SISTEMA ESTADÍSTICO INFERENCIAL APLICADO A LAS ENCUESTAS DEL CENACAD PARA FACILITAR LA TOMA DE DECISIONES

Gisselle Guerra[1], Jorge Sánchez[2], Fabricio Echeverría, Msc.[3]
[1-3] Facultad de Ingeniería en Electricidad y Computación – Escuela Superior Politécnica del Litoral (ESPOL), Campus "Gustavo Galindo V.", Km. 30.5 Vía Perimetral, Apartado 09-01-5863. Guayaquil, Ecuador
[1] Email: gguerra@la.dole.com
[2] Email: j.sanchez@ott-computer.de
[3] Email: pechever@uniplex.com

Resumen

Este artículo describe la implementación de métodos estadísticos para la toma de decisiones, basado en las evaluaciones realizadas en el CENACAD. En la búsqueda por obtener datos fidedignos se ha tomado la iniciativa de fusionar dos ramas muy importantes, como lo son: la Estadística y la Minería de Datos para encontrar de esta manera patrones de conocimiento que sirvan para resolver inquietudes y a su vez tomar decisiones educativas.
La estadística inferencial permite evaluar datos de manera tal que se puedan obtener conclusiones que ayuden a beneficiar a quienes los están estudiando, en este caso los directivos interesados en verificar la calidad educacional de los profesores y materias que se dictan dentro de la ESPOL. Los 4 modelos estadísticos que serán implementados son: análisis de correspondencia, escalado multidimensional, análisis factorial y análisis de conglomerados Los módulos desarrollados no buscan competir con otras herramientas existentes en el mercado sino más bien poder utilizar medios locales para poder reducir los costos si se implementaran estos modelos en el sistema. Finalmente, se presentan las conclusiones y recomendaciones para futuros trabajos relacionados a esta tesis.

Palabras Claves: *estadística inferencial, CENACAD, análisis de correspondencia, escalado multidimensional, análisis de factores, clusterización.*

Abstract

This article describes the implementation of statistical methods for taking decisions based on the evaluations made in CENACAD. In the search to collect trustworthy data we have fused two very important sciences, which are: Statistics and Data Mining to find knowledge patterns that provides information and serve to make educative decisions. The inferencial statistic allows to evaluate data so conclusions can be obtained and help the decisions takers, in this case the directors interested in verifying the educational quality of the professors who dictate courses within ESPOL. The four statistical models that will be implemented are: analysis of correspondence, multidimensional scaling, factorial analysis and cluster analysis. The developed modules do not try to compete with other existing statistics tools, but rather use local means to be able to reduce the costs if these models in the system were implemented. Finally, the conclusions and recommendations for future works related to this thesis appears at the end of the article

1. Introducción

La estadística y los sistemas informáticos son utilizados en la actualidad como herramientas principales para la toma de decisiones en temas de gran relevancia. Desde la economía hasta la arquitectura, pasando por la física y la astronomía, el uso de sistemas estadísticos ha servido para definir conocimientos exactos y claros, además de conclusiones exactas y significativas[1].

El CENACAD en la ESPOL fue creado como herramienta para recolectar datos sobre la satisfacción de los estudiantes sobre sus profesores, las unidades educativas y las materias que les son dictadas durante cada semestre. En este sistema, se realizan encuestas y se mide la percepción que tienen los alumnos sobre el modo de dictar clases de sus instructores.

2. Análisis del Problema

Basados en los resultados obtenidos en las diversas encuestas, el CENACAD elabora diferentes reportes para sumarizar y tratar de explicar los resultados. Sin embargo, se presentan problemas a la hora de analizar los datos, debido a la gran cantidad de datos que se tienen, hacen necesario algún tipo de análisis estadístico para entenderlos y poder así tomar decisiones correctas.

Es este problema el que pretende solucionar el presente trabajo. Se desarrollarán cuatro modelos estadísticos, cada uno tratando de explicar los datos de una manera distinta y comprensible para el usuario final, es decir, el tomador de decisiones.

Los cuatro métodos a implementarse son: Análisis de Correspondencia, Escalado Multidimensional, Análisis Factorial y Análisis de Conglomerados.

3. Diseño de la Solución

El sistema planteado presentará los reportes de todos los análisis en la misma interfaz web del CENACAD.

Los cuatro métodos son análisis estadísticos complejos y que requieren de una gran cantidad de cálculo y procesamiento computacional. Por esta razón, los análisis no pueden ser hechos en tiempo real (por demanda), sino que deben estar hechos previamente para así mostrar los reportes cuando se requieran.

Para solucionar este inconveniente, los procesos estadísticos serán programados para ejecutarse en el servidor durante las noches, y de esta manera, los reportes presentados estarán listos todos los días con datos actualizados y reales.

3.1 Análisis de Correspondencia

El Análisis de Correspondencia permite mostrar gráficamente datos de una tabla de frecuencia entre dos variables[2]. Los análisis que serán presentados son:

1. Estudiantes vs Respuestas
   Se trata de descubrir en cada paralelo relaciones entre cada estudiante y una respuesta específica.
2. Preguntas vs Respuestas
   En cada paralelo se busca encontrar con que respuesta tiene relación cada pregunta del cuestionario.

