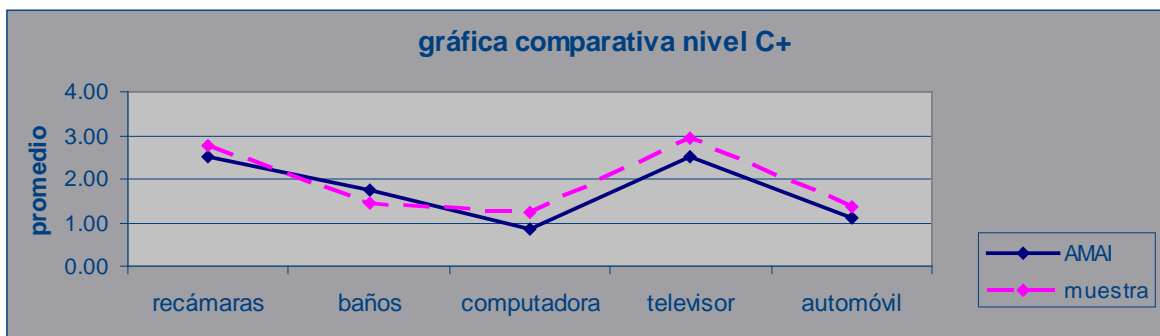
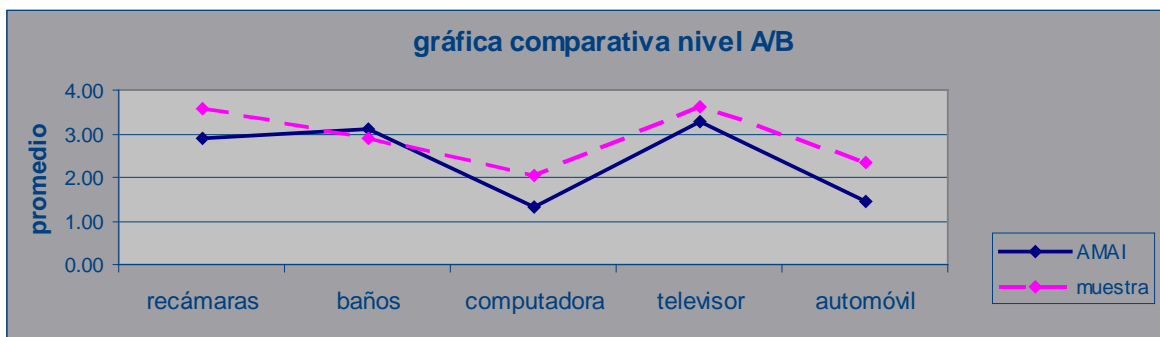
Posteriormente se calculó la distancia euclidiana entre las respuestas de cada estudiante y los promedios arriba mencionados. La distancia menor se consideró para agrupar a los estudiantes en un determinado nivel de la AMAI. Para ilustrar esto se toman dos ejemplos:

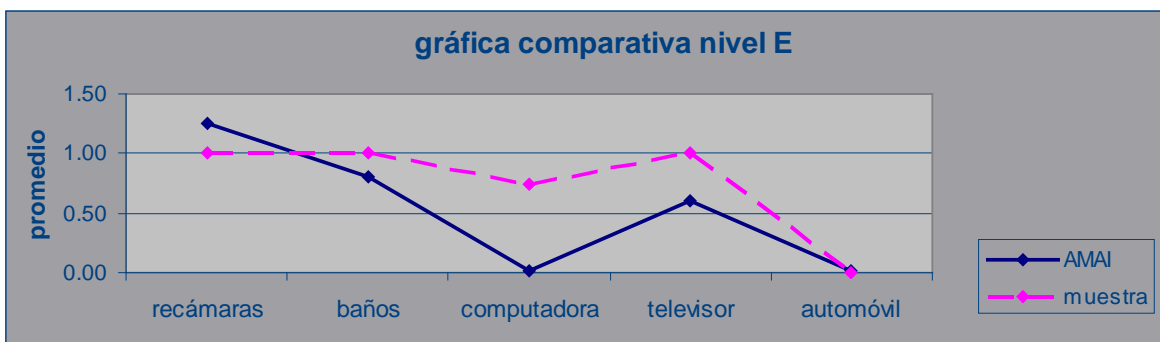
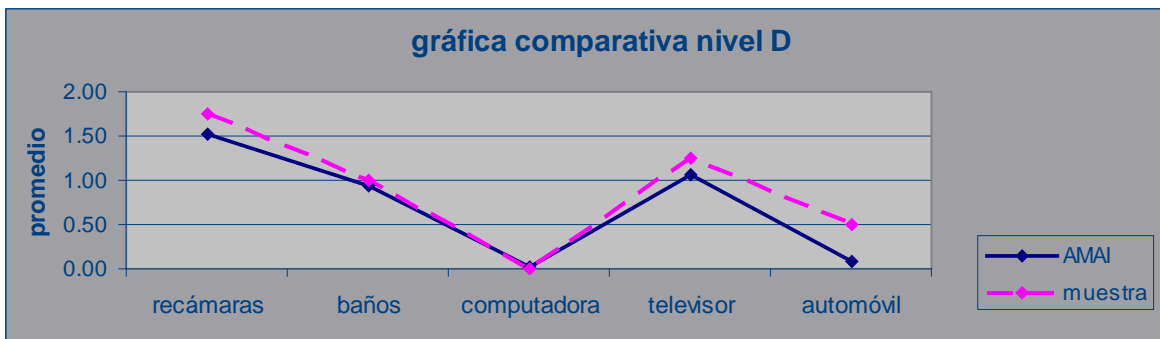
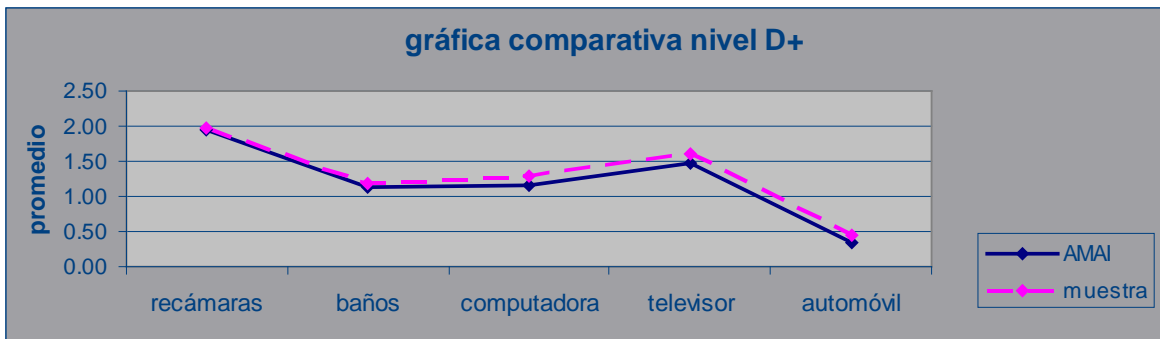
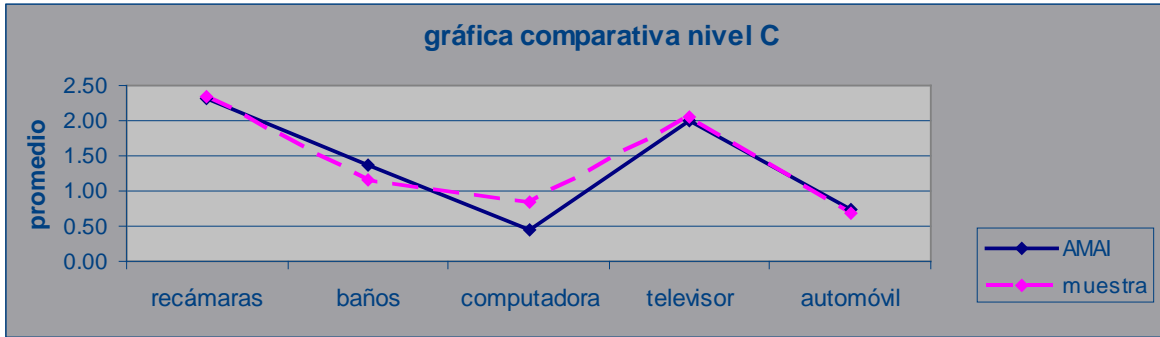
recámara	baño	computadora	televisor	automóvil	distancias euclidianas por nivel					
					A/B	C+	C	D+	D	E
4	2	1	4	4	3.083	3.598	4.267	4.974	5.667	6.113

recámara	baño	computadora	televisor	automóvil	distancias euclidianas por nivel					
					A/B	C+	C	D+	D	E
1	1	0	1	0	4.119	2.655	1.894	1.612	0.545	0.501

El primer estudiante tiene una distancia euclidiana menor con relación al nivel A/B, por lo que se clasificó en esta categoría. El segundo estudiante se puede clasificar en el nivel E porque es con el que tiene la distancia más corta.

Una vez que quedaron agrupados los estudiantes en cada nivel, se realizaron gráficas para comparar los promedios de la AMAI y los promedios de cada grupo de la muestra, las cuales son las siguientes:





Donde se aprecia que tienen características muy similares.

El promedio por área de estudio, sexo y nivel socioeconómico

En este apartado se consideró el promedio de calificaciones que los mismos alumnos expresaron en la encuesta. El interés está en conocer si existen diferencias significativas en el aprovechamiento:

- a) Entre alumnos de diferentes áreas de estudio según su elección de carrera
- b) Por sexo para determinar si es más alto en mujeres que en hombres o viceversa
- c) Por nivel socioeconómico

Los criterios para elegir otra universidad en caso de no ser aceptado por la UNAM

Los alumnos inscritos en bachillerato incorporado tienen la opción de presentar examen de admisión a la UNAM. En la pregunta 9 del cuestionario se planteó a los estudiantes que indicaran la importancia que le darían a determinados criterios, en caso de no ser aceptados y optar por otra universidad para continuar sus estudios.

Quizá la elección de otra universidad recaiga más en los padres que en los alumnos, y más aun si se trata de costos accesibles o becas; sin embargo, se desea conocer qué piensan los alumnos al respecto.

MÉTODO

Con la información recabada de la DGIRE se estructuró una base de datos clasificando los planteles por delegación, número de alumnos y colegiatura. La colegiatura más baja que se encontró fue de \$670.00 mensuales y la más alta fue de \$7,500.00.

Con la finalidad de obtener una muestra representativa se determinaron tres estratos de colegiatura. Esta clasificación se hizo bajo un criterio personal para delimitar un estrato bajo, uno intermedio y uno alto. Dentro de ellos se identificó el número de planteles y su población como se muestra en la siguiente tabla:

estrato (nivel de colegiatura)	escuelas	total alumnos
menos de \$2,000	37	3,172
\$2,000 - \$3,999	128	9,841
\$4,000 o más	37	3,029
	202	16,042

Selección de la muestra

Para determinar el tamaño de la muestra se consideró un nivel del 95% de confianza con un margen de error de 3% para estimar una proporción. De esta manera se estableció la siguiente relación:

$$Z_{\alpha/2} \sqrt{\frac{pq}{n}} = 0.03$$

Despejando n se tiene:

$$n = \frac{pqZ_{\alpha/2}^2}{0.03^2}$$

Se consideró la opción de asignar a p un valor de 0.5, que es el caso en que se obtiene un tamaño de muestra lo suficientemente grande para obtener el error máximo de estimación deseado; entonces el tamaño de la muestra queda:

$$n = \frac{(0.5)(0.5)(1.96)^2}{0.03^2}$$

$$n = 1067 \text{ alumnos}$$

Con este dato se hizo una distribución proporcional de la muestra dentro de los estratos de colegiatura. De las 202 escuelas con 16,042 alumnos inscritos, era posible entrevistar en promedio a 7 estudiantes por escuela, si de cada estrato se visita al 80% de escuelas, como se observa a continuación:

estrato (nivel de colegiatura)	escuelas	total alumnos	muestra				
			80%		proporción de alumnos	núm. de alumnos por nivel	alumnos por escuela
			escuelas	alumnos			
menos de \$2,000	37	3,172	30	2,538	0.20	213	7
\$2,000 - \$3,999	128	9,841	102	7,873	0.61	651	6
\$4,000 - \$7,500	37	3,029	30	2,423	0.19	203	7
	202	16,042	162	12,834	1.00	1,067	7

La elección del 80% de escuelas y alumnos es subjetiva, pero da una idea de cuántos alumnos entrevistar por plantel. Finalmente se decidió "redondear" la cifra a 10 alumnos por escuela para conformar la muestra.

