

Universidad Nacional Autónoma de México

“Por mi raza hablara el espíritu”



Facultad de Estudios Superiores Acatlán
Licenciatura en Matemáticas Aplicadas y Computación

**REDUCIENDO LA INCONFORMIDAD EN LOS GRUPOS
DE LA CARRERA DE ACTUARÍA**

Análisis multivariado

Práctica 1. Cluster

Peralta Cortés Jorge Alejandro

Mahil Maldonado Herrera

Índice

1. Introducción	2
2. Problemática	2
3. Técnicas estadísticas utilizadas	3
4. Análisis descriptivo	4
5. Análisis utilizando la técnica	6
6. Conclusión	8
7. Bibliografía	8
8. Anexo	8

Objetivo

Se requiere armar grupos para los alumnos de la carrera de actuaría, con el fin de que se reduzca lo más posible la inconformidad de los alumnos con el grupo que les toca.

1. Introducción

En este trabajo se intenta responder a la pregunta: ¿cómo podemos formar grupos que reduzcan la inconformidad de los alumnos?

Nuestro primer paso para este objetivo es obtener información que nos permita conocer las preferencias de los alumnos y así poder aplicar diferentes grupos a través de técnicas estadísticas, permitiendo reducir el factor de arbitrariedad en la formación de los diferentes grupos y formarlos de manera más objetiva, con detalles que no son tomados en cuenta por los funcionarios responsables.

Es importante resaltar que no se podrán satisfacer las necesidades de todos los alumnos, sino que se busca un equilibrio entre satisfacción y la factibilidad de formar los grupos, ya que por temas de personal y de espacio, podemos formar a los más cinco grupos con un máximo de cinco materias en cada uno.

2. Problemática

Cada semestre, las escuelas se enfrentan al problema de formar grupos para sus estudiantes. Esta tarea viene con la complejidad inherente de que los gustos, intereses, habilidades y horarios de los alumnos son muy variados.

Por esto mismo, es extremadamente difícil formar grupos que satisfagan a los alumnos o que se ajusten a sus necesidades, lo que provoca que cada semestre se presenten múltiples quejas y peticiones de cambio de grupo por parte de los mismos.

Además, lo bien que se ajuste un grupo a los alumnos puede contribuir a un aumento en el rendimiento académico, garantizando el éxito de los mismos, permitiendo formar profesionistas más capaces.

3. Técnicas estadísticas utilizadas

Para la realización de este trabajo se utilizó la técnica de cluster, que permite formar grupos de individuos que comparten características similares. Para formar estos clusters empleé el método k-means y agrupación por dendograma (método aglomerativo).

Descripción del método aglomerativo:

En este método se inicia tomando a cada individuo como un cúmulo, los cúmulos más similares se agrupan entre sí y así sucesivamente hasta que la disimilaridad entre distintos cúmulos va decreciendo, hasta llegar a tener un solo cúmulo.

Descripción del método k-means:

El entero k debe ser especificado, mediante un algoritmo partitivo divide los datos en k clusters (aleatoriamente separa los individuos en k grupos). Se corre el algoritmo para un rango de k grupos, por medio de un algoritmo iterativo hasta encontrar un óptimo con menor variabilidad dentro de los grupos (del centro del grupo a cada individuo) y al mismo tiempo que sea mayor la diferencia entre los grupos, el algoritmo, va sacando y metiendo individuos hasta encontrar un “óptimo”.

4. Análisis descriptivo

Hacemos una revisión rápida de la estructura de los datos, que tienen la siguiente estructura

index	Indivi...	Orden...	Clave	Nomb...
0	1	5	2030	Análisis d...
1	1	4	2031	Análisis ...
2	1	3	2049	Evaluació...
3	1	2	2046	Auditoría...
4	1	1	2047	Contabili...
5	2	5	2035	FINANZA...
6	2	4	2056	SEGURO ...
7	2	3	2036	FIANZAS
8	2	2	2040	MATEMÁ...
9	2	1	2046	AUDITOR...
10	3	5	2034	Economi...
11	3	4	2039	Modelos ...
12	3	3	2051	Finanzas ...
13	3	2	2049	Evaluació...
14	3	1	2044	Análisis d...

```
RangeIndex: 609 entries, 0 to 608
Data columns (total 4 columns):
#   Column                      Non-Null Count  Dtype
---  ---                      ---
0   Individuo                   609 non-null    int64
1   Orden de preferencia        609 non-null    int64
2   Clave                       608 non-null    float64
3   Nombre de la Materia        599 non-null    object
```