Luego de realizar los dos análisis, se mostrarán como gráficos bidimensionales con puntos y cuadrados de diferente color, y las distancias entre ellos representarán la correlación que tienen los conjuntos de datos.

3.2 Escalado Multidimensional

El Escalado Multidimensional sirve para medir distancias entre elementos del mismo tipo, y así encontrar datos o grupos atípicos[3].

De esta manera, se plantea realizar tres análisis diferentes:

1. Cursos de un profesor
   Se comparan los promedios que han obtenido todas las materias dictadas por un profesor en particular
2. Cursos de una materia
   Son comparados los promedios que han obtenido todos los cursos de una materia en particular
3. Unidades Académicas por encuesta
   Se comparan los promedios que han obtenido las unidades académicas en una encuesta específica.

3.3 Análisis Factorial

El Análisis Factorial nos permite reducir el número de variables que se debe estudiar a un grupo de nuevas variables o factores más pequeño. De esta manera se facilita el análisis global de los datos, y puede ser usado como datos de entrada para nuevos algoritmos[4].

Se busca determinar factores en cada paralelo, y relacionarlos con las preguntas para obtener las preguntas más relevantes en cada uno.

Además, luego de realizado el análisis, se guardarán los puntajes sobre los nuevos factores para usarlos como dato de entrada al Análisis de Conglomerados.

3.4 Análisis de Conglomerados

El Análisis de Conglomerados, también conocido como Clusterización, nos ayuda a encontrar grupos dentro de un conjunto de datos[5]. En el presente proyecto, se usará para determinar posibles grupos de estudiantes dentro de un paralelo específico.

Para lograr esto, se pueden tomar como variables las respuestas que los estudiantes han dado a las preguntas del cuestionario. Sin embargo, esto representa un problema ya que el número de variables es alto (34), por lo cual se decidió clusterizar en base a los factores obtenidos en el análisis anterior, que en promedio son 4 o 5 factores (variables) por paralelo.

De esta forma, se logrará realizar el proceso de clusterización de una manera más rápida, simplificando el algoritmo y su ejecución.

4. Implementación del Proyecto

El motor estadístico del proyecto está implementado en JAVA, pues es un lenguaje que ofrece gran eficiencia en el manejo de la memoria y los cálculos complejos.

La interfaz web (las páginas de los reportes de los diferentes análisis), ha sido implementada en PHP, usando un framework MVC llamado Mojavi, que es el utilizado en el CENACAD.

4.1 Análisis de Correspondencia

Tal como se mencionó en el sección 3.1 han sido realizados 2 estudios para los cuales se ha seguido el mismo fundamento matemático:

1. Se realiza la obtención de datos
2. Se calculan las frecuencias entre preguntas/formularios vs respuestas.
3. Se dividen los valores para la suma de frecuencia
4. Normalización de los datos en base al cuadrado de las frecuencias relativas
5. Se calculan en base a la matriz Z, y a las frecuencias relativas
6. Los datos se guardan en las tablas de hecho
7. Estos datos servirán para mostrar los reportes en la interfaz web

Luego de este proceso matemático es posible mostrar los reportes y resultados obtenidos por el estudio. En la página web se muestran 4 pestañas.

La primera tabla corresponde a las frecuencias Estudiantes vs Respuestas, es decir, en la tabla se muestra cuántas veces cada estudiante respondió la alternativa 1, 2, 3, 4 o 5 en el cuestionario

Tabla Estudiantes - Respuestas	Gráfico Estudiantes - Respuestas	Tabla Preguntas - Respuestas	Gráfico Preguntas - Respuestas

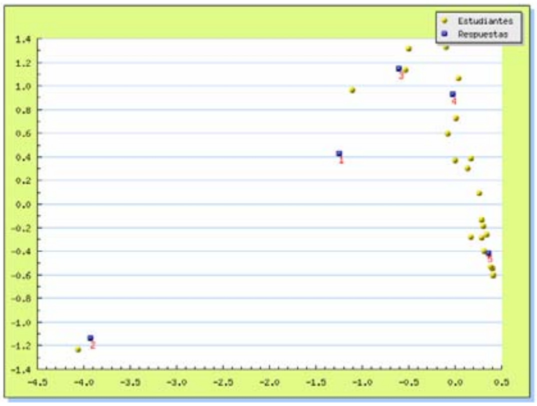
Tabla de Contingencia: Estudiantes - Respuestas

Estudiante	Respuesta					
	1	2	3	4	5	Total
1	1	0	8	7	18	34
2	0	0	0	6	28	34
3	0	0	1	7	26	34
4	0	0	4	11	19	34
5	2	29	3	0	0	34
6	0	0	1	7	26	34
7	0	0	0	0	34	34
8	0	0	27	2	5	34
9	0	0	1	6	27	34
10	0	0	0	0	34	34
11	0	1	1	5	27	34
12	0	2	14	15	3	34
13	2	5	17	9	1	34
14	4	0	2	16	12	34
15	0	0	0	0	34	34
16	1	0	3	19	11	34
17	0	0	0	0	34	34
18	0	0	0	12	22	34
19	0	0	1	16	17	34
20	0	0	0	29	5	34
21	0	0	0	0	34	34
22	0	0	3	2	29	34
23	0	0	0	1	33	34
24	0	0	0	0	34	34
25	0	0	1	0	33	34
26	0	0	3	0	31	34
27	0	0	4	29	1	34
28	0	0	0	0	34	34
Total	10	37	94	199	612	952

En la siguiente pestaña, se muestra el gráfico Estudiantes vs Respuestas. Los estudiantes se representan con puntos amarillos, y las respuestas, con cuadrados azules.