Proceso de muestreo

Con el software estadístico SPSS (versión 10.0) se realizó una elección aleatoria de las escuelas por visitar. En la base de datos previamente diseñada se consultaron los domicilios y se acudió a las instituciones elegidas para aplicar dentro del aula, también al azar, un cuestionario a diez alumnos que cursaban el último semestre de CCH o el último año de preparatoria.

Limitaciones al tomar la muestra

En el proceso de muestreo se presentaron dos limitantes que no permitieron que éste resultara totalmente aleatorio:

1. Una vez que el plantel aceptó la encuesta, en algunos casos no fue posible elegir libremente el aula porque las autoridades educativas decidían cuál grupo estaba disponible en ese momento, por ejemplo, que no tuviera actividades deportivas o un examen, o que el profesor fuera inaccesible, etc.
2. En los casos en que alguna institución no aceptó participar, se visitó otra que se encontrara cerca, tomando como base que las escuelas cercanas entre sí podrían tener características similares como colegiatura y nivel socioeconómico de los alumnos.

ESTADÍSTICAS DESCRIPTIVAS

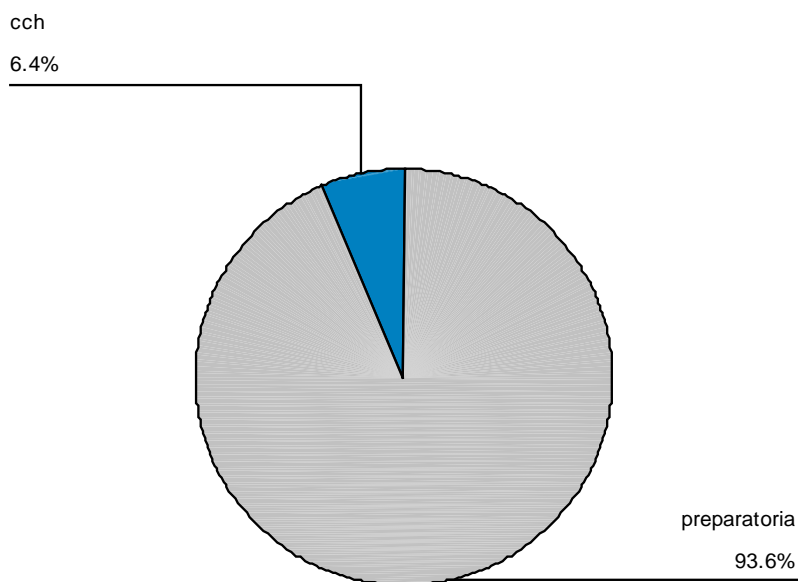
Generalidades

Se logró recabar un total de 900 cuestionarios debido a que algunas instituciones se negaron a participar. De ellos, el 93.6 % corresponde a alumnos de preparatoria (plan 1) y el resto a alumnos de CCH (plan 2), como lo indican la siguiente tabla y la gráfica:

plan

	Frecuencia	Porcentaje	porcentaje válido	porcentaje acumulado
preparatoria	842	93.6	93.6	93.6
cch	58	6.4	6.4	100.0
Total	900	100.0	100.0	

plan



La edad promedio de los alumnos oscila alrededor de los 18 años. De los 900 entrevistados, uno no respondió y la media es calculada sobre 899 alumnos:

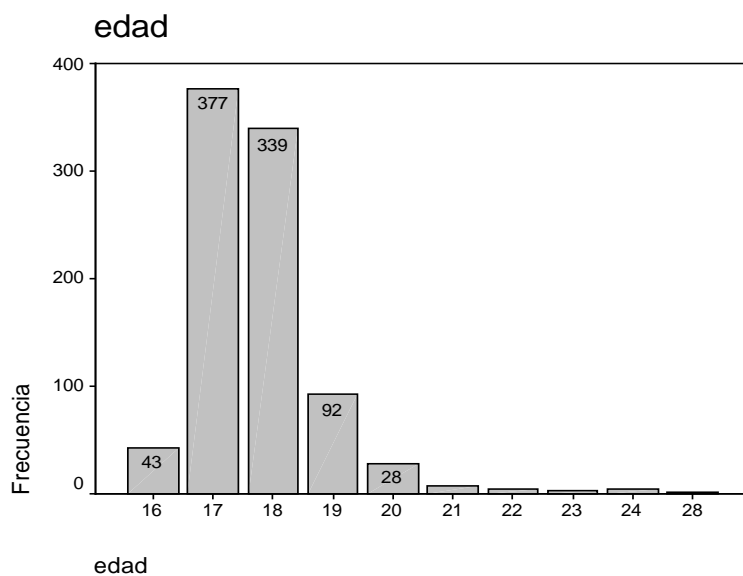
Estadísticas Descriptivas

	N	Mínimo	Máximo	media
edad	899	16	28	17.75

Sin embargo, el valor máximo que aparece es de 28 años, este dato es atípico para el nivel de bachillerato, por lo que resulta útil mostrar las frecuencias y el histograma de edades:

edad

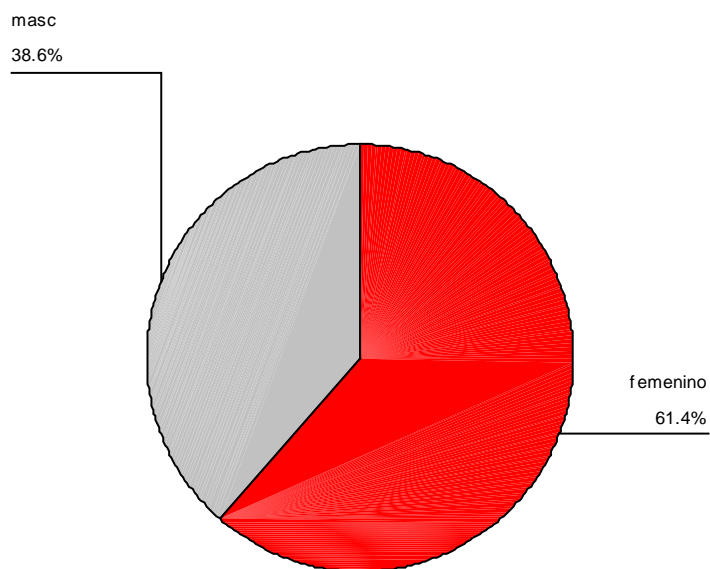
	Frecuencia	porcentaje	porcentaje válido	porcentaje acumulado
16	43	4.8	4.8	4.8
17	377	41.9	41.9	46.7
18	339	37.7	37.7	84.4
19	92	10.2	10.2	94.7
20	28	3.1	3.1	97.8
21	8	.9	.9	98.7
22	4	.4	.4	99.1
23	3	.3	.3	99.4
24	4	.4	.4	99.9
28	1	.1	.1	100.0
Total	899	99.9	100.0	
sin respuesta	1	.1		
Total	900	100.0		



La tabla de frecuencias indica que el porcentaje acumulado de los estudiantes de entre 16 y 20 años es del 97.8%; es decir, apenas el 2.2% de los encuestados es mayor de 20 años. En el histograma se aprecia que la mayoría tiene entre 17 y 18 años.

En cuanto al sexo de los alumnos se obtuvieron los siguientes resultados:

sexo				
	Frecuencia	Porcentaje	porcentaje válido	porcentaje acumulado
femenino	553	61.4	61.4	61.4
masculino	347	38.6	38.6	100.0
Total	900	100.0	100.0	

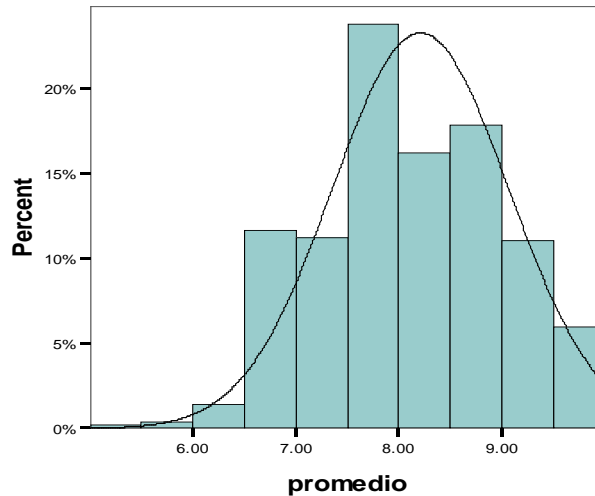


Y se observa que poco más del 60% de los entrevistados fueron mujeres.

Promedio de calificaciones

Los promedios de calificaciones global y por plan de estudio son los siguientes:

promedio global



promedio global (todos los alumnos)

	N	mínimo	máximo	media
promedio	855	5.00	10.00	8.2082

El promedio global de calificaciones en la muestra se ubica en 8.21, con una desviación estándar de 0.86. De los 900 alumnos entrevistados, 855 respondieron a esta pregunta. Con estos datos se puede estimar, mediante un intervalo de confianza del 95%, el promedio de calificación global de la población:

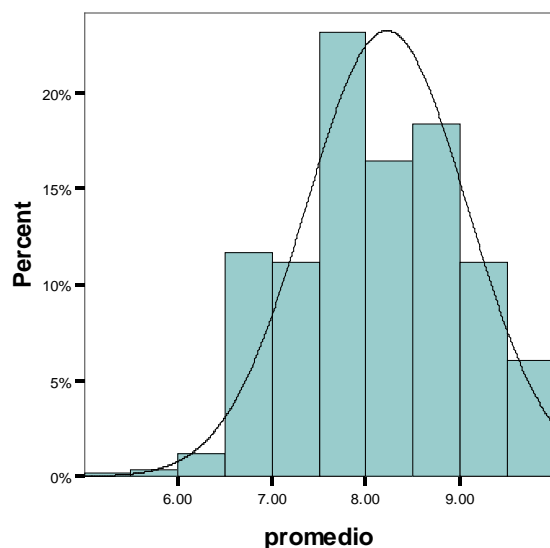
$$\bar{x} - z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

$$8.21 - (1.96) \left(\frac{0.86}{\sqrt{855}} \right) < \mu < 8.21 + (1.96) \left(\frac{0.86}{\sqrt{855}} \right)$$

$$8.14 < \mu < 8.26$$

Por lo tanto, se estima que el promedio de calificaciones de la población se ubica entre 8.14 y 8.26.

promedio preparatoria

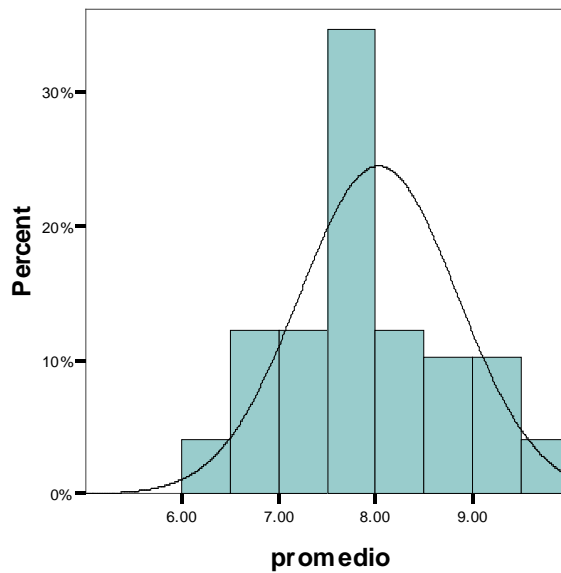


promedio de los estudiantes de preparatoria

	N	mínimo	máximo	media
promedio	806	5.00	10.00	8.2191

De la muestra, 842 estudiantes cursan la preparatoria, de ellos 806 indicaron su promedio de calificaciones. La calificación mínima manifestada fue de 5 y la máxima de 10. La media se ubica en 8.22.

promedio cch



promedio de los estudiantes de cch

	N	mínimo	máximo	media
promedio	49	6.50	9.80	8.0286

De 58 alumnos inscritos en CCH, 49 manifestaron su promedio de calificaciones, cuya media es de 8.03. Las calificaciones mínima y máxima fueron de 6.5 y 9.8 respectivamente.