Figura 1: Estructura de los datos

Las columnas de este conjunto de datos no son de utilidad para el análisis, por lo que de ellas solo nos interesa obtener el número de elementos únicos que hay en cada una de ellas, llegando a los siguientes resultados:

- **Estudiantes:** 124 elementos únicos
- **Materias:** 29 elementos únicos

Después de obtener el nivel de interes total que hay por materia se realiza un histograma de este interes y delimitamos una franja que en el percentil del 25 por ciento, que marcará las materias a omitir si su interes está por debajo de este valor

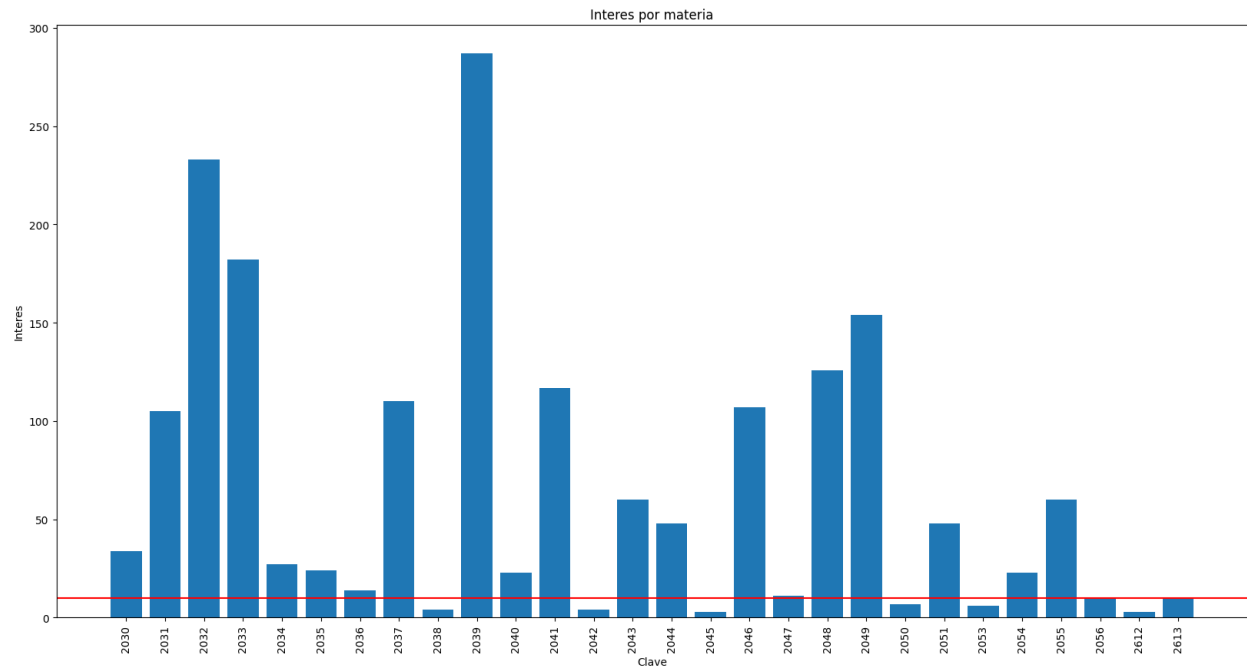


Figura 2: Histograma del interes por materia

Se adjunta además el análisis descriptivo de la variable de interes por materia:

```
count    28.000000
mean     65.714286
std      73.848715
min       3.000000
25%      10.000000
50%      30.500000
75%     107.750000
max     287.000000
dtype: float64
The number of courses with frequency greater than the 25% is: 22
```

Figura 3: Análisis descriptivo de la variable de interes por materia

5. Análisis utilizando la técnica

Se generó un dendograma para visualizar los grupos que se formaron, obteniendo el siguiente resultado:

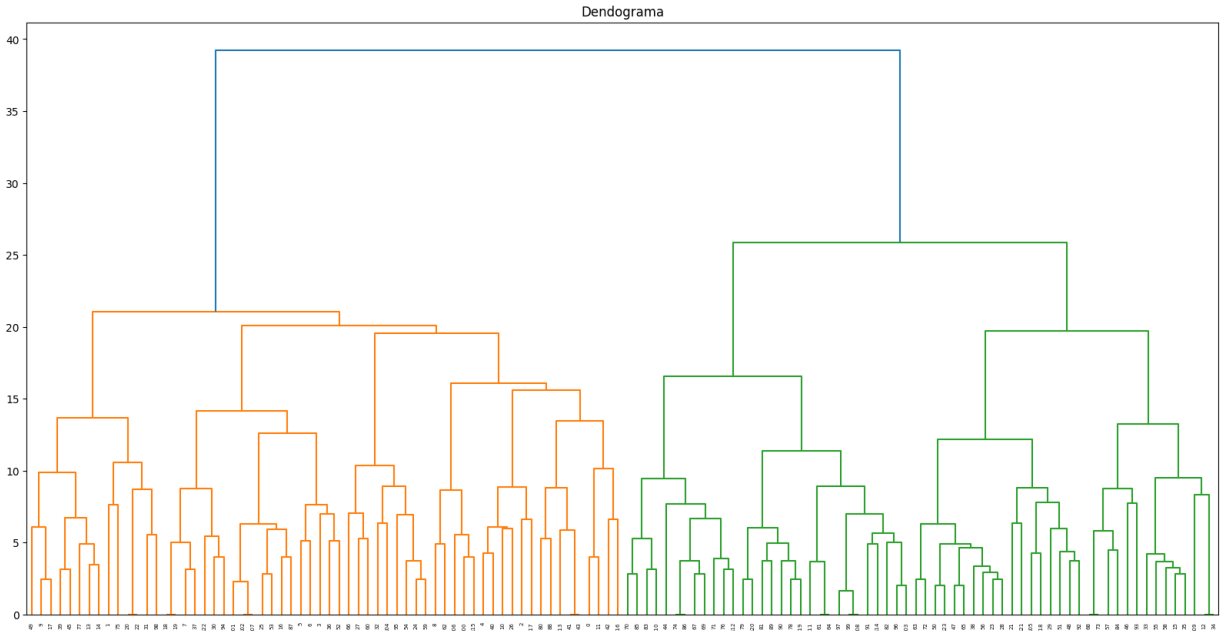


Figura 4: Dendograma

En este dendograma se puede observar que se formaron 2 grupos, ambos con muchos individuos, por lo que se decide realizar un análisis mediante el método de k-means, con el fin de obtener un número de grupos más adecuado.

Se realizó un diagrama de codo para determinar el número óptimo de clusters, obteniendo el siguiente resultado:

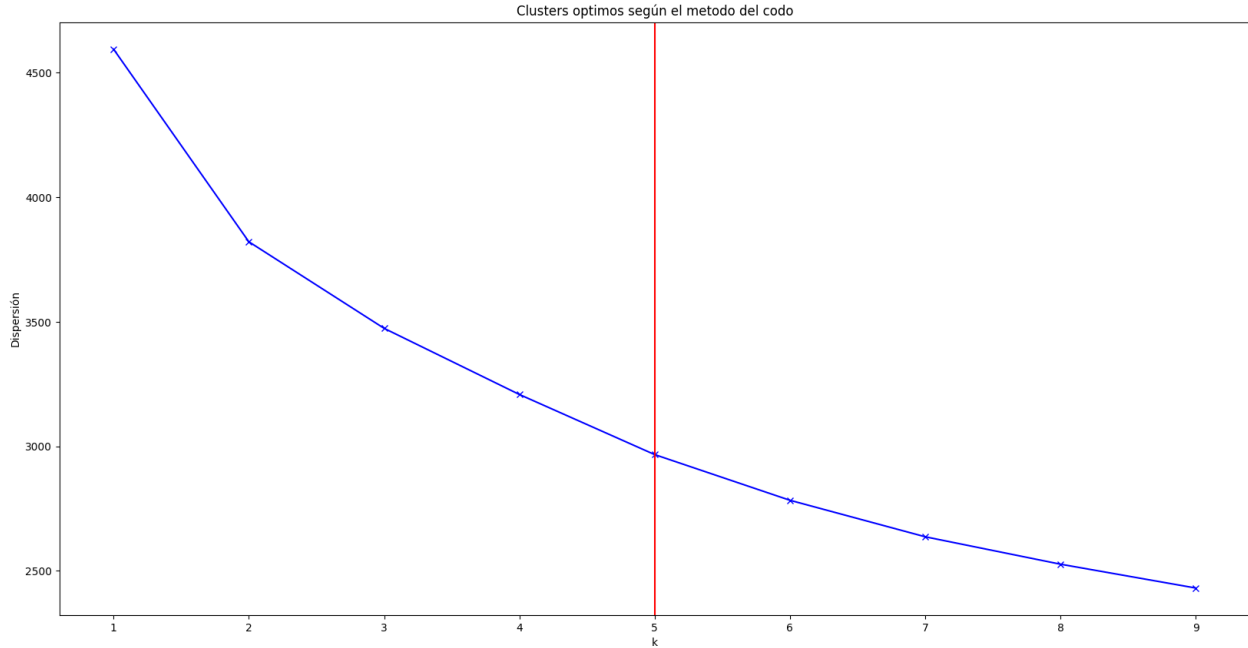


Figura 5: Diagrama de codo

El análisis mostró que por lo menos hasta la iteración 9 se siguen obteniendo mejoras en la suma de cuadrados, aunque la mejora respecto a nuestra limitante de 5 grupos no es tan significativa, por lo que se decide tomar este número de grupos para el análisis. Se realizó la agrupación de los datos, obteniendo los siguientes resultados:

```

4      27
0      26
2      26
3      25
1      20
Name: Cluster, dtype: int64

```

Figura 6: Cantidad de alumnos en cada grupo

Mediante un análisis de los centroides de cada grupo, se determinaron las materias que se impartirán en cada grupo, formando los siguientes grupos:

- **Grupo 0:** Derivados, Evaluación de Proyectos, Auditoria Actuarial, Análisis Multivariado, Modelos y simulación
- **Grupo 1:** Investigación de Operaciones II, Análisis de Regresión, Modelos y simulación, Muestreo, Evaluación de Proyectos
- **Grupo 2:** Estadística Bayesiana, Muestreo, Modelos y simulación, Análisis de Regresión, Series de Tiempo
- **Grupo 3:** Auditoria Actuarial, Modelos y simulación, Planeación financiera, Evaluación de Proyectos, Análisis de estados financieros

- **Grupo 4:** Análisis de Regresión, Modelos y simulación, Análisis Multivariado, Derivados, Series de Tiempo

6. Conclusión

A lo largo del análisis se vió que hay gran variación en el interes que tienen los alumnos por las materias, lo que planteó múltiples problemas a la hora de formar los grupos. Sin embargo, con el uso de técnicas estadísticas se logró formar grupos que puedan satisfacer a la mayoría de los alumnos, reduciendo la inconformidad de los mismos.

Este análisis puede ser de gran utilidad para las autoridades de la facultad, ya que les permite formar grupos de manera más objetiva, eficiente y con un menor grado de arbitrariedad, lo que puede contribuir a un aumento en el rendimiento académico de los alumnos.

Se debe resaltar que este análisis no es perfecto, ya que no se pueden satisfacer las necesidades de todos los alumnos retirando materias que son de baja demanda

Se presupone que este análisis es fácilmente replicable para otros semestres, lo que ahorraría tiempo y esfuerzo a las autoridades de la facultad.

7. Bibliografía

- Herrera, M. (2023). *Apuntes AR y AM*. Universidad Nacional Autónoma de México
- Microsoft Bing. (2023). Conversación personal sobre K-means.
- Towards Data Science. (24/10/2023). Understanding K-means Clustering in Machine Learning. Towards Data Science. Recuperado de [Towards Data Science](#)

8. Anexo

Para la realización de este trabajo se modificó la base de datos original, la cual estaba llena de errores de captura principalmente en los nombres de las materias.

Se removió la columna de “Nombre de la materia”, además se removieron las filas que contenían datos faltantes. Posteriormente se sustituyo los valores de “2511” por el valor de “2037” en la columna clave, ya que se trataba de un error de captura.

Posteriormente generé un dataframe con las veces que los alumnos eligieron cada materia, y el nivel de interes de las mismas.

Luego, se formó un nuevo conjunto de datos estableciendo la clave de las materias como columnas con el nivel de interes como valor. Asimismo se definió al individuo como el índice de las filas y se removieron aquellas materias que tuvieran un nivel de interes menor al del percentil 25.

Se adjunta el código utilizado para la limpieza de los datos y el análisis en el documento “Análisis Optativas”.