Tabla Estudiantes - Respuestas	Gráfico Estudiantes - Respuestas	Tabla Preguntas - Respuestas	Gráfico Preguntas - Respuestas

Gráfico: Estudiantes - Respuestas



Se presenta un ejemplo, donde se puede observar gran concentración de puntos amarillos (estudiantes) alrededor de la respuesta 5, lo cual sugiere que la mayoría de estudiantes tienen una percepción buena del paralelo.

La siguiente pestaña, es la tabla de frecuencias entre Preguntas y Respuestas que es parecida a la primera tabla, pero esta vez se muestran las preguntas.

Tabla Estudiantes - Respuestas	Gráfico Estudiantes - Respuestas	Tabla Preguntas - Respuestas	Gráfico Preguntas - Respuestas

Tabla de Contingencia: Preguntas - Respuestas

No.	Pregunta	Respuesta					
		1	2	3	4	5	Total
1	Asistí regularmente a clases	1	0	6	0	21	28
2	Asistí puntualmente a clases	2	0	4	0	22	28
3	Permanezco en clase durante toda la sesión programada	1	0	2	0	25	28
4	Participé de forma activa en clases, solicitando aclarar dudas, respondiendo preguntas, aportando con ejemplos	2	0	7	0	19	28
5	En general juzgo que mi esfuerzo y dedicación fueron apropiados	0	0	5	0	23	28
6	Al inicio del término proporcionó y explicó a los estudiantes la programación y políticas del curso.	0	1	2	8	17	28
7	Refleja una adecuada preparación de sus clases.	0	1	1	8	18	28
8	Cumple con la programación propuesta al inicio del curso.	0	2	1	7	18	28
9	Considera los conocimientos previos de los estudiantes para el dictado de la clase.	0	1	3	5	19	28
10	Fomenta la utilización del texto guía y los textos de referencia.	0	1	4	7	16	28
11	Presenta los contenidos de la clase de una manera comprensible.	0	1	3	6	18	28
12	Enfatiza durante la clase los puntos principales de los temas que expone.	0	1	1	8	18	28
13	Utiliza tecnologías de información para reforzar los contenidos de las clases.	0	1	3	8	16	28
14	Presenta ejemplos prácticos sobre lo tratado en clase.	0	1	1	9	17	28
15	Promueve el razonamiento de los temas tratados.	0	1	1	9	17	28
16	Contesta en forma satisfactoria las preguntas formuladas en clase.	0	1	2	8	17	28
17	Asigna actividades que requieren investigación por parte de los estudiantes.	0	1	3	6	18	28
18	Organiza durante la clase actividades de autoaprendizaje.	0	2	2	7	17	28
19	Estimula en el estudiante el pensamiento creativo y admite pensamientos diferentes.	1	1	2	5	19	28
20	Desarrolla los contenidos de la materia con un ritmo apropiado.	0	1	3	6	18	28
21	Utiliza técnicas de trabajo en grupo durante la clase.	1	1	2	6	18	28
22	Exalta las buenas costumbres y conciencia social durante la clase.	0	1	2	10	15	28
23	Fomenta la participación activa de los estudiantes en clase.	0	2	2	11	13	28
24	Es respetuoso y cordial en el trato con los estudiantes.	0	2	3	4	19	28
25	Estimula en la clase el comportamiento ético de los estudiantes.	0	2	2	6	18	28
26	Realiza evaluaciones periódicas (deberes, lecciones, proyectos, pruebas, etc.)	0	2	4	11	11	28
27	Cumple con las políticas de evaluación señaladas al inicio del curso.	0	1	4	7	16	28
28	Formula claramente las preguntas en las evaluaciones escritas.	0	1	2	7	18	28
29	Los temas en las evaluaciones son representativo de lo enseñado.	0	1	3	5	19	28
30	Hace conocer los resultados de las evaluaciones periódicas en plazos oportunos a sus estudiantes.	0	1	2	5	20	28
31	Califica procedimientos y resultados en las evaluaciones de los temas de examen.	1	1	2	4	20	28
32	Asiste puntualmente a clases (llega y se retira dentro del tiempo reglamentario)	0	2	3	5	18	28
33	Asiste regularmente a clases (frecuencia)	1	2	2	4	19	28
34	De acuerdo con sus respuestas anteriores, usted evalúa el desempeño del profesor(a) como:	0	1	5	7	15	28
Total		10	37	94	199	612	952

La cuarta y última pestaña, muestra el gráfico de Preguntas vs Respuestas fruto del Análisis de Correspondencia. En este gráfico las preguntas se muestran con puntos amarillos, y muestra su relación con las respuestas. Los puntos tienen el mismo comportamiento que en el gráfico anterior. Es decir, si existen preguntas cercanas a la respuesta 5, quiere decir que esas preguntas obtuvieron valoración alta en el paralelo.