Preferencia de carreras

En la pregunta 7, "¿Qué carrera vas a estudiar terminando tu bachillerato?", 889 alumnos mencionaron 61 carreras en total, 10 no saben qué carrera estudiar y uno no contestó. Los porcentajes se muestran a continuación:

	carrera	Frecuencia	porcentaje
1	medicina	90	10.0
2	derecho	79	8.8
3	ciencias de la comunicación/periodismo	76	8.4
4	administración	52	5.8
5	psicología	50	5.6
6	mercadotecnia/publicidad	44	4.9
7	relaciones internacionales	43	4.8
8	diseño gráfico	39	4.3
9	gastronomía	34	3.8
10	turismo/hotelería	25	2.8
11	actuaría	24	2.7
12	contaduría/finanzas	23	2.6
13	ingeniería mecatrónica	22	2.4
14	pedagogía	20	2.2
15	ingeniería en sistemas	19	2.1
16	arquitectura	17	1.9
17	ingeniería industrial	16	1.8
18	ciencias políticas	14	1.6
19	comercio internacional	13	1.4
20	ingeniería química	12	1.3
21	químico farmacobiólogo	12	1.3
22	ing. Mecánica (automotriz)	12	1.3
23	no sabe	10	1.1
24	veterinaria	10	1.1
25	economía	9	1.0
26	música	9	1.0
27	letras	9	1.0
28	piloto aviador	9	1.0
29	teatro/actuación	8	0.9
30	biología/oceanografía	8	0.9
31	diseño textil	7	0.8
32	filosofía	6	0.7
33	cine	6	0.7
34	informática	5	0.6
35	odontología	5	0.6
36	diseño industrial	5	0.6
37	diseño y comunicación visual	5	0.6
38	nutrición	5	0.6
39	ing. civil	4	0.4
40	ing. En telemática	4	0.4
41	artes plásticas	4	0.4
42	ing. En alimentos	4	0.4
43	historia	3	0.3
44	enfermería	3	0.3
45	ciencias genómicas	2	0.2
46	matemáticas aplicadas y computación	2	0.2
47	ing. Agrícola (agronomía)	2	0.2
48	antropología	2	0.2
49	intérprete traductor	2	0.2
50	ing. En sonido y producción musical	2	0.2
51	ing. Mecánico eléct.	2	0.2
52	sobrecargo	1	0.1
53	relaciones industriales	1	0.1
54	ing. metalúrgica	1	0.1
55	diseño de interiores	1	0.1
56	ing. en geofísica	1	0.1
57	ing. Electrónica	1	0.1
58	ing. Petrolera	1	0.1
59	arqueología	1	0.1
60	física	1	0.1
61	trabajo social	1	0.1
62	ing. Biomédica	1	0.1
63	no contestó	1	0.1
	Total	900	100.0

En la tabla se observa que medicina, derecho y ciencias de la comunicación/periodismo tienen las preferencias más altas, mientras que las carreras menos solicitadas son, entre otras, ingeniería biomédica, trabajo social, física y arqueología. Por otro lado, los estudiantes que no saben qué carrera estudiar representan el 1.1% de los encuestados.

La estimación por intervalo de la proporción total de alumnos que prefieren estudiar cierta carrera se puede realizar con un intervalo de confianza del 95%:

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Por ejemplo, suponiendo normalidad, la proporción estimada para la carrera de medicina es de:

$$.10 - (1.96) \sqrt{\frac{(.10)(.90)}{889}} < p < .10 + (1.96) \sqrt{\frac{(.10)(.90)}{889}}$$

$$.08 < p < .12$$

El mismo procedimiento se puede seguir para las demás carreras. Sin embargo, dado que las estimaciones para los valores de p son muy pequeñas, deben considerarse con reservas, pues este procedimiento no resulta confiable cuando p es muy cercano a cero o a 1.

Las estimaciones de las tres carreras más solicitadas se muestran en la siguiente tabla:

preferencias más altas	estimación de la proporción	error estándar	intervalo del 95% de confianza	
			valor inferior	valor superior
medicina	0.10	0.010062	0.08	0.12
derecho	0.088	0.009501	0.07	0.11
ciencias de la comunicación/periodismo	0.084	0.009303	0.07	0.10

Si se considera que la población de estudio consta de 16,042 alumnos, se puede estimar que entre el 8 y el 12% de ellos desean estudiar medicina; es decir, aproximadamente entre 1,284 y 1,926 estudiantes del Distrito Federal, inscritos en el SI, optan por esta carrera.

De las 553 mujeres encuestadas, más del 11% prefieren estudiar medicina. Las preferencias de las diez carreras más solicitadas por mujeres son las siguientes:

carreras más solicitadas por mujeres		porcentaje
1	medicina	11.2
2	ciencias de la comunicación/periodismo	9.0
3	derecho	8.3
4	psicología	7.4
5	relaciones internacionales	6.0
6	administración	5.4
7	mercadotecnia/publicidad	5.2
8	diseño gráfico	4.7
9	gastronomía	4.3
10	turismo/hotelería	3.6

Para el caso de los hombres, las preferencias son:

carreras más solicitadas por hombres		porcentaje
1	derecho	9.5
2	medicina	8.1
3	ciencias de la comunicación/periodismo	7.5
4	administración	6.3
5	ingeniería en sistemas	4.6
6	ingeniería mecatrónica	4.3
7	mercadotecnia/publicidad	4.3
8	diseño gráfico	3.7
9	ciencias políticas	3.2
10	contaduría/finanzas	3.2

Aunque en diferente orden para hombres y mujeres, en estas tablas se puede apreciar que las tres carreras con más demanda son medicina, derecho y ciencias de la comunicación.

Con respecto a los motivos que expresan los alumnos para elegir su carrera, se encontraron los siguientes:



Con estos datos es posible estimar, con un intervalo de confianza del 95%, las proporciones de la población de estudiantes que manifiestan el motivo por el que elige su carrera:

motivo	estimación de la proporción	error estándar	intervalo de confianza del 95%	
			valor inferior	valor superior
gusto	0.553	0.016573	0.52	0.59
interés	0.283	0.015015	0.25	0.31
ayudar a la gente	0.064	0.008158	0.05	0.08
aptitudes/facilidad	0.059	0.007854	0.04	0.07
amplio campo de trabajo/tiene futuro	0.021	0.004779	0.01	0.03
no contestó	0.017	0.004309	0.01	0.03

De esta información destaca que entre el 52 y el 59% de los encuestados decide estudiar una carrera por gusto hacia ella.

En la tabla anterior se descartan dos motivos debido a que su proporción es muy pequeña y el procedimiento para estimar su intervalo no es confiable, pues no cumplen con el requisito de que

$n\hat{p}$ y $n\hat{q}$ sean ambos por lo menos igual a 5, como se aprecia enseguida:

motivo	proporción de la muestra	n	$n\hat{p}$	$n\hat{q}$
poca demanda	0.001	900	0.9	899.1
trabajar con mi papá	0.001	900	0.9	899.1

DEMANDA DE CARRERAS POR SEXO Y POR NIVEL SOCIOECONÓMICO

Demanda de carreras por sexo

Continuando con el análisis, primero se realizó una prueba de χ^2 con el software JMP (versión 7.0) para identificar si el sexo de los alumnos y el área de la carrera que desean estudiar son independientes entre sí. Después se realizó un análisis de correspondencia para ver la asociación entre estas dos variables. Dado que se desea conocer cómo es la demanda de carreras, del total de 900 alumnos de la muestra se excluyeron 11 estudiantes (10 que no saben qué carrera estudiar y uno que no respondió la pregunta), por lo que los cálculos se realizaron sobre 889 estudiantes.

a) Tabla de contingencia para la prueba de χ^2 :

Contingency Analysis of areacarr By sexo

Count Expected	femenino	masculino	
fisico mate	72 92.1822	77 56.8178	149
biológicas y salud	133 116.929	56 72.0709	189
sociales	216 222.722	144 137.278	360
human y artes	84 75.4781	38 46.5219	122
otras	45 42.6884	24 26.3116	69
	550	339	889

Test	ChiSquare	Prob>ChiSq
Pearson	20.764	0.0004

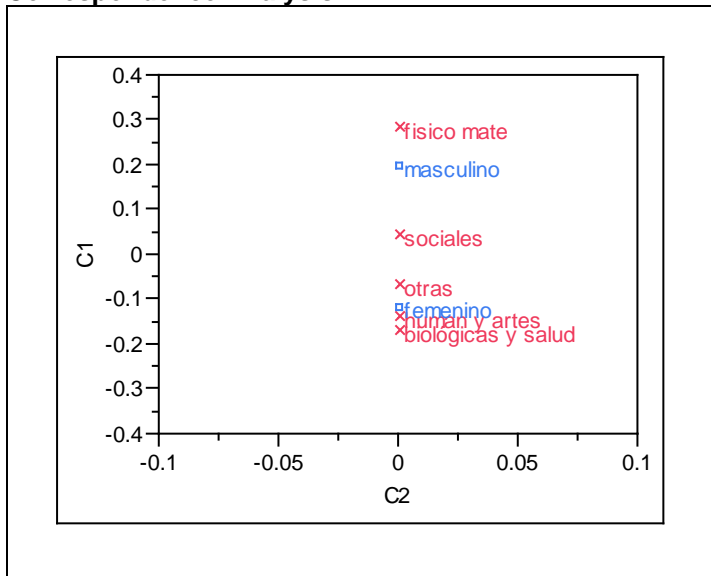
Si $\chi^2 > \chi^2_{\alpha}$ se rechaza la hipótesis nula sobre independencia. En este caso, se realizó la prueba con un nivel de significancia de 0.05 cuyo valor crítico es $\chi^2_{0.05} = 9.488$ con 4 grados de libertad. En la segunda tabla se observa que el valor de χ^2 de Pearson es 20.764 y, dado que $20.764 > 9.488$, entonces se rechaza la hipótesis nula por lo que el sexo y el área de la carrera que desean estudiar los alumnos no son independientes. También se puede rechazar H_0 si el valor

$p < \alpha$. De acuerdo con la segunda tabla, $p = 0.0004$, y como este valor es menor al nivel de significancia de 0.05 se rechaza la hipótesis nula.

b) Análisis de correspondencia

En el análisis de correspondencia (ver anexo) las frecuencias observadas de la tabla de contingencia se transforman en valores estandarizados, tomando en cuenta su peso dentro de cada renglón y cada columna, es decir, sus frecuencias relativas. La transformación se hace de tal forma que los datos puedan ser representados en una sola gráfica que refleje la relación entre las dos categorías de la tabla de contingencia. La gráfica obtenida es la siguiente:

Correspondence Analysis



x areacarr
□ sexo

Singular Value	Inertia	Portion	Cumulative
0.15283	0.02336	1.0000	1.0000

areacarr	c1
fisico mate	0.2789
biológicas y salud	-0.1751
sociales	0.0384
human y artes	-0.1438
Otras	-0.0690

sexo	c1
femenino	-0.1200
masculino	0.1947

La variable *sexo* se muestra con una "X" y la variable *área-carrera* con un "cuadro" y es posible apreciar que las carreras del área de físico matemáticas están más asociadas a los hombres. Por otra parte, las carreras del área de humanidades y artes y del área de biológicas y la salud aparecen más asociadas a las mujeres.

Además de la gráfica, JMP muestra la siguiente información:

Singular Value

En este caso es igual a 0.15283 y representa el valor singular de $D_r^{-1/2}(P - rc')D_c^{-1/2}$ donde:

P es la matriz de frecuencias relativas, llamada matriz de correspondencia.

r es el vector ($n \times 1$) que contiene los valores de la suma de cada renglón de P

c es el vector ($m \times 1$) con los valores de la suma de cada columna de P

El número de valores singulares está asociado con el rango de la matriz de la tabla de contingencia, el cual es el mínimo entre $(a - 1, b - 1)$ donde a es el número de renglones y b el número de columnas. Dado que la variable *sexo* tiene dos categorías entonces el rango es igual a 1, por lo tanto, sólo existe un valor singular.