4.2 Escalado Multidimensional

Para realizar los análisis del Escalado Multidimensional se debe de cumplir con los siguientes requisitos matemáticos:

1. Obtención de los datos
2. Se calcula la diferencia entre promedios de los elementos
3. La distancia se toma como el valor absoluto de la diferencia calculada
4. Matriz de similitud o covarianzas entre las distancias de los elementos
5. Se eligen las 2 columnas más representativas de la matriz generada
6. Los puntajes se calculan en base a los 2 valores propios mayores
7. Se obtienen las coordenadas para poder realizar el gráfico de distancias
8. Los datos se guardan en las tablas de hecho.
9. Estos datos se consultarán en la interfaz web.

Luego de obtenidos todos los datos anteriores, se realizan los reportes de cada uno de los análisis que a pesar de ser parecidos en estructura cada uno tiene una interpretación distinta. En este escrito solo se presentará el primero de los análisis

Escalado de Materias de un Profesor

Este reporte es realizado por profesor. Se presenta todos los cursos que el profesor ha dictado en una tabla, ordenados por promedio

Reporte: **ESCALADO MULTIDIMENSIONAL: PROFESORES**

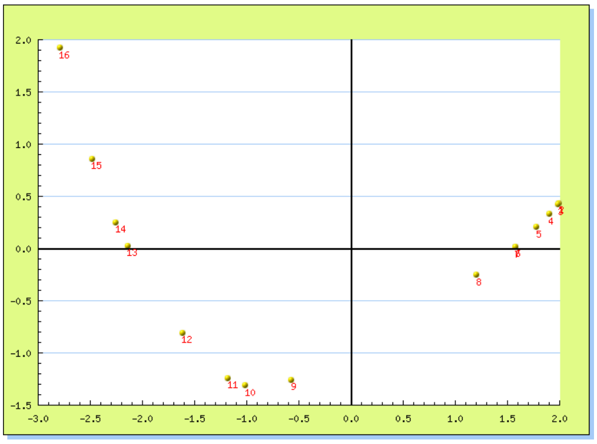
Profesor: EDGAR EDUARDO, CERVANTES BERNABE

Explicación del reporte

CURSOS QUE EL PROFESOR HA DICTADO

No.	Año	Término	Materia	Paralelo	Encuesta	Promedio
1	2006	1S	FUND.DE INGENIERIA	001	Encuesta de Materias Teóricas 2006 - 1S	**100.00**
2	2006	2S	FUND.DE INGENIERIA	001	Encuesta de Materias Teóricas 2006 - 2S	**100.00**
3	2006	2S	INGENIERIA I	001	Encuesta de Materias Teóricas 2006 - 2S	**99.85**
4	2007	1S	INGENIERIA II	001	Encuesta de Materias Teóricas 2007 - 1S	**99.08**
5	2004	2S	OCEANOGRAFIA DESCRIPTIVA	001	Encuesta de Materias teóricas 2004 - 2S	**98.25**
6	2006	1S	ING.PARA ACUACULTURA II	001	Encuesta de Materias Teóricas 2006 - 1S	**97.14**
7	2007	1S	ING. PARA ACUACULTURA I	001	Encuesta de Materias Teóricas 2007 - 1S	**97.11**
8	2005	2S	OCEANOGRAFIA DESCRIPTIVA	001	Encuesta de Materias Teóricas 2005 - 2S	**95.55**
9	2005	2S	MANEJO RECURSOS NATURALES	002	Encuesta de Materias Teóricas 2005 - 2S	**88.70**
10	2006	1S	ING. PARA ACUACULTURA I	001	Encuesta de Materias Teóricas 2006 - 1S	**86.94**
11	2005	1S	GEOGRAFÍA FÍSICA Y AMBIENTAL	003	Encuesta de Materias Teóricas 2005 - 1S	**86.20**
12	2004	2S	INGENIERIA I	001	Encuesta de Materias teóricas 2004 - 2S	**84.11**
13	2006	2S	ING.PARA ACUACULTURA II	001	Encuesta de Materias Teóricas 2006 - 2S	**81.16**
14	2005	2S	INGENIERIA I	001	Encuesta de Materias Teóricas 2005 - 2S	**80.37**
15	2005	1S	INGENIERIA II	001	Encuesta de Materias Teóricas 2005 - 1S	**78.00**
16	2005	1S	FUND.DE INGENIERIA	001	Encuesta de Materias Teóricas 2005 - 1S	**72.04**

Luego se muestra el gráfico resultante fruto del Escalado Multidimensional. Generalmente los datos se colocan en forma de parábola. Mientras más a la derecha los puntos, significa que ese paralelo ha obtenido mayor puntaje.



4.3 Análisis Factorial

Los resultados obtenidos en este análisis servirán de base para realizar el análisis de conglomerados.