Estas matrices y el procedimiento del análisis de correspondencia se explican en el anexo.

Inertia

Es el cuadrado del valor o los valores singulares y representa la variación en los datos.

Portion

Se refiere a la porción que representa la inercia con respecto a la inercia total. En este caso los datos sólo tienen un valor singular y por ende un solo valor de inercia, por lo que la porción que representa es de 1. Esto también indica que las variables pueden ser graficadas en un solo eje, llamado C1. Los valores de C1 para cada categoría de *área-carrera* y *sexo* son sus respectivas coordenadas.

Demanda de carreras por nivel socioeconómico

Una vez que los estudiantes se clasificaron dentro de los niveles socioeconómicos de la AMAI, donde A/B es el más alto y E el más bajo, se realizó una prueba de χ^2 con el software JMP, cuya tabla de contingencia es la siguiente:

Contingency Analysis of areacarr By nse AMAI

Count Expected	fisico mate	biológicas y salud	sociales	human y artes	otras	totales
A/B	76 78.2711	96 99.2835	202 189.111	54 64.0877	39 36.2463	467
C+	51 45.0855	52 57.189	103 108.931	42 36.9156	21 20.8785	269
C	13 13.0731	19 16.5827	30 31.5861	11 10.7042	5 6.05399	78
D+	8 11.3971	20 14.4567	22 27.5366	15 9.33183	3 5.27784	68
D	0 0.50281	1 0.6378	1 1.21485	0 0.4117	1 0.23285	3
E	1 0.67042	1 0.85039	2 1.6198	0 0.54893	0 0.31046	4
totales	149	189	360	122	69	889

Test	ChiSquare	Prob>ChiSq
Pearson	19.245	0.5060

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

De acuerdo con la leyenda que arroja JMP, el 20% de las celdas tienen una frecuencia esperada menor a 5. De acuerdo con Weimer (1993), al menos el 80% de las celdas de la tabla de contingencia deben tener frecuencias esperadas mayores o iguales a 5. En la tabla se aprecia que los niveles D y E no cumplen esta condición. En realidad son diez celdas que representan el 33% del total ($10 \div 30 = 0.33$). Esto se puede solucionar agrupando las observaciones de los dos niveles en uno solo pero antes es necesario comparar sus características. Los niveles socioeconómicos de la AMAI tienen los siguientes promedios por hogar:

nse	promedio AMAI				
	recámaras	baños	computadora	televisor	automóvil
D	1.53	0.94	0.03	1.07	0.08
E	1.25	0.81	0.01	0.61	0.01

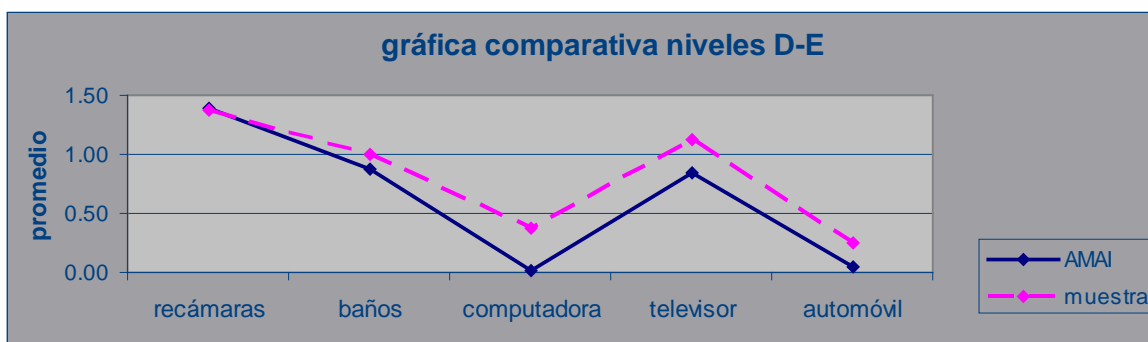
Si se calcula el promedio entre los dos niveles y al resultado se le llama D-E, queda:

nse	promedio AMAI				
	recámaras	baños	computadora	televisor	automóvil
D-E	1.39	0.88	0.02	0.84	0.05

Para determinar si la fusión sería adecuada, los alumnos de la muestra del nivel E se agruparon con los del nivel D y se calcularon los promedios:

nse	promedio (muestra)				
	recámaras	baños	computadora	televisor	automóvil
D-E	1.38	1.00	0.38	1.13	0.25

La gráfica de estas dos últimas tablas es la siguiente:



Donde se aprecia que las características son muy similares.

Con esta nueva clasificación, llamada ahora AMAI 2, se obtiene la siguiente prueba χ^2 :

Contingency Analysis of areacarr By nse AMAI 2

Count Expected	fisico mate	biológicas y salud	sociales	human y artes	otras	totales
A/B	76 78.2711	96 99.2835	202 189.111	54 64.0877	39 36.2463	467
C+	51 45.0855	52 57.189	103 108.931	42 36.9156	21 20.8785	269
C	13 13.0731	19 16.5827	30 31.5861	11 10.7042	5 6.05399	78
D+	8 11.3971	20 14.4567	22 27.5366	15 9.33183	3 5.27784	68
D-E	1 1.17323	2 1.48819	3 2.83465	0 0.96063	1 0.54331	7
totales	149	189	360	122	69	889

Test	ChiSquare	Prob>ChiSq
Pearson	15.978	0.4545

Aquí se observa que con la agrupación D-E apenas se cumple la condición de que el 20% de las celdas ($5 \div 25 = .2$) tienen frecuencias esperadas menores a 5.

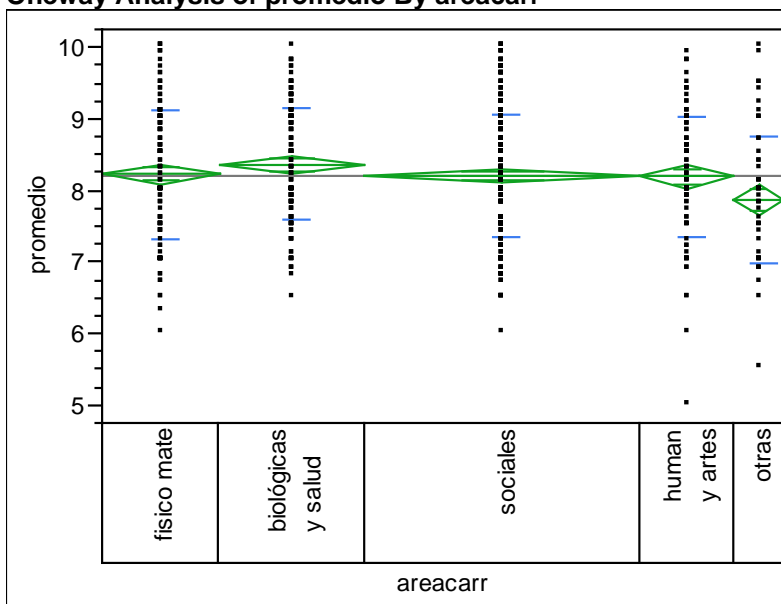
El valor de $p = 0.4545$ es mayor que el nivel de significancia de 0.05, por lo tanto no se rechaza la hipótesis nula; es decir, el área de la carrera que eligen los estudiantes de la muestra es independiente de su nivel socioeconómico.

PROMEDIO DE CALIFICACIONES POR ÁREA DE ESTUDIO, SEXO Y NIVEL SOCIOECONÓMICO

Promedio por área de estudio

Para comparar el promedio de calificaciones de los alumnos se realizó un análisis de varianza, así como pruebas de homogeneidad de varianzas y una comparación de medias por el procedimiento de Tukey-Kramer:

Oneway Analysis of promedio By areacarr



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Prob > F
areacarr	4	10.78241	2.69560	0.0052
Error	840	608.70559	0.72465	
C. Total	844	619.48800		

Tests that the Variances are Equal

Level	Count	Std Dev
fisico mate	146	0.9053718
biológicas y salud	180	0.7780580
sociales	341	0.8565351
human y artes	117	0.8542149
otras	61	0.8888410

Test	F Ratio	DF	Prob > F
Bartlett	1.0306	4	0.3897

Comparisons for all pairs using Tukey-Kramer HSD

Level			Mean
biológicas y salud	A		8.3566667
físico mate	A	B	8.2253425
sociales	A	B	8.1909091
human y artes	A	B	8.1880342
otras		B	7.8786885

Levels not connected by same letter are significantly different.

Con el análisis de varianza se prueba la hipótesis nula de que las medias de calificaciones por área-carrera son iguales. En la gráfica los rombos representan un intervalo de confianza del 95% para cada media y se observa que la media más baja corresponde al área de Otras (gastronomía, turismo/hotelería, sobrecargo y piloto aviador).

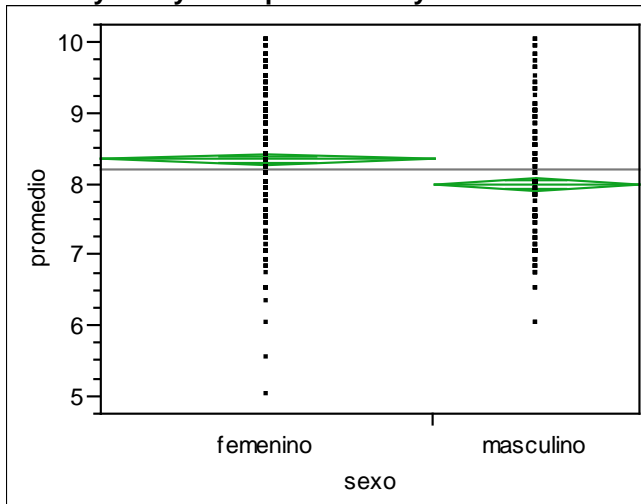
La tabla de análisis de varianza indica que H_0 se rechaza con un valor p de 0.0052; por lo tanto, al menos dos de las medias son significativamente diferentes. Sin embargo, cuando no existe igualdad entre varianzas, la prueba de anova no es válida. Para saber si las varianzas son iguales se puede realizar una prueba de homogeneidad de varianzas. En este caso se realizó la prueba de Bartlett, cuyo estadístico tiene una distribución χ^2 bajo el supuesto de normalidad en las poblaciones. Los grados de libertad son cuatro ($k - 1$, donde k es el número de niveles o categorías). En la segunda tabla se muestra el número de alumnos por área así como la desviación estándar de sus calificaciones con respecto al promedio del área. El valor del estadístico de Bartlett es el resultado de multiplicar los grados de libertad por el valor F que arroja JMP, es decir: $4 * 1.0306 = 4.1224$. Este valor se compara con el valor crítico correspondiente a 4 grados de libertad y un nivel de significancia del 5% que es igual a 9.488. Dado que el estadístico de Bartlett es menor que el valor crítico, se acepta la hipótesis nula sobre igualdad de varianzas. De hecho, el valor $p = 0.3897$ es mayor que el nivel de significancia, por lo que es evidencia suficiente para aceptar la hipótesis nula. Una vez que se identifica la homogeneidad de varianzas, es posible conocer cuáles medias son significativamente diferentes utilizando la prueba de comparación de medias de Tukey, en cuya tabla se asignan las letras A y B a todas las categorías que tienen medias iguales. Si en dos o más categorías sólo aparece la letra A o la B, entonces las medias de esas categorías son

significativamente diferentes. En este caso son el área de Biológicas y de la Salud con la media más alta, y el área de Otras con la media más baja.

Promedio por sexo

El análisis de varianza correspondiente es:

Oneway Analysis of promedio By sexo



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Prob > F
sexo	1	27.23377	27.2338	<.0001
Error	843	592.25423	0.7026	
C. Total	844	619.48800		

Tests that the Variances are Equal

Level	Count	Std Dev
femenino	523	0.8372180
masculino	322	0.8397573

Test	F Ratio	DF	p-Value
Bartlett	0.0036	1	0.9519

Comparisons for all pairs using Tukey-Kramer HSD

Level		Mean
femenino	A	8.3500956
masculino	B	7.9804348

Levels not connected by same letter are significantly different.

En la gráfica se observa que el promedio de calificaciones de las mujeres es más alto que el de los hombres. La hipótesis nula sobre igualdad de medias se rechaza con un valor p menor a 0.0001 y

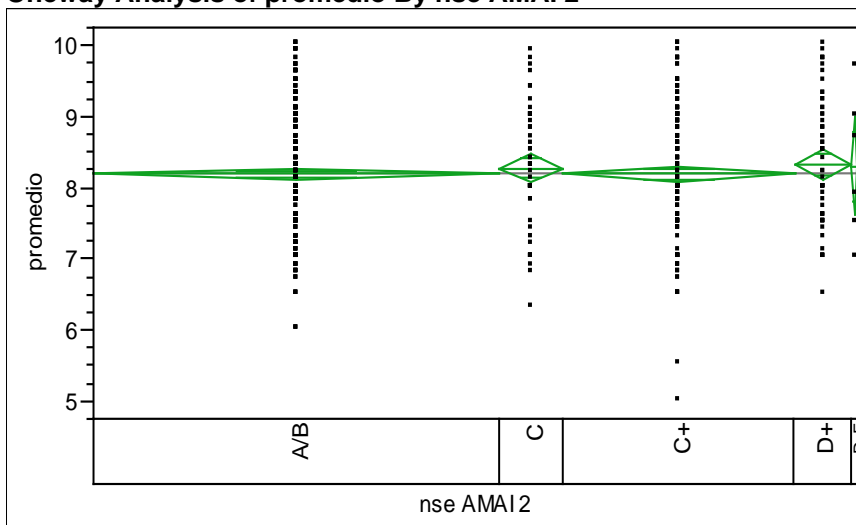
en la prueba de Bartlett se acepta la hipótesis nula sobre homogeneidad de varianzas con un valor p igual a 0.9519, por lo que la prueba de anova resulta confiable.

En la tabla de Tukey se puede ver que las medias de las calificaciones por sexo no están asociadas con las mismas letras (A y B), lo cual indica que son significativamente diferentes: la media más alta corresponde a los estudiantes del sexo femenino.

Promedio por nivel socioeconómico

El análisis de varianza y la comparación de medias de Tukey son:

Oneway Analysis of promedio By nse AMAI 2



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Prob > F
nse AMAI 2	4	1.35708	0.339269	0.7644
Error	840	618.13092	0.735870	
C. Total	844	619.48800		

Tests that the Variances are Equal

Level	Count	Std Dev
A/B	449	0.846453
C	71	0.817473
C+	256	0.876261
D+	63	0.893261
D-E	6	1.009950

Test	F Ratio	DF	Prob > F
Bartlett	0.2976	4	0.8797

Comparisons for all pairs using Tukey-Kramer HSD

Level		Mean
D+	A	8.3174603
D-E	A	8.3000000
C	A	8.2788732
C+	A	8.1968750
A/B	A	8.1888641

Levels not connected by same letter are significantly different.

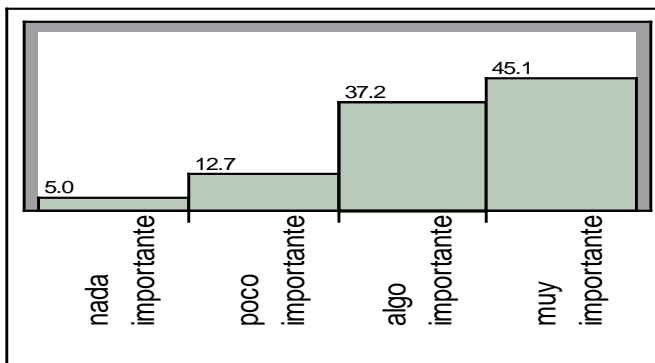
Gráficamente se observan los promedios de calificaciones muy similares entre nivel socioeconómico.

La hipótesis nula sobre igualdad de medias no se rechaza con un valor p de 0.7644. La prueba de Bartlett arroja un valor de p igual a 0.8797, por lo que las varianzas no tienen diferencias significativas. Por otra parte, la comparación de medias de Tukey indica que todos los promedios de calificaciones están asociados por la letra A; es decir, las medias de calificaciones por nivel socioeconómico no son significativamente diferentes.

CRITERIOS PARA ELEGIR OTRA UNIVERSIDAD EN CASO DE NO SER ACEPTADO POR LA UNAM

En este apartado se identificaron las frecuencias absolutas de las respuestas que dieron los estudiantes a cada característica en la pregunta 9 del cuestionario. Los alumnos que no respondieron sobre alguna característica se omitieron en el análisis. A continuación se muestran los histogramas sobre las características que resultaron más importantes para los alumnos, su prueba de independencia con otras variables y, en su caso, su análisis de correspondencia:

Costos accesibles



Para el 45% de los encuestados son muy importantes los costos accesibles al elegir otra universidad. Esta característica podría tener alguna relación con el nivel socioeconómico, lo cual se puede determinar a partir de una prueba de independencia:

Contingency Analysis of costos By nse AMAI 2

Count Expected	nada importante	poco importante	algo importante	muy importante	totales
A/B	30 23.1269	83 59.9198	186 176.08	173 212.873	472
C+	10 13.2784	23 34.4031	99 101.097	139 122.222	271
C	3 3.87082	3 10.029	26 29.471	47 35.6292	79
D+	1 3.33185	5 8.63252	22 25.3675	40 30.6682	68
D-E	0 0.39198	0 1.01559	2 2.98441	6 3.60802	8
totales	44	114	335	405	898

Test	ChiSquare	Prob>ChiSq
Pearson	44.821	<.0001

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

Debido a que 6 de las 20 celdas de la tabla, es decir el 30%, tienen una frecuencia esperada menor a 5, la prueba no resulta confiable. El nivel D-E, que es el más bajo en la escala socioeconómica, cuenta con sólo 8 observaciones, si se excluye este nivel, la prueba queda:

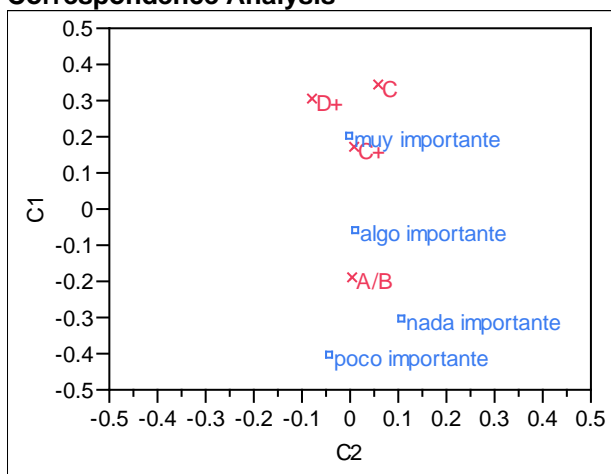
Contingency Analysis of costos By nse AMAI 2

Count Expected	nada importante	poco importante	algo importante	muy importante	totales
A/B	30 23.3348	83 60.4584	186 176.602	173 211.604	472
C+	10 13.3978	23 34.7124	99 101.397	139 121.493	271
C	3 3.90562	3 10.1191	26 29.5584	47 35.4169	79
D+	1 3.3618	5 8.71011	22 25.4427	40 30.4854	68
totales	44	114	333	399	890

Test	ChiSquare	Prob>ChiSq
Pearson	41.354	<.0001

Y se rechaza la hipótesis nula de independencia con un valor p menor a 0.0001. La asociación de las dos variables se puede identificar con la gráfica del análisis de correspondencia:

Correspondence Analysis



nse AMAI 2



costos

Details

Singular Value	Inertia	Portion	Cumulative
0.21314	0.04543	0.9777	0.9777
0.02864	0.00082	0.0177	0.9954
0.01467	0.00022	0.0046	1.0000

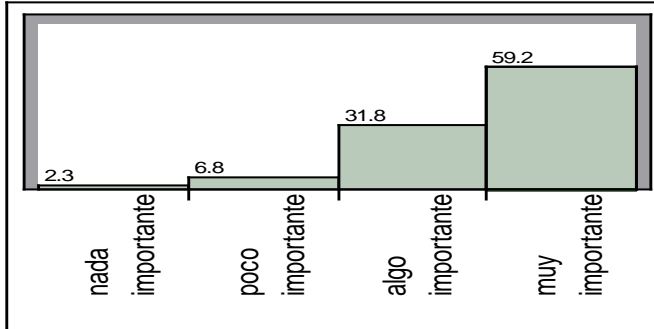
nse AMAI 2	c1	c2	c3
A/B	-0.1944	-0.0002	0.0034
C	0.3400	0.0543	0.0298
C+	0.1640	0.0058	-0.0189
D+	0.3012	-0.0847	0.0170

costos	c1	c2	c3
nada importante	-0.3063	0.1031	0.0301
poco importante	-0.4050	-0.0443	0.0131
algo importante	-0.0629	0.0086	-0.0179
muy importante	0.2020	-0.0059	0.0079

Donde se aprecia que para los estudiantes del nivel A/B, el más alto en la escala socioeconómica, los costos accesibles son nada o poco importantes.

La inercia total se obtiene con tres valores singulares pero la inercia asociada a los dos primeros abarcan una porción acumulada del 99.5%, lo que significa que la gráfica es representativa en dos ejes: C1 y C2. Los valores de C1 y C2 se muestran en las dos últimas tablas y corresponden a las coordenadas de cada categoría del nivel socioeconómico y de los costos.

Prestigio



Esta característica es muy importante para el 59% de los encuestados. La prueba de independencia entre el prestigio y el área de la carrera es la siguiente:

Contingency Analysis of prestiigi By areacarr

Count Expected	nada importante	poco importante	algo importante	muy importante	
físico mate	5 3.35964	7 10.0789	49 47.3709	88 88.1905	149
biológicas y salud	4 4.23901	16 12.717	61 59.77	107 111.274	188
sociales	2 8.0947	19 24.2841	101 114.135	237 212.486	359
human y artes	5 2.75085	15 8.25254	47 38.7869	55 72.2097	122
otras	4 1.55581	3 4.66742	24 21.9369	38 40.8399	69
	20	60	282	525	887

Test	ChiSquare	Prob>ChiSq
Pearson	30.950	0.0020

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

La prueba no es confiable debido al número de casillas que tienen frecuencias esperadas menores a 5. Sin embargo, el porcentaje de los estudiantes que piensa que el prestigio no es importante representa únicamente el 2.3%. Si se omite esta categoría, la prueba queda:

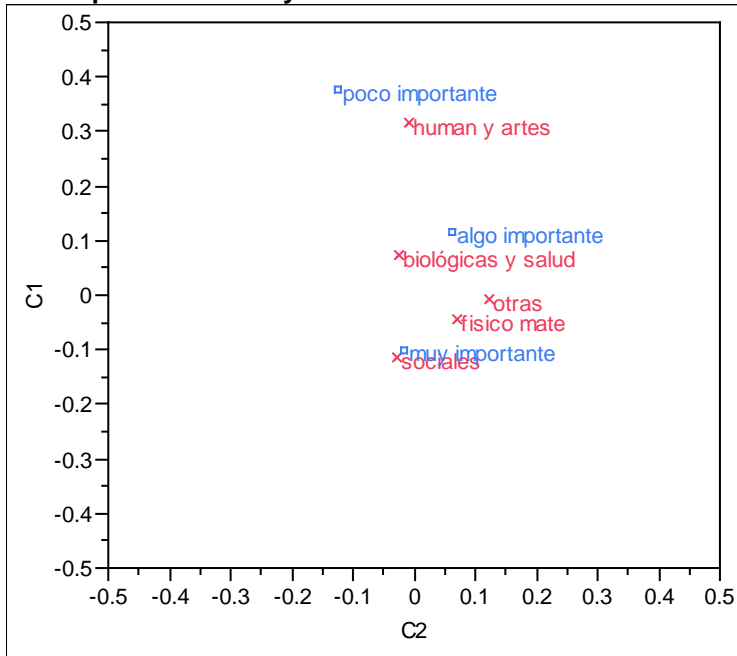
Contingency Analysis of prestiigi By areacarr