Obtener los principales factores de las preguntas que se encuentran en el formularios se logra de la siguiente manera:

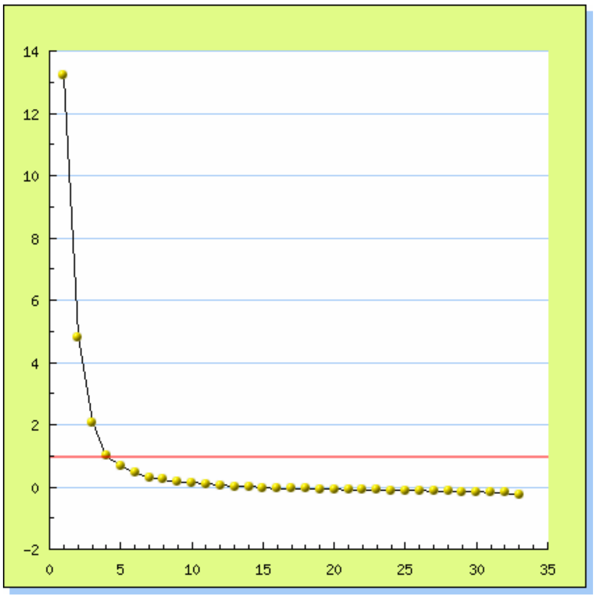
1. Obtención de los datos
2. Se usa el método de Factor Principal
3. Se usa la regla de Kaiser (1960)
4. Se extraen factores con valor propio > 1
5. Usando el método Varimax
6. En base a la Matriz de Comunalidades
7. Las preguntas que influyan en más de un 50%
8. Se calculan los puntajes de los formularios en base a los factores extraidos
9. Los datos se guardan en las tablas de hecho.
10. Estos datos servirán para futuros análisis.

La estructura del reporte es similar en estructura a la del Análisis de Correspondencia, manteniendo así una consistencia y coherencia visual en todos los reportes del sistema.

Lo primero que se muestra en el reporte es una tabla con los factores obtenidos junto con su valor propio y el porcentaje de varianza que explican. Todos los factores con valor propio mayor a 1 son resaltados, pues son éstos los que van a ser extraídos.

Factor	Valor	Varianza	Varianza acumulada
1	13.219	0.568	0.568
2	4.816	0.207	0.775
3	2.075	0.089	0.864
4	1.000	0.043	0.907
5	0.676	0.029	0.936
6	0.451	0.019	0.956
7	0.297	0.013	0.969
8	0.249	0.011	0.979
9	0.167	0.007	0.986
10	0.147	0.006	0.993
11	0.097	0.004	0.997
12	0.051	0.002	0.999
13	0.019	0.001	1.000
14	0.001	0.000	1.000
15	-0.035	-0.002	1.000
16	-0.048	-0.002	1.000
17	-0.052	-0.002	1.000
18	-0.062	-0.003	1.000
19	-0.069	-0.003	1.000
20	-0.079	-0.003	1.000
21	-0.088	-0.004	1.000
22	-0.091	-0.004	1.000
23	-0.099	-0.004	1.000
24	-0.107	-0.005	1.000
25	-0.110	-0.005	1.000
26	-0.121	-0.005	1.000
27	-0.127	-0.005	1.000
28	-0.140	-0.006	1.000
29	-0.151	-0.006	1.000
30	-0.158	-0.007	1.000
31	-0.174	-0.007	1.000
32	-0.183	-0.008	1.000
33	-0.262	-0.011	1.000

De igual manera, se muestra el gráfico de sedimentación (**¡Error! No se encuentra el origen de la referencia.**), que no es más que el gráfico de los factores y su correspondiente valor propio



Luego de esto se muestra una tabla que corresponde a la matriz de comunalidades, es decir, la relación que tiene cada factor con las preguntas. Aquí se puede ver cuáles preguntas están más relacionadas con los factores, mediante sus pesos. Mientras mayor sea el peso o la comunalidad de esa pregunta sobre el factor, significa que mayor es la relación entre ambos.

Pregunta	Factor 1	Factor 2	Factor 3	Factor 4
Asistí regularmente a clases	0.1519	1.1054	0.0230	0.1526
Asistí puntualmente a clases	0.1544	1.0919	0.0802	0.0560
Permanezco en clase durante toda la sesión programada	0.0345	1.0163	0.2571	0.0151
Participé de forma activa en clases, solicitando aclarar dudas, respondiendo preguntas, aportando con ejemplos	0.1309	1.1199	0.0999	0.0330
En general juzgo que mi esfuerzo y dedicación fueron apropiados	0.1383	1.0040	0.0273	0.1095
De acuerdo con sus respuestas anteriores, usted evalúa el desempeño del profesor(a) como:	0.4580	0.2165	0.3384	0.1033
Al inicio del término proporcionó y explicó a los estudiantes la programación y políticas del curso.	0.4316	0.1769	0.3597	0.1401
Refleja una adecuada preparación de sus clases.	0.4044	0.1182	0.3725	0.2539
Cumple con la programación propuesta al inicio del curso.	0.4545	0.1283	0.3459	0.1459
Relaciona los conocimientos previos de los estudiantes para el desarrollo de nuevos contenidos.	0.5175	0.1827	0.3695	0.1896
Presenta los contenidos de la clase de una manera comprensible.	0.6640	0.1004	0.4291	0.1611
Enfatiza durante la clase los puntos principales de los temas que expone.	0.4909	0.1269	0.3537	0.3500
Utiliza material de apoyo didáctico para reforzar los contenidos de las clases.	0.2906	0.1756	0.2813	0.4591
Presenta ejemplos apropiados para la comprensión de lo tratado en clase.	0.4417	0.1263	0.2315	0.5203
Promueve el razonamiento de los temas tratados.	0.7389	0.1077	0.2651	0.1677
Desarrolla los contenidos de la materia con un ritmo apropiado.	0.7511	0.2034	0.1223	0.2865
Fomenta el trabajo en equipo.	0.4581	0.2220	0.2012	0.3692
Tiene predisposición para aclarar dudas y ofrecer asesorías dentro y fuera de clases.	0.2286	0.0470	0.6407	0.1285
Facilita la participación activa de los estudiantes en clase	0.4561	0.1537	0.4298	0.2022
Es respetuoso y cordial en el trato con los estudiantes.	0.1164	0.0332	0.7784	0.1171
Estimula en la clase la formación de valores éticos y las buenas costumbres de los estudiantes.	0.0984	0.0540	0.7254	0.3199
Realiza evaluaciones periódicas (deberes, lecciones, proyectos, pruebas, etc.).	0.7161	0.2481	0.0630	0.3466
Formula claramente las preguntas en las evaluaciones.	0.3979	0.1312	0.1890	0.5159
Los temas en las evaluaciones son representativos del contenido del curso.	0.2373	0.1544	0.2464	0.6397
Califica procedimientos y resultados en las evaluaciones de los temas de examen.	0.3220	0.1062	0.1795	0.6826
Cumple con las políticas de evaluación señaladas para el curso.	0.2532	0.1884	0.2957	0.3947
Hace conocer los resultados de las evaluaciones periódicas en plazos oportunos a sus estudiantes.	0.2913	0.1227	0.3022	0.4119
Asiste puntualmente a clases (llega y se retira dentro del tiempo reglamentario).	0.4824	0.1533	0.2246	0.1630
Asiste regularmente a clases (frecuencia).	0.2226	0.0792	0.3435	0.3208
Contesta en forma satisfactoria las preguntas formuladas en clase.	0.2684	0.0696	0.5155	0.5038
Asigna actividades que requieren investigación por parte de los estudiantes.	1.0343	0.1934	0.0197	0.0550
Organiza durante la clase actividades de autoaprendizaje.	1.0429	0.2422	0.0800	0.1118
Promueve en el estudiante el pensamiento crítico.	0.7076	0.0686	0.2027	0.1977

Finalmente, se presenta un vínculo para ver el siguiente reporte, el cual será analizado en detalle en el siguiente capítulo, pero que utiliza los resultados de este Análisis Factorial como datos de entrada para su estudio.

4.4 Análisis de Conglomerados

Como se analizó en la sección anterior, los datos que servirán para realizar el estudio, serán los puntajes sobre los factores principales obtenidos en el capítulo anterior.

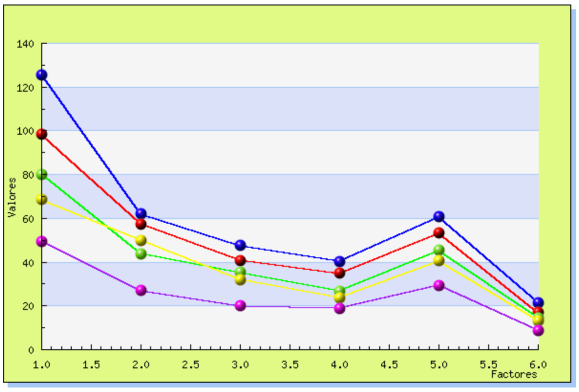
1. Se escogen los n primeros formularios como centros
2. n es el número de grupos en los que se está clusterizando.
3. Cada formulario se asigna al grupo más cercano
4. Si los centros han cambiado, se repite la asignación de formularios
5. Si los centros no han cambiado, se continúa con el proceso
6. Se usa el criterio de la SCDG
7. Si no se cumple el criterio, se aumenta un grupo y se recalcula.
8. Los grupos se ordenan en base a los centros
9. Los datos se guardan en las tablas de hecho.

10. Estos datos se usarán para las consultas por la web

Lo primero que se muestra en el reporte es una tabla que contiene los grupos en los que se dividió el paralelo, el número de estudiantes que contiene cada grupo, y el color con el que se lo va a representar en los siguientes gráficos.

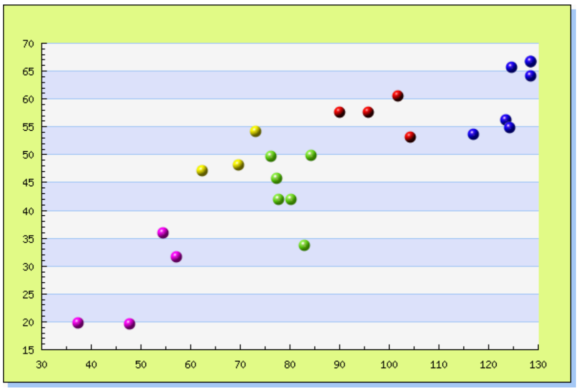
Grupo	Símbolo	# estudiantes
1		8
2		4
3		6
4		3
5		4

A continuación, se muestra en un gráfico bi-dimensional los centroides de los grupos (**¡Error! No se encuentra el origen de la referencia.**). En el eje de las X se encuentran los factores extraidos, y en el eje de las Y se especifica el puntaje alcanzado por el centroide de ese grupo, en ese factor



.