Count Expected	poco importante	algo importante	muy importante	totales
físico mate	7 9.9654	49 46.8374	88 87.1972	144
biológicas y salud	16 12.7336	61 59.8478	107 111.419	184
sociales	19 24.7059	101 116.118	237 216.176	357
human y artes	15 8.09689	47 38.0554	55 70.8478	117
otras	3 4.49827	24 21.1419	38 39.3599	65
totales	60	282	525	867

Test	ChiSquare	Prob>ChiSq
Pearson	19.782	0.0112

Con un valor p de 0.0112 se rechaza la hipótesis nula, es decir, el prestigio y el área de la carrera no son independientes. La gráfica de análisis de correspondencia es:

Correspondence Analysis



areacarr



prestigi

Details

Singular Value	Inertia	Portion	Cumulative
0.14274	0.02038	0.8930	0.8930
0.04941	0.00244	0.1070	1.0000

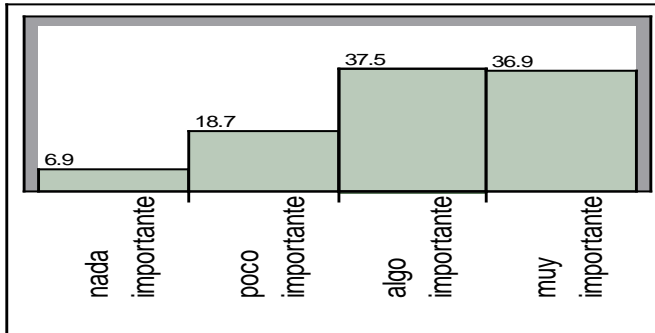
areacarr	c1	c2
fisico mate	-0.0464	0.0687
biológicas y salud	0.0691	-0.0293
sociales	-0.1178	-0.0309
human y artes	0.3138	-0.0105
otras	-0.0108	0.1193

prestigi	c1	c2
poco importante	0.3757	-0.1262
alguno importante	0.1127	0.0595
muy importante	-0.1035	-0.0175

Es posible observar que para los estudiantes del área de sociales es muy importante el prestigio, lo que no sucede con los estudiantes de humanidades y artes porque para ellos resulta poco importante.

La inercia total se representa adecuadamente en dos ejes, con una porción acumulada del 100%.

Que otorgue becas



El 37.5% de los encuestados considera algo importante el otorgamiento de becas y el 36.9% muy importante. Al realizar una prueba de independencia entre el nivel socioeconómico y las becas se obtiene:

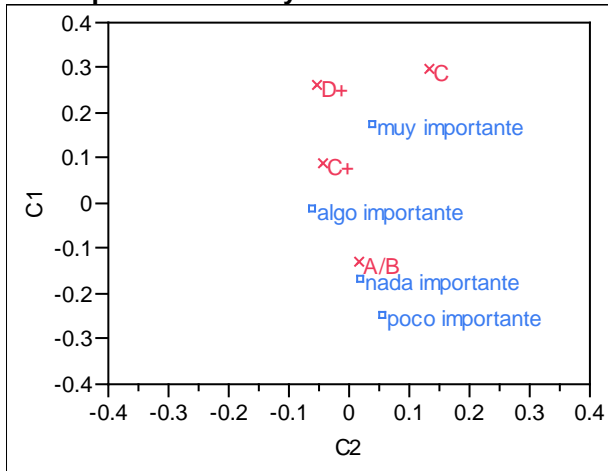
Contingency Analysis of becas By nse AMAI 2

Count	nada importante	poco importante	algo importante	muy importante	
Expected					
A/B	37 32.387	109 88.6659	177 177.863	149 173.084	472
C+	17 18.5264	41 50.7199	107 101.744	105 99.0101	270
C	4 5.4207	10 14.8403	24 29.7694	41 28.9696	79
D+	3 4.66592	7 12.7739	27 25.6243	31 24.9359	68
	61	167	335	326	889

Test	ChiSquare	Prob>ChiSq
Pearson	24.117	0.0041

Con un valor p de 0.0041 se rechaza la hipótesis de independencia. Del análisis de correspondencia se obtiene la siguiente gráfica:

Correspondence Analysis



nse AMAI 2



becas

Details

Singular Value	Inertia	Portion	Cumulative
0.15648	0.02449	0.9026	0.9026
0.05099	0.00260	0.0958	0.9985
0.00647	0.00004	0.0015	1.0000

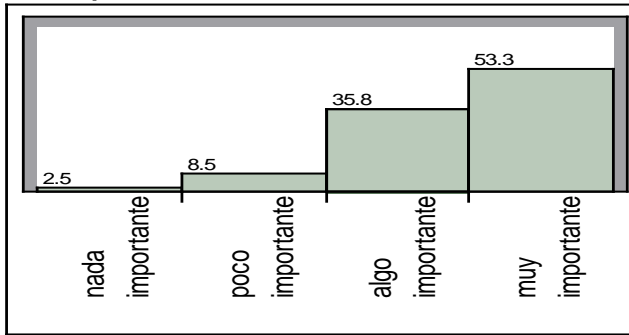
nse AMAI 2	c1	c2	c3
A/B	-0.1348	0.0140	-0.0017
C	0.2913	0.1313	0.0025
C+	0.0857	-0.0481	0.0068
D+	0.2572	-0.0583	-0.0184

becas	c1	c2	c3
nada importante	-0.1671	0.0156	0.0227
poco importante	-0.2476	0.0531	-0.0055
algo importante	-0.0145	-0.0646	-0.0013
muy importante	0.1730	0.0362	-0.0001

Y se aprecia que para los estudiantes del nivel socioeconómico A/B resulta nada o poco importante que se otorguen becas.

Los dos primeros valores singulares están asociados con el 99.85% de la inercia total, por lo que la gráfica es representativa con dos ejes.

Que imparta idiomas



Poco más de la mitad de los alumnos de la muestra ve muy importante, para elegir una universidad, que ésta imparta idiomas. La prueba de independencia entre esta variable y el área de la carrera elegida proporciona el siguiente resultado:

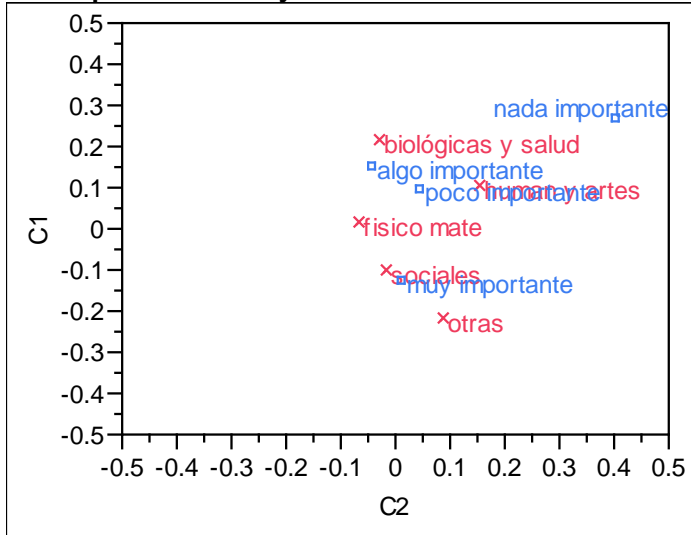
Contingency Analysis of idiomas By areacarr

Count	nada importante	poco importante	algo importante	muy importante	
Expected					
fisico mate	2	14	55	77	148
	3.67494	12.5282	52.9526	78.8442	
biológicas y salud	6	16	85	81	188
	4.66817	15.9142	67.2641	100.153	
sociales	6	28	116	210	360
	8.93905	30.474	128.804	191.783	
human y artes	6	13	43	59	121
	3.00451	10.2427	43.2923	64.4605	
otras	2	4	18	45	69
	1.71332	5.84086	24.6874	36.7585	
	22	75	317	472	886

Test	ChiSquare	Prob>ChiSq
Pearson	22.429	0.0330

Con un valor p de 0.033 se rechaza H_0 . La asociación entre las dos variables se puede apreciar con el análisis de correspondencia:

Correspondence analysis



areacarr



idiomas

Details

Singular Value	Inertia	Portion	Cumulative
0.13863	0.01922	0.7591	0.7591
0.07006	0.00491	0.1939	0.9530
0.03449	0.00119	0.0470	1.0000

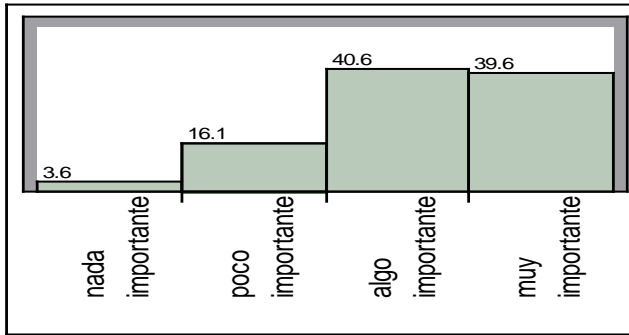
areacarr	c1	c2	c3
físico mate	0.0115	-0.0692	0.0472
biológicas y salud	0.2095	-0.0323	-0.0380
sociales	-0.1054	-0.0218	0.0012
human y artes	0.1022	0.1515	0.0354
otras	-0.2250	0.0846	-0.0661

idiomas	c1	c2	c3
nada importante	0.2659	0.3992	-0.0610
poco importante	0.0954	0.0402	0.1091
algo importante	0.1494	-0.0470	-0.0148
muy importante	-0.1279	0.0066	-0.0046

Los estudiantes del área de sociales son los que dan mayor importancia a los idiomas, en segundo lugar están los del área de Otras (gastronomía, turismo/hotelería, sobrecargo y piloto aviador).

En este caso, también los ejes C1 y C2 son suficientes para representar adecuadamente la asociación entre las dos variables.

Que cuente con actividades culturales



Esta variable está relacionada con el área de la carrera que desea estudiar el alumno, como se puede observar enseguida:

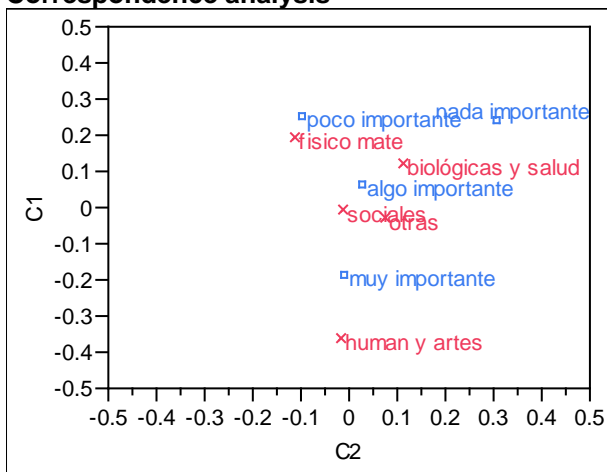
Contingency Analysis of cultural By areacarr

Count Expected	nada importante	poco importante	algo importante	muy importante	
físico mate	5 5.38149	36 24.0485	59 60.5418	49 59.0282	149
biológicas y salud	11 6.75395	31 30.1817	82 75.9819	63 74.0824	187
sociales	10 12.9661	55 57.9424	154 145.869	140 142.222	359
human y artes	2 4.40632	10 19.6907	41 49.5711	69 48.3318	122
otras	4 2.4921	11 11.1366	24 28.0361	30 27.3352	69
	32	143	360	351	886

Test	ChiSquare	Prob>ChiSq
Pearson	32.010	0.0014

Las dos variables no son independientes con un valor p de 0.0014.

Correspondence analysis





areacarr



cultural

Details

Singular Value	Inertia	Portion	Cumulative
0.16564	0.02744	0.7594	0.7594
0.07252	0.00526	0.1456	0.9050
0.05858	0.00343	0.0950	1.0000

areacarr	c1	c2	c3
físico mate	0.1893	-0.1148	0.0523
biológicas y salud	0.1181	0.1089	-0.0008
sociales	-0.0092	-0.0152	-0.0579
human y artes	-0.3649	-0.0215	0.0293
otras	-0.0359	0.0701	0.1387

cultural	c1	c2	c3
nada importante	0.2415	0.3054	0.1531
poco importante	0.2502	-0.1002	0.0587
algo importante	0.0607	0.0234	-0.0648
muy importante	-0.1862	-0.0110	0.0286

Las actividades culturales son poco importantes para los estudiantes del área físico matemáticas y son muy importantes para los del área de humanidades y artes.