Para poder mostrar de una manera clara los grupos de estudiantes y la estructura del curso respecto a los grupos, también se muestra un gráfico de los estudiantes, proyectados sobre los 2 primeros factores (los más importantes), y cada estudiante se pinta del color que le corresponde según el grupo al que pertenece.



5. Pruebas y resultados experimentales

El sistema desarrollado permite a los profesores y directivos acceder y evaluar la información adicional sin tener que migrar los datos a otras herramientas de análisis estadístico ni aprender a usar dichas herramientas.

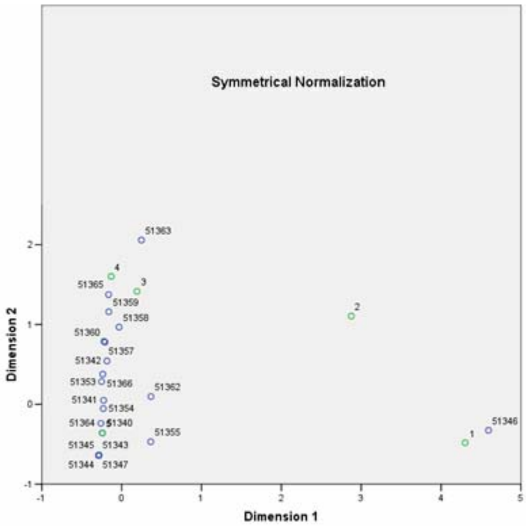
Una de las herramientas que provee análisis similares es el SPSS. Con este sistema se han realizado las pruebas de verificación y validación de los análisis que se hicieron en este proyecto de tesis.

A continuación se mostrarán las respuestas obtenidas en comparación con el SPSS.

5.1 Análisis de Correspondencia

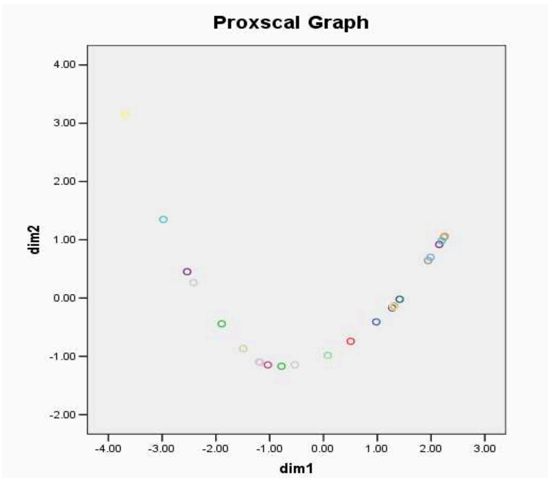
Para probar la validez del algoritmo desarrollado, los datos de algunos paralelos fueron extraídos, y luego analizados con el SPSS, usando la opción de Análisis de Correspondencia del mismo, y luego proyectando los datos resultantes en un gráfico.

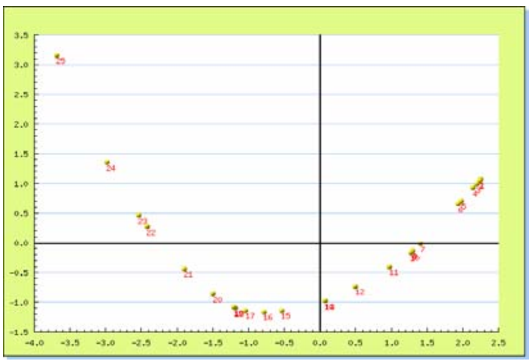
Como se puede observar, los gráficos son iguales y demuestran que el proceso se implementó correctamente. Estas mismas pruebas fueron hechas con algunos paralelos, encontrando siempre coincidencias entre el proceso desarrollado por el SPSS y por nuestro algoritmo.



5.2 Escalado Multidimensional

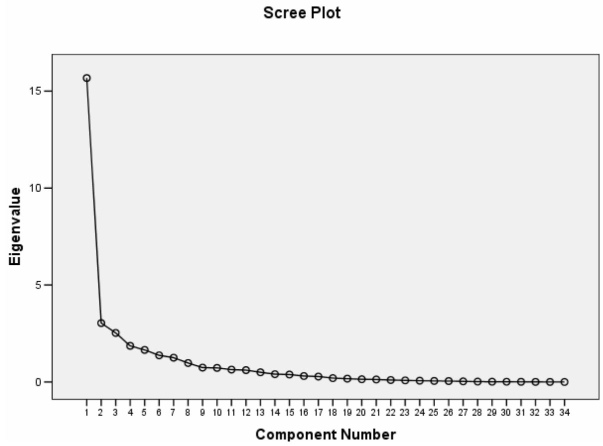
La validez del algoritmo desarrollado, se realizaron pruebas, donde fueron comparados los resultados de nuestro sistema con análisis hechos en la herramienta antes mencionada (SPSS) de la misma información, obteniendo los mismos datos de respuesta, probando así la efectividad del método.