La gráfica es adecuada con dos ejes cuyos valores singulares están asociados al 90.5% de la inercia total.

CONCLUSIONES

Considerando las limitaciones expuestas sobre el muestreo realizado, se obtuvieron las siguientes conclusiones:

En cuanto a la demanda de carreras:

- Las carreras del área físico matemáticas están más asociadas a los hombres.
- Por su parte, las mujeres son más propensas a cursar una carrera del área de humanidades y artes y del área de biológicas y de la salud.
- La elección de carrera que desea cursar el alumno es independiente de su nivel socioeconómico.

Con respecto a los promedios:

- Los estudiantes que desean cursar las carreras de turismo y hotelería, piloto aviador, gastronomía y sobrecargo son los que tienen los promedios más bajos.
- Los promedios más altos los tienen los alumnos que quieren estudiar una carrera del área de biológicas y de la salud.
- Los promedios por sexo son significativamente diferentes, las mujeres tienen los más altos, cuya media es de 8.35, por arriba de la media de los hombres que es de 7.98.
- Los promedios por nivel socioeconómico no son significativamente diferentes.

En cuanto a los criterios para seleccionar una universidad en caso de no ser aceptado por la UNAM se encontró que:

- Para los estudiantes del nivel socioeconómico más alto no es importante que al elegir una universidad ésta cuente con costos accesibles. Resultó ser más importante para los alumnos de los otros niveles.
- El prestigio es una característica muy importante para el 59% de los estudiantes encuestados. Los que le dan mayor importancia son aquellos que desean cursar una carrera del área de sociales y los que le dan poca importancia son aquellos que desean realizar sus estudios en el área de humanidades y artes.
- El otorgamiento de becas es más importante para los alumnos de los niveles socioeconómicos C+, C y D+. Para los de nivel socioeconómico A/B (el más alto) no es importante.
- Que la universidad imparta idiomas es muy importante para el 53% de los alumnos. Los estudiantes que le dan mayor importancia son los del área de sociales. Por su parte, los estudiantes del área de humanidades y artes consideran esta característica poco importante.
- Al elegir una universidad, las actividades culturales resultaron muy importantes para quienes desean estudiar una carrera de humanidades y artes. Los que le dan menos importancia son los estudiantes orientados hacia el área de físico matemáticas.

ANEXO

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA**

Cuestionario sobre demanda de carreras universitarias en instituciones incorporadas a la UNAM.

El presente cuestionario tiene como finalidad, única y exclusivamente, conocer la preferencia de los alumnos que cursan el bachillerato incorporado a la UNAM sobre las carreras universitarias que se ofrecen en el país.

Instrucciones:

Lee cada pregunta y responde sinceramente y con toda confianza lo que se pide. No es necesario indicar tu nombre, pues las respuestas forman parte de un estudio global.

1. Edad: _____ 2. Sexo: ___Femenino ___Masculino
3. Inscrito en: ___Preparatoria ___CCH
4. Si estás inscrito en Preparatoria, marca con una X el área que cursas actualmente:

ÁREA	
I. FÍSICO MATEMÁTICAS Y DE INGENIERÍAS	<input type="checkbox"/>
II. CIENCIAS BIOLÓGICAS Y DE LA SALUD	<input type="checkbox"/>
III. CIENCIAS SOCIALES	<input type="checkbox"/>
IV. HUMANIDADES Y ARTES	<input type="checkbox"/>

5. Si estás inscrito en CCH, marca con una X las materias que cursas actualmente:

CÁLCULO INTEGRAL Y DIFERENCIAL II	<input type="checkbox"/>
ESTADÍSTICA Y PROBABILIDAD II	<input type="checkbox"/>
CIBERNÉTICA Y COMPUTACIÓN II	<input type="checkbox"/>
BIOLOGÍA IV	<input type="checkbox"/>
FÍSICA IV	<input type="checkbox"/>
QUÍMICA IV	<input type="checkbox"/>
TEMAS SELECTOS DE FILOSOFÍA II	<input type="checkbox"/>
ADMINISTRACIÓN II	<input type="checkbox"/>
ANTROPOLOGÍA II	<input type="checkbox"/>
CIENCIAS DE LA SALUD II	<input type="checkbox"/>
CIENCIAS POLÍTICAS Y SOCIALES II	<input type="checkbox"/>
DERECHO II	<input type="checkbox"/>
ECONOMÍA II	<input type="checkbox"/>
GEOGRAFÍA II	<input type="checkbox"/>
PSICOLOGÍA II	<input type="checkbox"/>
TEORÍAS DE LA HISTORIA II	<input type="checkbox"/>
GRIEGO II	<input type="checkbox"/>
LATÍN II	<input type="checkbox"/>
LECTURA Y ANÁLISIS DE TEXTOS LITERARIOS II	<input type="checkbox"/>
TALLER DE COMUNICACIÓN II	<input type="checkbox"/>
TALLER DE DISEÑO AMBIENTAL II	<input type="checkbox"/>
TALLER DE EXPRESIÓN GRÁFICA II	<input type="checkbox"/>

6. ¿Cuál es tu actual promedio general de calificaciones?
7. ¿Qué carrera vas a estudiar terminando tu bachillerato?
8. ¿Por qué esa carrera?
9. En caso de no ser aceptado por la UNAM para cursar tu carrera, indica con una X el grado de importancia que tiene para ti cada una de las siguientes características, suponiendo que deseas elegir otra Universidad:

característica	Nada importante	poco importante	algo importante	muy importante
cercanía a mi domicilio/trabajo				
costos accesibles				
prestigio				
que el plan de estudios sea corto				
que otorgue becas				
que imparta otros idiomas				
que tenga cafetería				
que cuente con librería				
que tenga servicio médico				
que cuente con actividades culturales				
que cuente con actividades deportivas				
otra (específica):				

10. De la siguiente lista, indica con una cruz la cantidad que existe de cada uno en tu casa:

	0	1	2	3	4 o más
recámaras					
baño					
computadoras					
televisor					
automóvil					

11. ¿Tienes servicio de internet en tu casa?

_____NO _____SI (marca con una X):

↓

tarjeta prepagada	
contrato con empresa telefónica	
Televisión por cable	
banda ancha	
otro	

¿cuál? _____

→ **FIN DE LA ENTREVISTA** ←

¡Muchas gracias!

ANÁLISIS DE CORRESPONDENCIA

A partir de una tabla de contingencia se puede realizar una prueba χ^2 para determinar con un nivel de significancia si las variables en cuestión son independientes entre sí. El resultado de la prueba puede indicar que las variables no son independientes, pero se desconoce de qué manera se asocian entre sí. Una técnica gráfica que ayuda a conocer esta asociación es el análisis de correspondencia, el cual consiste en transformar las frecuencias relativas de la tabla de contingencia para obtener las coordenadas de cada columna y de cada renglón, de tal forma que se puedan plasmar en una sola gráfica.

Para ello se definen las siguientes matrices:

P de tamaño $n \times m$ como la matriz de frecuencias relativas. A esta se le llama matriz de correspondencia.

r es el vector ($n \times 1$) que contiene los valores de la suma de cada renglón de P

c es el vector ($m \times 1$) con los valores de la suma de cada columna de P

D_r es la matriz diagonal con los valores de r

D_c es la matriz diagonal con los valores de c

La Matriz P se puede descomponer en factores; es decir, se puede obtener a partir de la multiplicación de tres matrices:

$$P = AD_{\alpha}B'$$

A esto se le llama descomposición generalizada en valores singulares (SVD por sus siglas en inglés) y las matrices se definen como sigue:

A es la matriz ($n \times q$) cuyas columnas son los vectores singulares izquierdos

D_{α} es la matriz ($q \times q$) diagonal que contiene los valores singulares

B es la matriz ($m \times q$) cuyas columnas son los vectores singulares derechos

Y se cumple:

$$A'D_r^{-1}A = B'D_c^{-1}B = I$$

Donde I es la matriz identidad.

Para encontrar las matrices A , D_α y B es posible desarrollar primero la descomposición ordinaria en valores singulares de una matriz Q , la cual se define como:

$$Q = D_r^{-1/2} P D_c^{-1/2}$$

La SVD de Q es:

$$Q = U D V'$$

Donde:

D es una matriz diagonal ($k \times k$) cuyos elementos $\lambda_1, \lambda_2, \dots, \lambda_k$ son las raíces cuadradas positivas de los eigenvalores distintos de cero de $Q'Q$ o QQ'

U es la matriz ($n \times k$) cuyas columnas son los eigenvectores normalizados de QQ'

V es la matriz ($p \times k$) cuyas columnas son los eigenvectores normalizados de $Q'Q$

Dado que las columnas de U y de V son eigenvectores de matrices simétricas, entonces U y V son mutuamente ortogonales y se cumple: $U'U = V'V = I$

De esta manera:

$$A = D_r^{1/2} U$$

$$B = D_c^{1/2} V$$

$$D_\alpha = D$$

Y así se obtiene la descomposición generalizada de P :

$$P = A D_\alpha B'$$

Las coordenadas para los renglones de la tabla de contingencia se obtienen de la siguiente matriz:

$$F = D_r^{-1}AD$$

y las coordenadas para las columnas de la tabla de contingencia son los valores de:

$$G = D_c^{-1}BD$$

Para graficar en dos dimensiones se toman las dos primeras columnas de las matrices F y G , pero también se puede graficar en tres dimensiones. La gráfica será buena en la medida que represente la mayor parte de la inercia, cuyo valor indica la variación en los datos.

Inercia Total

La distancia χ^2 en una tabla de contingencia se puede expresar con matrices de la siguiente manera:

$$\chi^2 = \sum_{i=1}^a np_{i.} (r_i - c)' D_c^{-1} (r_i - c)$$

Donde:

i = número de renglón

a = total de renglones

n = total de observaciones

$P_{i.}$ = suma del renglón i de la matriz P

C = vector con los valores de la suma de cada columna de P

La matriz que contiene cada r_i , llamada R , se obtiene:

$$R = D_r^{-1} P = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_a \end{pmatrix} = \begin{pmatrix} \frac{p_{11}}{p_{1.}} & \frac{p_{12}}{p_{1.}} & \dots & \frac{p_{1b}}{p_{1.}} \\ \frac{p_{21}}{p_{2.}} & \frac{p_{22}}{p_{2.}} & \dots & \frac{p_{2b}}{p_{2.}} \\ \frac{p_{31}}{p_{3.}} & \frac{p_{32}}{p_{3.}} & \dots & \frac{p_{3b}}{p_{3.}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{a1}}{p_{a.}} & \frac{p_{a2}}{p_{a.}} & \dots & \frac{p_{ab}}{p_{a.}} \end{pmatrix}$$

Para calcular la inercia total se multiplican ambos miembros de la ecuación de χ^2 por $\frac{1}{n}$ y se

obtiene:

$$\text{Inercia total: } \frac{\chi^2}{n} = \sum_{i=1}^a p_{i.} (r_i - c)' D_c^{-1} (r_i - c)$$

Debido a que $r_i - c$ mide la distancia entre r_i y su promedio c , y $\sum_{i=1}^a p_{i.} = 1$, al dividir

χ^2 por n el resultado es una medida de la variación de los datos.

La inercia total es igual a la suma de los cuadrados de los valores singulares distintos de 1:

$$\frac{\chi^2}{n} = \sum_{i=1}^k \lambda_i^2$$

y las dimensiones necesarias para representar adecuadamente la matriz P se definen por la contribución que tiene cada valor singular distinto de 1 a la inercia total. En la medida que estos representen la mayor parte de la inercia total se utilizarán dos o tres dimensiones para graficar. Por ejemplo, la contribución de los dos primeros valores singulares distintos de 1 es:

$$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{i=1}^k \lambda_i^2}$$

Si el resultado representa la mayor parte, entonces sólo se utilizarán dos dimensiones para la gráfica. El rango de la matriz Q (de la cual se obtiene la SVD) es el mínimo de $(a - 1)$ y $(b - 1)$ y es igual al número de valores singulares distintos de 1 que contribuyen a la inercia total.