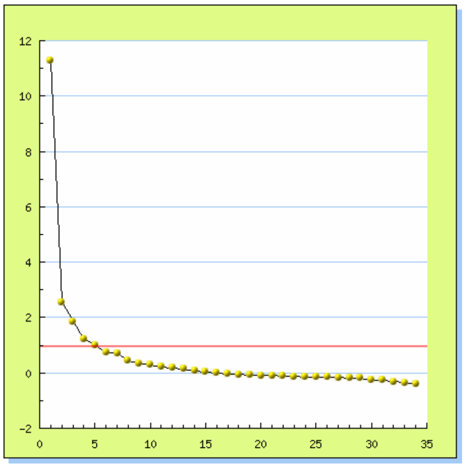




5.3 Análisis Factorial

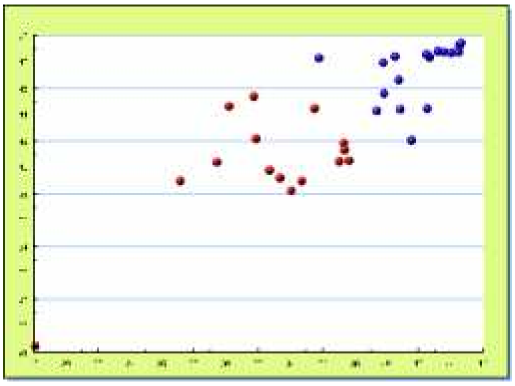
Para probar la validez de este análisis, fue también implementado en Matlab 6.1 y los resultados ploteados en SPSS 13, obteniendo los mismos datos y probando así que el algoritmo estaba correctamente implementado

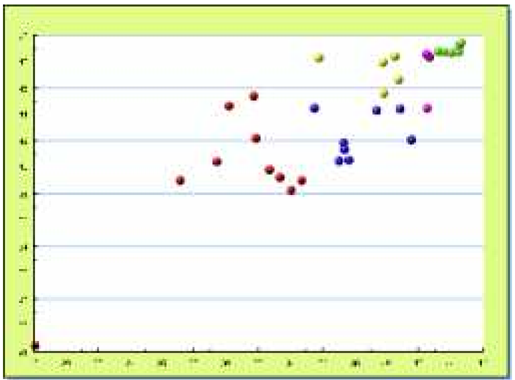


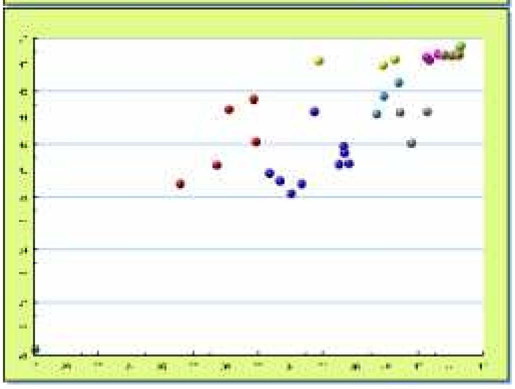


5.4 Análisis de Conglomerados

Se realizaron pruebas con distinto número de grupos, para comprobar visualmente que los grupos estén siguiendo un orden lógico. En las siguientes imágenes, se puede observar el proceso de clusterización en cada iteración, esto quiere decir, que en cada gráfico se aumenta en uno el número de grupos, para observar el comportamiento del agrupamiento.







Como se observa en las figuras el proceso de clusterización siguió un orden lógico, por lo cual se puede concluir que el algoritmo es válido y los resultados son confiables.

6. Conclusiones

La combinación de las técnicas de estadística inferencial junto con la minería de datos implementadas en esta tesis permite generar información y elaborar reportes que ayuden en la toma de decisiones para mejorar el rendimiento educacional que por años ha mantenido la ESPOL.

Los módulos desarrollados (análisis de correspondencia, escalado multidimensional, factorial y clusterización) reducen el número de variables que deben ser analizadas para emitir un juicio de valor por parte de cualquier directivo de alguna entidad educativa o empresarial, por lo podrían ser implementados en otras ramas tanto educativas como sociales.

Los reportes gráficos presentan la información de una manera más comprensible y amigable para el usuario que las tablas de resultados. Del desarrollo de este tema concluimos que es posible presentar gráficos/biplots que complementan los resultados mostrados numéricamente en las tablas de datos.

7. Referencias

[1] PEÑA, DANIEL, “Análisis de Datos Multivariantes”, McGraw-HILL. 2002, 171-180 p.

[2] SALVADOR FIGUERAS, M, "Análisis de Correspondencias", <http://www.5campus.com/leccion/correspondencias>. 2003, 1-2 p.

[3] PEÑA, DANIEL, “Análisis de Datos Multivariantes”, McGraw-HILL. 2002, 195-206 p.

[4] BELLIDO VASQUEZ, P, “Estadísticas para marketing (1) El Análisis Factorial”, <http://www.ilustrados.com/publicaciones/EpyukylkEFAVid hbVk.php>. 2003.

[5] SALVADOR FIGUERAS, M, “Análisis de conglomerados o cluster” <http://www.5campus.org/leccion/cluster>. 2001.