Ejemplo

La tabla de contingencia del área de la carrera por sexo, de la página 21, se transcribe aquí para tomarla como ejemplo:

Contingency Analysis of areacarr By sexo

Count Expected	femenino	masculino	
Fisico mate	72 92.1822	77 56.8178	149
biológicas y salud	133 116.929	56 72.0709	189
sociales	216 222.722	144 137.278	360
human y artes	84 75.4781	38 46.5219	122
Otras	45 42.6884	24 26.3116	69
	550	339	889

Test	ChiSquare	Prob>ChiSq
Pearson	20.764	0.0004

Las matrices correspondientes, obtenidas con el software *Mathematica* versión 5.0, son:

$$P = \begin{array}{|l} 0.081 \\ 0.15 \\ 0.24 \\ 0.094 \\ 0.051 \end{array} \begin{array}{l} 0.087 \\ 0.063 \\ 0.16 \\ 0.043 \\ 0.027 \end{array}$$

$$r = \begin{array}{|l} 0.168 \\ 0.213 \\ 0.405 \\ 0.137 \\ 0.0776 \end{array}$$

$$c = \begin{array}{c|c} 0.619 & \\ \hline 0 & 0.381 \end{array} \quad |$$

$$D_r = \begin{array}{c|ccccc} 0.168 & 0 & 0 & 0 & 0 \\ \hline 0 & 0.213 & 0 & 0 & 0 \\ 0 & 0 & 0.405 & 0 & 0 \\ 0 & 0 & 0 & 0.137 & 0 \\ 0 & 0 & 0 & 0 & 0.0776 \end{array} \quad |$$

$$D_c = \begin{array}{c|c} 0.619 & 0 \\ \hline 0 & 0.381 \end{array} \quad |$$

$$D_r^{1/2} = \begin{array}{c|ccccc} 0.409 & 0 & 0 & 0 & 0 \\ \hline 0 & 0.461 & 0 & 0 & 0 \\ 0 & 0 & 0.636 & 0 & 0 \\ 0 & 0 & 0 & 0.370 & 0 \\ 0 & 0 & 0 & 0 & 0.279 \end{array} \quad |$$

$$D_r^{-1/2} = \begin{array}{c|ccccc} 2.44 & 0 & 0 & 0 & 0 \\ \hline 0 & 2.17 & 0 & 0 & 0 \\ 0 & 0 & 1.57 & 0 & 0 \\ 0 & 0 & 0 & 2.70 & 0 \\ 0 & 0 & 0 & 0 & 3.59 \end{array} \quad |$$

$$D_c^{1/2} = \begin{array}{c|c} 0.787 & 0 \\ \hline 0 & 0.618 \end{array} \quad |$$

$$D_c^{-1/2} = \begin{array}{c|c} 1.27 & 0 \\ \hline 0 & 1.62 \end{array} \quad |$$

$$Q = D_r^{-1/2} P D_c^{-1/2} \text{ queda:}$$

$$\begin{array}{c|c} \hline 0.252 & 0.343 \\ \hline 0.413 & 0.221 \\ \hline 0.485 & 0.412 \\ \hline 0.324 & 0.187 \\ \hline 0.231 & 0.157 \\ \hline \end{array}$$

$$Q' = \begin{array}{c|ccccc} \hline 0.252 & 0.413 & 0.485 & 0.324 & 0.231 \\ \hline 0.343 & 0.221 & 0.412 & 0.187 & 0.157 \\ \hline \end{array}$$

$$U = \begin{array}{ccccc} 0.409395 & 0.747044 & -0.514954 & -0.00639728 & -0.0954098 \\ 0.461084 & -0.528176 & -0.33644 & -0.550121 & -0.304315 \\ 0.636356 & 0.160077 & 0.750584 & -0.0438192 & -0.0642601 \\ 0.37045 & -0.348601 & -0.200839 & 0.829735 & -0.111573 \\ 0.278595 & -0.125735 & -0.133857 & -0.0833449 & 0.938995 \end{array}$$

$$V = \begin{array}{cc} -0.786557 & -0.617517 \\ -0.617517 & 0.786557 \end{array}$$

$$V' = \begin{array}{cc} -0.786557 & -0.617517 \\ -0.617517 & 0.786557 \end{array}$$

$$D = \begin{array}{cc} 0. & 0. \\ 0. & 0.152827 \\ 0. & 0. \\ 0. & 0. \\ 0. & 0. \end{array}$$

$$A = D_r^{1/2} U \text{ queda:}$$

$$\begin{array}{ccccc} -0.167604 & 0.305836 & -0.21082 & -0.00261901 & -0.0390603 \\ -0.212598 & -0.243533 & -0.155127 & -0.253652 & -0.140315 \\ -0.404949 & 0.101866 & 0.477639 & -0.0278846 & -0.0408924 \\ 0.137233 & -0.129139 & -0.0744007 & 0.307375 & -0.041332 \\ 0.0776153 & -0.0350291 & -0.0372919 & -0.0232195 & 0.261599 \end{array}$$

$$B = D_c^{1/2} V \text{ queda:}$$

$$\begin{array}{cc} -0.618673 & -0.485713 \\ -0.381327 & 0.485713 \end{array}$$

$$B' = \begin{pmatrix} 0.618673 & -0.381327 \\ -0.485713 & 0.485713 \end{pmatrix}$$

Para calcular F se necesita:

$$D_r^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 4.70 & 0 & 0 & 0 \\ 0 & 2.47 & 0 & 0 \\ 0 & 0 & 7.29 & 0 \\ 0 & 0 & 0 & 12.9 \end{pmatrix}$$

$F = D_r^{-1}AD$ queda:

$$K = \begin{pmatrix} 1. & 0.278871 \\ 1. & -0.175064 \\ 1. & 0.0384439 \\ 1. & -0.143813 \\ 1. & -0.0689734 \end{pmatrix}$$

Para calcular G se necesita:

$$D_c^{-1} = \begin{pmatrix} 1.62 & 0 \\ 0 & 2.62 \end{pmatrix}$$

$G = D_c^{-1}BD$ queda:

$$C = \begin{pmatrix} -1. & -0.119983 \\ -1. & 0.194662 \end{pmatrix}$$

La matriz de los valores singulares es:

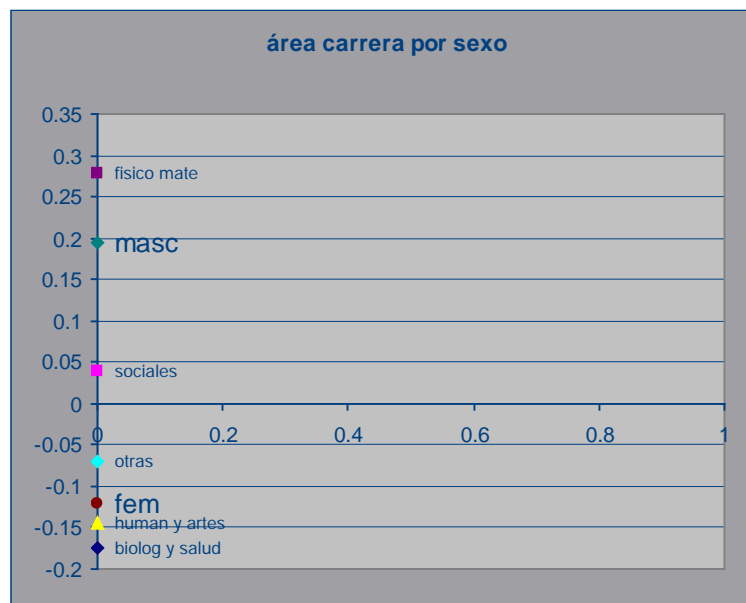
$$\begin{pmatrix} 1 \\ 0.152827 \end{pmatrix}$$

Por lo tanto, la primera columna de F y la primera columna de G están asociadas al valor singular 1. La inercia total es representada por un solo valor singular:

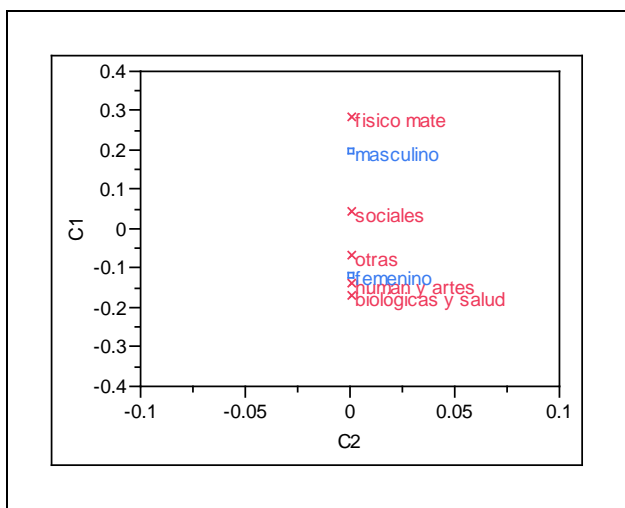
valores singulares	inercia	contribución
0.152827	0.023	1.000

Esto concuerda con el rango de la matriz Q : $\text{mínimo}(5 - 1, 2 - 1) = 1$ y significa que la gráfica refleja apropiadamente la variación con un solo eje, cuyas coordenadas son los valores de la segunda columna de las matrices F y G .

La gráfica de estas coordenadas en una hoja de cálculo de Excel es la siguiente:



Esta gráfica refleja las mismas coordenadas de la obtenida por el software JMP (página 22):



Para que las matrices F y G no contengan los valores asociados al valor singular 1, se puede sustituir la matriz P por $P - rc'$ en la matriz Q :

$$Q = D_r^{-1/2} (P - rc') D_c^{-1/2}$$

Entonces Q queda:

$$Q = \begin{bmatrix} 0.071 & 0.09 \\ 0.05 & -0.063 \\ 0.015 & 0.019 \\ 0.033 & -0.042 \\ 0.12 & -0.015 \end{bmatrix}$$

La SVD de Q es (UDV') :

U

$$U = \begin{bmatrix} 0.747 & -0.258 & -0.609 & 0.055 & -0.03 \\ 0.528 & -0.251 & -0.575 & -0.513 & -0.254 \\ 0.16 & -0.783 & 0.511 & -0.279 & -0.153 \\ 0.349 & -0.445 & -0.162 & 0.804 & -0.092 \\ 0.126 & -0.245 & -0.107 & -0.102 & 0.95 \end{bmatrix}$$

D

$$D = \begin{bmatrix} 0.153 & 0. \\ 0. & 0. \\ 0. & 0. \\ 0. & 0. \\ 0. & 0. \end{bmatrix}$$

V'

$$V' = \begin{bmatrix} 0.618 & 0.787 \\ 0.787 & 0.618 \end{bmatrix}$$

La matriz F resultante es:

$$F = \begin{bmatrix} 0.279 & 0. \\ 0.175 & 0. \\ 0.038 & 0. \\ 0.144 & 0. \\ 0.069 & 0. \end{bmatrix}$$

La matriz G es:

$$G = \begin{bmatrix} -0.12 & 0. \\ 0.195 & 0. \end{bmatrix}$$

Y el valor singular es:

1.02827

En el software JMP la primera columna de las matrices F y G corresponden al eje C1 y la segunda columna al eje C2.

BIBLIOGRAFÍA

COX, TREVOR F. *An introduction to multivariate data analysis*. USA, Hodder Arnold, 2005

EVERITT, BRIAN S. *Cluster analysis*. Fourth edition. USA, Arnold, 2001

LEVIN, RICHARD I. *Estadística para administradores*. México, Prentice-Hall, 2000

MONTGOMERY, DOUGLAS C. *Introduction to Linear Regression Analysis*. USA, John Wiley and sons, inc., 1992

MORRISON, DONALD F. *Multivariate statistical methods*. Fourth edition, USA, Thompson, 2005

RENCHE, ALVIN C. *Methods of multivariate analysis*. Second edition, USA, Wiley-Interscience, 2002

TAMHANE, AJIT C. y DUNLOP DOROTHY D. *Statistics and Data Analysis from elementary to intermediate*. USA, Prentice-Hall, 2000

WALPOLE, RONALD E. y MYERS, RAYMOND H. *Probabilidad y Estadística*. México, McGraw-Hill, 1992

WEIMER, RICHARD C. *Estadística*. México, CECSA, 1993